



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΣΗΜΑΤΩΝ, ΕΛΕΓΧΟΥ ΚΑΙ ΡΟΜΠΟΤΙΚΗΣ

ΕΡΓΑΣΤΗΡΙΟ ΟΡΑΣΗΣ ΥΠΟΛΟΓΙΣΤΩΝ, ΕΠΙΚΟΙΝΩΝΙΑΣ ΦΩΝΗΣ
ΚΑΙ ΕΠΕΞΕΡΓΑΣΙΑΣ ΣΗΜΑΤΟΣ

Οπτική Μοντελοποίηση Ανθρώπινου Προσώπου με Εφαρμογές σε Αναγνώριση

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

Επαμεινώνδα Π. Αντωνάκου

Επιβλέπων: Πέτρος Μαραγκός
Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούλιος 2011



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΣΗΜΑΤΩΝ, ΕΛΕΓΧΟΥ ΚΑΙ ΡΟΜΠΟΤΙΚΗΣ
ΕΡΓΑΣΤΗΡΙΟ ΟΡΑΣΗΣ ΥΠΟΛΟΓΙΣΤΩΝ, ΕΠΙΚΟΙΝΩΝΙΑΣ ΦΩΝΗΣ
ΚΑΙ ΕΠΕΞΕΡΓΑΣΙΑΣ ΣΗΜΑΤΟΣ

**Μοντελοποίηση και Αυτόματη Αναγνώριση
Εκφράσεων του Προσώπου**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

Επαμεινώνδα Π. Αντωνάκου

Επιβλέπων: Πέτρος Μαραγκός
Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 22 Ιουλίου 2011.

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

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

.....
Γεράσιμος Ποταμιάνος
Διευθυντής Ερευνών, Ε.Κ.Ε.Φ.Ε.

Αθήνα, Ιούλιος 2011.

.....
(Επαμεινώνδας Π. Αντωνάκος)

(Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Ηλεκτρονικών Υπολογιστών)

© Επαμεινώνδας Π. Αντωνάκος, 2011
Με επιφύλαξη παντός δικαιώματος. All rights reserved.

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

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

Περίληψη

Η Οπτική Μοντελοποίηση του ανθρώπινου προσώπου στοχεύει στη δημιουργία ενός όσο το δυνατόν πιο γενικευμένου μοντέλου του σχήματος και της υφής του προσώπου. Επιπλέον, καταφέρνει να ταιριάζει το μοντέλο προσώπου που δημιουργεί σε νέες στατικές εικόνες ή βίντεο. Το πρόβλημα αυτό έχει αναπτυχθεί πολύ τα τελευταία χρόνια και υπάρχουν αλγόριθμοι και τεχνικές που το αντιμετωπίζουν σε πραγματικό χρόνο.

Μια ιδιαίτερη υποκατηγορία του γενικότερου προβλήματος σε θέματα αναγνώρισης είναι η αυτόματη αναγνώριση των εκφράσεων του προσώπου και κατ' επέκταση των συναισθημάτων με εφαρμογές στις περιοχές της μελέτης συμπεριφοράς, αναγνώρισης δράσεων, επικοινωνίας ανθρώπου-ρομπότ, γραφικής με υπολογιστές, συναισθηματικής υπολογιστικής και αναγνώρισης φωνής.

Στόχος της διπλωματικής εργασίας είναι η ανάπτυξη τεχνικών και αλγορίθμων στην κατεύθυνση της οπτικής μοντελοποίησης του ανθρώπινου προσώπου σε πραγματικό χρόνο με χρήση τεχνικών της Όρασης Υπολογιστών και της Αναγνώρισης Προτύπων. Η προσπάθεια έχει ως στόχο τη βελτίωση της επίδοσης τεχνικών βασισμένων σε ειδικά, αντικειμενοστραφή μοντέλα, τα Active Appearance Models (AAMs), ώστε να επιτευχθεί το στοιχείο του πραγματικού χρόνου. Επιπλέον, στόχος είναι η ανάδειξη της σημασίας της οπτικής μοντελοποίησης σε θέματα αναγνώρισης. Δίνεται έμφαση στην ανάδειξη χαρακτηριστικών της καθολικότητας των εκφράσεων του ανθρώπινου προσώπου και στη διεξαγωγή πειραμάτων στην κατεύθυνση της αυτόματης αναγνώρισης συναισθημάτων με βάση την αναγνώριση των εκφράσεων προσώπου, χρησιμοποιώντας το σύστημα κωδικοποίησης Facial Action Coding System (FACS) που εισάχθηκε από τους Ekman και Friesen. Πιθανές εφαρμογές περιλαμβάνουν την αυτόματη αναγνώριση συναισθηματικής κατάστασης ενός ανθρώπου σε πραγματικό χρόνο, τη σύνθεση φωνής με συναισθηματική χροιά και την εξαγωγή μοντέλων συναισθηματικής προσοχής και σημαντικότητας από δεδομένα ταινιών.

Λέξεις κλειδιά: Σημεία ενδιαφέροντος, Παραμορφώσιμα μοντέλα, Ταίριασμα μοντέλου σε πραγματικό χρόνο, Εντοπισμός και αναγνώριση προσώπων, Αναγνώριση δράσεων, Αναγνώριση συναισθημάτων.

Abstract

Human Face Modeling targets on the creation of a general model of the shape and texture of human face. In addition to that, it manages to fit the face model on new static images or video. The problem has been developed during the past years and there are algorithms and techniques that introduce real time solutions.

A subcategory of the genaral problem on the domain of recognition is the automatic recognitian of facial expressions and emotions with applications in the regions of behaviour study, actions recognition, human-robot interaction, graphics with computers, emotional computing and voice recognition.

The purpose of this diploma thesis is the development of techniques and algorithms on the direction of human face visual modeling in real time with the usage of Computer Vision and Pattern Recognition techniques. The attempt targets on the improvement of the performance of techniques based on specially designed object-oriented models, Active Appearance Models (AAMs), so that the characteristic of real time is achieved. Additionally, one of the goals is the pointing of the importance of face modeling on topics of recognition. Emphasis is given on the highlight of features of the universality of human facial expressions and the realization of experiments on the direction of autimatic emotion recognition based on facial expressions recognition, using the system Facial Action Coding System (FACS) that was introduced by Ekman and Friesen. Possible applications include automatic real-time human emotion recognition, voice synthesis with emotional tinge and extraction of models of emotional attention and significance from movies data.

Keywords: Landmark points, Adaptive models, Real-time model fitting, Face detection and recognition, Actions recognition, Emotion recognition.

Πρόλογος και Ευχαριστίες

Η Όραση Υπολογιστών και η Αναγνώριση Προτύπων είναι δυο ιδιαιτέρως αναπτυσσόμενοι επιστημονικοί τομείς, άμεσα συνδεδεμένοι με τον τομέα της Ρομποτικής. Η μεγάλη τεχνολογική πρόοδος της Ρομποτικής απαιτεί την ύπαρξη αλγορίθμων και τεχνικών που να λειτουργούν σε πραγματικό χρόνο και να εξυπηρετούν την Επικοινωνία Ανθρώπου-Ρομπότ.

Το κίνητρο για την ενασχόλησή μου με τη συγκεκριμένη διπλωματική εργασία ήταν αφενός το μεγάλο ενδιαφέρον που υιοθέτησα για τα μαθήματα της Όρασης Υπολογιστών και της Αναγνώρισης Προτύπων και αφετέρου η προσωπική μου ευχαρίστηση για υλοποίηση πραγματοποιήσιμων, εφαρμόσιμων και πρακτικών συστημάτων. Αποτελεί μεγάλη προσωπική ικανοποίηση η μετατροπή της θεωρίας και των τεχνικών σε ένα αποδοτικό σύστημα που καταφέρνει να υλοποιήσει τους αλγορίθμους σε πραγματικό χρόνο.

Θέλω λοιπόν να ευχαριστήσω θερμά τον Καθηγητή Πέτρο Μαραγκό, επιβλέποντα της παρούσας διπλωματικής εργασίας, για την καθοδήγησή του και την εμπιστοσύνη που μου έδειξε κατά τη διάρκεια της υλοποίησής της. Θέλω επίσης να τον ευχαριστήσω γιατί οι διαλέξεις του στα μαθήματα της Όρασης Υπολογιστών και της Αναγνώρισης Προτύπων αλλά και η ελευθερία που μου έδωσε στην επιλογή του θέματος, συνέβαλλαν στη διαμόρφωση των επιστημονικών μου ενδιαφερόντων. Ακόμη, θέλω να ευχαριστήσω όλους τους υποψήφιους διδάκτορες του Εργαστηρίου Όρασης Υπολογιστών, Επικοινωνίας Λόγου και Επεξεργασίας Σημάτων, για τη βοήθειά τους σε οτιδήποτε τους ζήτησα.

Τέλος, θέλω να αφιερώσω την παρούσα διπλωματική εργασία στους γονείς μου Πέτρο και Ελένη και στον αδερφό μου Γιώργο και να τους ευχαριστήσω για τη στήριξή τους.

Περιεχόμενα

1 Τεχνική Εισαγωγή	16
1.1 Ορισμός του Προβλήματος	16
1.2 Περιεχόμενο της Παρούσας Διπλωματικής Εργασίας	18
I Ανίχνευση Προσώπου με Προαιρετική Ανίχνευση Δέρματος σε Πραγματικό Χρόνο	20
2 Εισαγωγή	21
2.1 Σχετική Εργασία στην Ανίχνευση Προσώπου	21
2.2 Προεπισκόπηση	23
3 Ανίχνευση Δέρματος	24
3.1 Μοντελοποίηση Δέρματος	24
3.1.1 Gaussian Κατανομή	24
3.1.2 Μίγμα Gaussian Κατανομών	25
3.2 Χρωματικός Χώρος Δέρματος	25
3.2.1 Παρουσίαση Χρωματικών Χώρων	25
3.2.2 Επιλογή Χρωματικού Χώρου για Ανίχνευση Δέρματος	27
3.3 Πειραματικά Αποτελέσματα	29
4 Ανίχνευση Προσώπου με Χρήση Haar Χαρακτηριστικών και AdaBoost Αλγορίθμου	33
4.1 Ολοκληρωτική Εικόνα	33
4.2 Χαρακτηριστικά Haar	34
4.3 Εξαγωγή Χαρακτηριστικών με τον Αλγόριθμο AdaBoost	35
4.4 Ταξινομητής Cascade	37
4.5 Πειραματικά Αποτελέσματα	37
5 Συνδυασμός Ανιχνευτή Προσώπου και Ανιχνευτή Δέρματος	43
5.1 Τελικό Σύστημα Ανίχνευσης Προσώπου	43
5.2 Πειραματικά Αποτελέσματα	44

II Οπτική Μοντελοποίηση Προσώπου με Active Appearance Models σε Πραγματικό Χρόνο 47

6 Εισαγωγή	48
6.1 Προεπισκόπηση	48
7 Εκπαίδευση Μοντέλου AAM	50
7.1 Μοντέλο Σχήματος Προσώπου	50
7.1.1 Landmark Σημεία	50
7.1.2 Ευθυγράμμιση Σχημάτων	52
7.1.2.1 Ευθυγράμμιση Δυο Σχημάτων (Procrustes Analysis) . .	52
7.1.2.2 Ευθυγράμμιση Πολλαπλών Σχημάτων (Generalized Procrustes Analysis)	54
7.1.3 Principal Component Analysis (PCA)	54
7.1.3.1 Αριθμός Αξόνων Μεταβλητότητας	56
7.1.4 Σύνθεση Σχημάτων Προσώπου	57
7.2 Μοντέλο Υφής Προσώπου	60
7.2.1 Ορισμός Υφής Προσώπου	60
7.2.2 Χαρτογράφηση Διανυσμάτων Υφής Προσώπου	60
7.2.2.1 Delaunay Τριγωνισμός	61
7.2.2.2 Piecewise Affine Warp	62
7.2.2.3 Δειγματοληψία Εικόνας με Διγραμμική Παρεμβολή . .	65
7.2.3 Φωτομετρική Ομαλοποίηση	68
7.2.4 Principal Component Analysis (PCA) Χαμηλής Μνήμης	69
7.2.5 Σύνθεση Υφής Προσώπου	71
7.3 Ανεξάρτητο Μοντέλο Προσώπου	74
7.3.1 Μοντέλο Σχήματος Προσώπου και Πόζα Προσώπου	74
7.3.1.1 Μετασχηματισμός Ομοιότητας	74
7.3.1.2 Σύνθεση Συνάρτησης Warp και Μετασχηματισμού Ομοιότητας	76
7.3.2 Μοντέλο Υφής Προσώπου και Συνθήκες Φωτεινότητας	76
7.3.2.1 Ομοπαραλληλικός Μετασχηματισμός Υφής	77
7.3.2.2 Δειγματοληψία Εικόνας και Ομοπαραλληλικός Μετασχηματισμός Υφής	78
7.3.3 Σύνθεση Προσώπου	78
8 Ταίριασμα Μοντέλου AAM	80
8.1 Ορισμός Προβλήματος και Συμβολισμοί	80
8.2 Ταίριασμα Μοντέλου με Inverse-Compositional Αλγόριθμο	82
8.2.1 Project-Out Inverse Compositional Αλγόριθμος	82
8.2.2 Simultaneous Inverse Compositional Αλγόριθμος	83
8.2.3 Βήμα Ανανέωσης Inverse Compositional Αλγορίθμου	83
8.3 Προσαρμοστικά Inverse Compositional Μοντέλα AAM	86
8.3.1 Θεωρητική Ανάλυση	86
8.3.1.1 Τάξη Αλγορίθμου	88

8.3.1.2 Πολυπλοκότητα Αλγορίθμου	89
8.3.2 Ψευδοκάδικας Αλγορίθμου	90
8.3.3 Ανάλυση Βημάτων Εκτός Επαναληπτικού Βρόχου	91
8.3.3.1 Gradient Ιδιοδιανυσμάτων Υφής	91
8.3.3.2 Ιακωβιανή Συνάρτησης Warp και Μετασχηματισμού Ομοιότητας	93
8.3.3.3 Πίνακας Προβολής και Πρότυπα Κίνησης	96
8.3.4 Ανάλυση Βημάτων Επαναληπτικού Βρόχου	101
8.3.4.1 Υπολογισμός Μεταβολών Παραμέτρων Σχήματος και Υφής Προσώπου	101
8.3.4.2 Ανανέωση Παραμέτρων Υφής Προσώπου	102
8.3.4.3 Ανανέωση Παραμέτρων Σχήματος Προσώπου με Priors .	102
8.4 Ανανέωση Διανύσματος Μέσης Υφής Προσώπου	105
8.5 Πειραματικά Αποτελέσματα	106
8.5.1 Παράμετροι Επιλογής Χρήστη	106
8.5.2 Ποσοστό Επιτυχίας	110
8.5.3 Χρονικό Κόστος	113
8.5.4 Παράδειγμα Εκτέλεσης Αλγορίθμου Ταιριάσματος AAM	117
III Εφαρμογή Μοντελοποίησης Προσώπου σε Αναγνώριση Εκφράσεων και Συναισθημάτων	122
9 Εισαγωγή	123
9.1 Προεπισκόπηση	123
10 Βάση δεδομένων CK+	124
10.1 Περιγραφή Χαρακτηριστικών, Πλεονεκτήματα και Μειονεκτήματα	124
10.2 Οπτική Μοντελοποίηση Προσώπου με τη Βάση Δεδομένων CK+	126
11 Σύστημα Κωδικοποίησης FACS	128
11.1 Χρησιμότητα Συστήματος FACS	128
11.2 Δομή του Συστήματος FACS	128
11.3 Κατηγοριοποίηση Συναισθημάτων	130
12 Αυτόματη Αναγνώριση Εκφράσεων Προσώπου και Συναισθημάτων	132
12.1 Εξαγωγή Χαρακτηριστικών	132
12.2 Ταξινόμηση με Support Vector Machine	133
12.3 Πειραματικά Αποτελέσματα Αναγνώρισης Εκφράσεων Προσώπου	133
12.4 Πειραματικά Αποτελέσματα Αναγνώρισης Συναισθημάτων	134
13 Συμπεράσματα	137
13.1 Μελλοντική Εργασία	138

Κατάλογος Σχημάτων

1.1 Διάγραμμα ροής συστήματος	18
3.1 Χρωματικός χώρος HSV	26
3.2 Τρισδιάστατη διασπορά χρώματος δέρματος στον χρωματικό χώρο HSV . .	28
3.3 Δισδιάστατη διασπορά χρώματος δέρματος στον χρωματικό χώρο HSV . .	29
3.4 Παράδειγμα ανίχνευσης δέρματος σε εικόνα της βάσης δεδομένων IMM . .	30
3.5 Παράδειγμα ανίχνευσης δέρματος σε εικόνα από τη webcam	31
3.6 Ανίχνευση δέρματος με πιθανοτικό αλγόριθμο για διάφορες τιμές του κατωφλίου D_{thres}	32
4.1 Ολοκληρωτική Εικόνα	34
4.2 Viola-Jones Χαρακτηριστικά που μοιάζουν με Haar	34
4.3 Κατανομές πιθανότητας θετικών και αρνητικών δειγμάτων προσώπων . .	35
4.4 Χαρακτηριστικά επιλεγμένα από τον αλγόριθμο AdaBoost	36
4.5 Ταξινομητής Cascade	38
4.6 Ποσοστό επιτυχίας του ανιχνευτή προσώπου στη βάση δεδομένων IMM για διάφορες τιμές κατωφλίου	40
4.7 Παραδείγματα ανίχνευσης προσώπου στη βάση δεδομένων IMM	40
4.8 Παραδείγματα ανίχνευσης προσώπου σε εικόνες με πολλά πρόσωπα ή/και πολύπλοκο παρασκήνιο	41
4.9 Παραδείγματα ανίχνευσης προσώπου με εμπόδια	42
5.1 Συνολικό σύστημα ανιχνευτή προσώπου	44
5.2 Παράδειγμα συνδυασμού ανιχνευτή προσώπου και ανιχνευτή δέρματος .	46
7.1 Εικόνα προσώπου που το σχήμα έχει ορισθεί με 58 landmark σημεία . .	51
7.2 Αρίθμηση landmark σημείων	51
7.3 Ευθυγράμμιση σχημάτων με Generalized Procrustes Analysis	55
7.4 Προσέγγιση σχήματος προσώπου ορισμένο από τη βάση δεδομένων με παραμόρφωση του μέσου σχήματος με ένα γραμμικό συνδυασμό ιδιοδιανυσμάτων του μοντέλου σχήματος	56
7.5 Ποσοστό μεταβλητότητας που αντιστοιχεί σε κάθε ιδιοτιμή του μοντέλου σχήματος προσώπου	58
7.6 Τα πρώτα πέντε πιο σημαντικά ιδιοδιανύσματα (modes of variation) του μοντέλου σχήματος προσώπου	59
7.7 Παράδειγμα χαρτογράφησης εικόνας υφής στο μέσο σχήμα προσώπου .	61

7.8 Τριγωνισμός Delaunay	62
7.9 Τριγωνισμός Delaunay σε σχήμα προσώπου	62
7.10 Συνάρτηση piecewise affine warp	63
7.11 Διγραμμική παρεμβολή	66
7.12 Παραδείγματα χαρτογράφησης υφής με διάφορες πόζες και εκφράσεις .	67
7.13 Παράδειγμα αποτυχημένης χαρτογράφησης με Delaunay τριγωνισμό .	68
7.14 Προσέγγιση υφής προσώπου ορισμένη από τη βάση δεδομένων με παραμόρφωση της μέσης υφής με ένα γραμμικό συνδυασμό ιδιοδιανυσμάτων του μοντέλου υφής	70
7.15 Τα πρώτα πέντε πιο σημαντικά ιδιοδιανύσματα (modes of variation) του μοντέλου υφής προσώπου	72
7.16 Ποσοστό μεταβλητότητας που αντιστοιχεί σε κάθε ιδιοτιμή του μοντέλου υφής προσώπου	73
7.17 Παραδείγματα μετασχηματισμού ομοιότητας	75
7.18 Παραδείγματα ομοπαραλληλικού μετασχηματισμού υφής	77
7.19 Παραδείγματα σύνθεσης προσώπου	79
 8.1 Βήμα Ανανέωσης Inverse Compositional Αλγορίθμου	84
8.2 Εικόνες gradient μέσης υφής προσώπου	92
8.3 Εικόνες gradient ιδιοδιανυσμάτων υφής προσώπου	92
8.4 Ιανωβιανές Σχήματος landmark σημείων 1, 10, 30, 40 και 50	93
8.5 Ιανωβιανές συνάρτησης warp για τα πέντε πιο σημαντικά ιδιοδιανύσματα σχήματος	94
8.6 Ιανωβιανές μετασχηματισμού ομοιότητας για τις δυο πρώτες παραμέτρους	95
8.7 Πρότυπα κίνησης και οι προβολές τους για τις παραμέτρους του μετασχηματισμού ομοιότητας	97
8.8 Πρότυπα κίνησης και οι προβολές τους για τις παραμέτρους σχήματος .	98
8.9 Απεικόνιση στιγμιοτύπου προτύπου κίνησης $M_\lambda(:, 3)$ μεταβάλλοντας τις πέντε πρώτες παραμέτρους υφής	99
8.10 Απεικόνιση στιγμιοτύπου προβολής προτύπου κίνησης $M'_\lambda(:, 3)$ μεταβάλλοντας τις πέντε πρώτες παραμέτρους υφής	100
8.11 Ανανέωση μέσης υφής προσώπου από το διάνυσμα A_0 στο A'_0	105
8.12 Καμπύλη απαιτούμενης μνήμης RAM ως προς την τιμή της καθολικής κλίμακας	108
8.13 Παραδείγματα αρχικοποίησης του αλγορίθμου ταιριάσματος για $s_{initial} = 1, 0.9, 0.8, 0.7$	109
8.14 Παράδειγμα εκτέλεσης αλγορίθμου ταιριάσματος με τιμές αρχικοποίησης $s_{initial} = 1$ και 0	110
8.15 Ποσοστό επιτυχίας ως προς την καθολική κλίμακα s_{global} για $r = m$. . .	111
8.16 Ποσοστό επιτυχίας ως προς την καθολική κλίμακα s_{global} για $r = \frac{m}{2}$. . .	112
8.17 Ποσοστό επιτυχίας ως προς την καθολική κλίμακα s_{global} για $r = 0$. . .	112
8.18 Χρονικό κόστος ανά επανάληψη για $r = m, r = \frac{m}{2}$ και $r = 0$	114
8.19 Παραδείγματα αρχικοποίησης του αλγορίθμου ταιριάσματος για $s_{initial} = 1, 0.9, 0.8, 0.7$	115

8.20 Συνολικό χρονικό κόστος αλγορίθμου ταιριάσματος για $r = m$, $r = \frac{m}{2}$ και $r = 0$	116
8.21 Απόδοση frames/second αλγορίθμου ταιριάσματος για $r = m$, $r = \frac{m}{2}$ και $r = 0$	116
8.22 Παράδειγμα εκτέλεσης αλγορίθμου ταιριάσματος AAM	119
8.23 Σφάλμα σχήματος προσώπου για το παράδειγμα του σχήματος 8.22	120
8.24 Μέγιστη μετατόπιση landmark σημείων για το παράδειγμα του σχήματος 8.22	120
8.25 Παράδειγμα ταιριάσματος υφής προσώπου	121
8.26 Σφάλμα υφής προσώπου για το παράδειγμα του σχήματος 8.25	121
10.1 Παράδειγμα ακολουθίας εικόνων από ουδέτερη έκφραση σε έκπληξη της βάσης δεδομένων CK+	125
10.2 Μέσο σχήμα και μέση υφή προσώπου της βάσης δεδομένων CK+	126
10.3 Στιγμιότυπα του μοντέλου σχήματος προσώπου CK+ για τα πέντε πιο σημαντικά ιδιοδιανύσματα	127
10.4 Στιγμιότυπα του μοντέλου υφής προσώπου CK+ για τα πέντε πιο σημαν- τικά ιδιοδιανύσματα	127
11.1 Παραδείγματα των 7 βασικών συναισθημάτων από τη βάση δεδομένων CK+	131

Κατάλογος Πινάκων

3.1 Χρόνοι εκτέλεσης αλγορίθμων ανίχνευσης δέρματος ανά εικόνα	31
7.1 Συμβολισμοί μοντέλου σχήματος προσώπου	57
7.2 Συμβολισμοί μοντέλου υφής προσώπου	71
8.1 Συμβολισμοί μοντέλου AAM	81
8.2 Μεγέθη πινάκων προσαρμοστικού αλγορίθμου ταιριάσματος AAM	91
8.3 Σχέση καθολικής κλίμακας και ανάλυσης του μέσου σχήματος προσώπου	107
8.4 Επεξήγηση εικόνων της βάσης δεδομένων IMM	111
8.5 Ποσοστό επιτυχίας αλγορίθμου ταιριάσματος με μέγιστη τιμή λάθους σχήματος 10 pixels	113
8.6 Τιμές παραμέτρου M_{displ} ως προς την καθολική κλίμακα και την τάξη του αλγορίθμου	117
8.7 Προτεινόμενες τιμές παραμέτρων συστήματος αναγνώρισης προσώπου .	117
11.1 Περιγραφή και συχνότητα εμφάνισης των δράεων AUs στη βάση δεδομένων CK+	129
11.2 Αναλυτική κωδικοποίηση συναισθημάτων με τις δράσεις AUs	130
11.3 Απλοποιημένη κωδικοποίηση συναισθημάτων με τις δράσεις AUs	130
12.1 Ποσοστά επιτυχίας αναγνώρισης εκφράσεων προσώπου με χρήση σχήματος προσώπου ως διάνυσμα χαρακτηριστικών	135
12.2 Ποσοστά επιτυχίας αναγνώρισης συναισθημάτων με χρήση σχήματος προσώπου ως διάνυσμα χαρακτηριστικών	136

Λιστ οφ Αλγορίτημας

1	Αλγόριθμος AdaBoost	36
2	Ταξινομητής Cascade	39
3	Τελικό σύστημα ανίχνευσης προσώπου με συνδυασμό ανιχνευτή προσώπου Viola-Jones και ανιχνευτή δέρματος	45
4	Procrustes Analysis	53
5	Generalized Procrustes Analysis (GPA)	54
6	Συνάρτηση Piecewise Affine Warp	65
7	Χαρτογράφηση διανυσμάτων υφής στο μέσο σχήμα προσώπου	66
8	Φωτομετρική Ομαλοποίηση	69
9	Σύνθεση συνάρτησης warp και μετασχηματισμού ομοιότητας	76
10	Αλγόριθμος Project-Out Inverse Compositional	82
11	Αλγόριθμος Simultaneous Inverse Compositional	83
12	Βήμα Ανανέωσης Inverse Compositional Αλγορίθμου	85
13	Προσαρμοστικός Αλγόριθμος Ταιριάσματος AAM	90
14	Υπολογισμός προβολών προτύπων κίνησης χωρίς την αποθήκευση του πίνακα προβολής	96

Κεφάλαιο 1

Τεχνική Εισαγωγή

1.1 Ορισμός του Προβλήματος

Κατά τη διάρκεια των τελευταίων δεκαετιών, η Οπτική Μοντελοποίηση του αινθρώπινου προσώπου έχει επικεντρώσει σε μεγάλο βαθμό το ενδιαφέρον της επιστημονικής κοινότητας, καθώς είναι ένα πρόβλημα με μεγάλη πληθώρα εφαρμογών σε αρκετούς τομείς. Το πρόβλημα της Οπτικής Μοντελοποίησης εφαρμόζει θεωρίες από τους επιστημονικούς τομείς της Όρασης Υπολογιστών (Computer Vision) και της Αναγνώρισης Προτύπων (Pattern Recognition), ενώ οι κυριότερες επιστημονικές εφαρμογές του είναι η αναγνώριση/ταυτοποίηση προσώπου, η αυτόματη αναγνώριση εκφράσεων προσώπου και η αυτόματη αναγνώριση ανθρώπινων συναισθημάτων, οι οποίες είναι εξαιρετικά σημαντικές στους τομείς της Μάθησης Μηχανής (Machine Learning) και της Επικοινωνίας Ανθρώπου-Μηχανής (Human-Robot Interaction). Η σημασία των επιστημονικών εφαρμογών της μοντελοποίησης προσώπου έχει αυξηθεί ιδιαίτερα τα τελευταία χρόνια με την ανάπτυξη εμπορικών εφαρμογών, όπως συστήματα οδηγικής ασφάλειας βασισμένα στο πρόσωπο (π.χ. έλεγχος αν τα μάτια του οδηγού κοιτάζουν ευθεία, έλεγχος συναισθηματικής κατάστασης οδηγού), συστήματα εικονικής πραγματικότητας για ηλεκτρονικά παιχνίδια ή άλλα γραφικά, αυτόματη δημιουργία animations σε κινηματογραφικές ταινίες, συστήματα διάγνωσης για διάφορους τομείς υγείας, συστήματα αυτοματισμών χώρου και ρομπότ οικιακής χρήσης.

Η Οπτική Μοντελοποίηση του αινθρώπινου προσώπου είναι το πρόβλημα της δημιουργίας ενός παραμετρικού και προσαρμοστικού μοντέλου προσώπου και του ταιριάσματος του μοντέλου αυτού σε νέα πρόσωπα. Η μοντελοποίηση είναι απαραίτητο να δημιουργεί κατάλληλα μοντέλα τόσο για το σχήμα του προσώπου όσο και για την υφή του, ενώ το ταίριασμα του μοντέλου οφείλει να επιτυγχάνεται και σε εικόνες με σημαντική απόκλιση από το μοντέλου που δεν έχουν συμπεριληφθεί στη διαδικασία της εκπαίδευσης. Στην παρούσα διπλωματική εργασία, η μοντελοποίηση γίνεται με χρήση των Ενεργών Μοντέλων Εμφάνισης (Active Appearance Models), τα οποία στη συνέχεια θα αναφέρονται με την αγγλική τους ανομασία ή τη συντομογραφία της (AAM).

Η χρήση των παραμορφώσιμων μοντέλων που έχουν τη δυνατότητα να ταιριάζουν σε αντικείμενα (αινθρώπινο πρόσωπο) εικόνων εισόδου είναι σύνηθης στον τομέα της

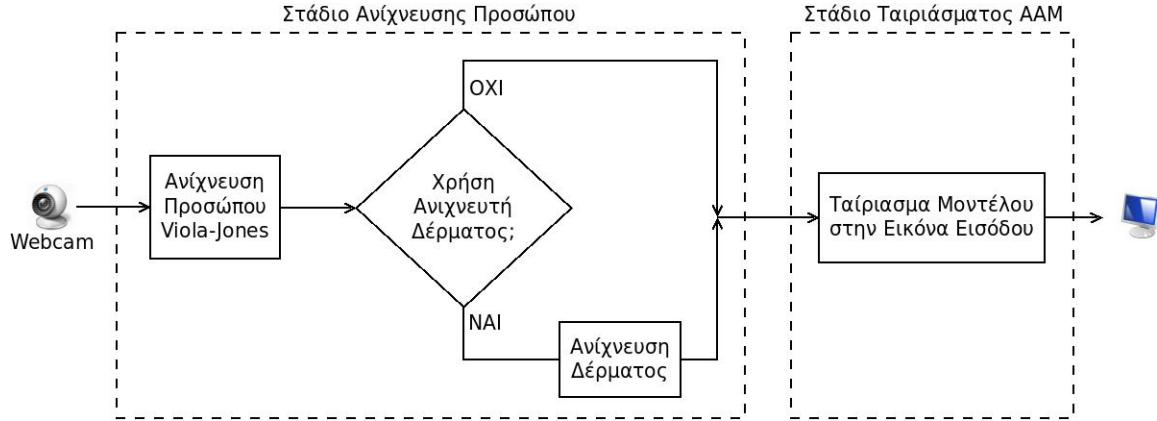
Όρασης Υπολογιστών. Έχουν αναπτυχθεί αρκετές προσεγγίσεις και τεχνικές δημιουργίας παραμορφώσιμων μοντέλων. Τα Active Contours ή Snakes [Ka87] είναι καμπύλες ελαχιστοποίησης ενέργειας που παραμορφώνονται σύμφωνα με εσωτερικές και εξωτερικές δυνάμεις. Τα Active Blobs [ScIs03] προκαλούν παραμόρφωση σύμφωνα με φυσικές ιδιότητες, όπως η ελαστικότητα και η ακαμψία. Η διαφορά των Active Blobs με τα AAMs είναι ότι τα πρώτα παραμορφώνουν ένα στατικό μοντέλο υφής και τα δεύτερα αλλάζουν ταυτόχρονα το σχήμα και την υφή κατά τη διάρκεια της διαδικασίας ταιριάσματος (fitting). Τα Active Shape Models (ASMs) [TaCo92], όπως και τα Active Appearance Models, ανήκουν στην κατηγορία Point Distribution Models (PDM), δηλαδή είναι μέθοδοι βασισμένες σε landmark σημεία όπου η εκπαίδευση της μεταβλητότητας του σχήματος γίνεται με Principal Component Analysis (PCA).

Τα Active Appearance Models (AAM) [CET01, MaBa04] ήταν η εξέλιξη των ASMs. Διαφέρουν στο ότι χρησιμοποιούν ένα πλήρες μοντέλο σχήματος και υφής. Μπορούν να μοντελοποιήσουν την μεταβλητότητα της υφής διαφόρων αντικειμένων όπως προσώπων, χειρών ή αυτοκινήτων από μια κατάλληλη βάση δεδομένων εικόνων. Εφαρμόζονται τόσο για σύνθεση όσο και για ανάλυση εικόνων σε εφαρμογές όπως η παρακολούθηση αντικειμένων σε video (object tracking), η αναγνώριση προσώπου (face recognition) και η σύνθεση προσώπου. Έχουν αναπτυχθεί αρκετές επεκτάσεις των AAM. Τα Constrained Active Appearance Models [CoTa01] χρησιμοποιούν πιθανοτική μέθοδο με maximum a-posteriori (MAP) διατύπωση που επιτρέπει τον ορισμό prior περιορισμών στα σημεία που ελέγχουν το σχήμα.

Το πιο σημαντικό κομμάτι της αναγνώρισης αντικειμένων με Active Appearance Models είναι το ταίριασμα του μοντέλου σε νέες εικόνες εισόδου. Οι αλγόριθμοι Project-Out Inverse Compositional (POIC) [MaBa04] και Simultaneous Inverse Compositional (SIC) [GMB05] βασίζονται στη μέθοδο Inverse Compositional Image Alignment που μετατρέπει το ταίριασμα των AAMs σε μια Steepest Descent (SD) μέθοδο. Η αρχική διατύπωση των AAM ανέφερε ότι η ανανέωση της πόζας του προσώπου έπρεπε να γίνεται συνθετικά (compositionally), αλλά η ανανέωση των παραμέτρων του σχήματος και της υφής να γίνεται προσθετικά (additively). Ωστόσο, οι Baker και Matthews έδειξαν ότι και η ανανέωση των παραμέτρων σχήματος πρέπει να γίνεται συνθετικά (compositionally). Όλες οι προηγούμενες εργασίες χρησιμοποιούσαν έναν σταθερό AAM πίνακα ταιριάσματος fitting matrix για την επαναληπτική ανανέωση των παραμέτρων του μοντέλου κατά το ταίριασμά του σε μια νέα εικόνα. Τα αποτελέσματα όμως δεν είναι τόσο καλά όταν κάποιος δουλεύει με μοντέλα που επιτρέπουν μεγάλη απόκλιση εμφάνισης από το μέσο πρότυπο υφής, όπως AAMs που έχουν εκπαιδευτεί σε μεγάλες βάσεις εικόνων πολλαπλών ανθρώπων [BaHa05, CoTa06]. Οι Παπανδρέου και Μαραγκός [PaMa08] προτείνουν δυο προσαρμογές για τον αλγόριθμο ταίριασματος: (a) μια προσαρμογή στον AAM fitting matrix και (β) έναν τρόπο ανανέωσης του μέσου προτύπου υφής. Ακόμη αναλύουν πως η prior πληροφορία μπορεί να ενσωματωθεί στον inverse compositional αλγόριθμο και να καθοδηγήσει τη διαδικασία ταίριασματος του AAM.

Τονίζεται ότι πριν το στάδιο του ταίριασματος του μοντέλου σε νέες εικόνες απαιτείται η Ανίχνευση Προσώπου (Face Detection). Η Ανίχνευση προσώπου στην προκειμένη περίπτωση λειτουργεί ως το στάδιο αρχικοποίησης του αλγορίθμου ταίρι-

άσματος AAM. Στην παρούσα διπλωματική εργασία αναλύεται η συνδυασμένη χρήση της μεθόδου ανίχνευσης προσώπου των Viola-Jones με μια μέθοδο ανίχνευσης δέρματος (Skin Detection) με στόχο την όσο το δυνατόν μικρότερη συχνότητα ανίχνευσης false-positive προσώπων.



Σχήμα 1.1: Διάγραμμα ροής συστήματος

Στο Σχήμα 1.1 φαίνεται το διάγραμμα ροής του συνολικού συστήματος που σχεδιάστηκε στα πλαίσια της παρούσας διπλωματικής εργασίας. Η υλοποίηση έχει γίνει κυρίως σε MATLAB με χρήση C++ μόνο στα σημαντικά και χρονικά πολύπλοκα τμήματα των αλγορίθμων. Έχει σχεδιαστεί πλήρες GUI με αρκετές επιλογές, γεγονός που δημιουργεί μια εκτελέσιμη εφαρμογή έτοιμη για χρήση. Η σχεδίαση των γραφικών της εφαρμογής έγινε με τέτοιο τρόπο ώστε να είναι εύχρηστη, απλή και ευκατανόητη. Έχει δωθεί μάλιστα η δυνατότητα χρήσης δυο βάσεων δεδομένων εικόνων, για μεγαλύτερη εποπτεία αποτελεσμάτων.

1.2 Περιεχόμενο της Παρούσας Διπλωματικής Εργασίας

Το περιεχόμενο της παρούσας διπλωματικής εργασίας αποτελείται από τρία μέρη:

- Στο πρώτο μέρος παρουσιάζονται δυο τεχνικές **Ανίχνευσης Δέρματος** και ένας αλγόριθμος **Ανίχνευσης Προσώπου** βασισμένος σε Haar χαρακτηριστικά που λειτουργούν σε πραγματικό χρόνο. Ακόμη, δημιουργείται ένα τελικό σύστημα ανίχνευσης προσώπου που χρησιμοποιεί τους προηγούμενους ανιχνευτές δέρματος και προσώπου.
- Στο δεύτερο μέρος αναλύεται το θέμα της **Οπτικής Μοντελοποίησης** του σχήματος και της υφής του ανθρώπινου προσώπου με Active Appearance Models. Υπάρχει εκτενής ανάλυση και απόδειξη των μαθηματικών σχέσεων τόσο για το στάδιο της εκπαίδευσης όσο και για το στάδιο του ταιριάσματος του μοντέλου σε νέες εικόνες σε πραγματικό χρόνο.

- Στο *τρίτο μέρος* εξηγούνται οι εφαρμογές της Οπτικής Μοντελοποίησης σε θέματα Αναγνώρισης και παρουσιάζονται δυο πειράματα **αναγνώρισης εκφράσεων προσώπου και συναισθημάτων** με αυτόματη εξαγωγή χαρακτηριστικών.
- Η διπλωματική εργασία συνοδεύεται από λογισμικό με πλήρως σχεδιασμένο GUI.

Μέρος Ι

**Ανίχνευση Προσώπου με
Προαιρετική Ανίχνευση Δέρματος σε
Πραγματικό Χρόνο**

Κεφάλαιο 2

Εισαγωγή

Η ανίχνευση προσώπου (*face detection*) είναι ιδιαίτερα σημαντική διαδικασία για αρκετές εφαρμογές της όρασης Υπολογιστών. Μπορεί να εφαρμοστεί σε μεμονωμένες εικόνες, ανιχνεύοντας ένα ή και περισσότερα πρόσωπα που υπάρχουν σε αυτές, αλλά και σε ακολουθίες video. Στη δεύτερη περίπτωση το πρόβλημα μεταφράζεται σε *ανίχνευση και παρακολούθηση προσώπου* (*face detection and tracking*), καθώς απαιτείται η συνεχόμενη ανίχνευση ενός προσώπου κατά την παρουσία του σε όλα τα frames του video. Σε αυτή τη διπλωματική εργασία η ανίχνευση προσώπου χρησιμοποιείται σαν το στάδιο της αρχικοποίησης για την αναγνώριση προσώπου με Active Appearance Models και κατ' επέκταση την αναγνώριση των εκφράσεών του.

Η μέθοδος ανίχνευσης και παρακολούθησης προσώπου που παρουσιάζεται σε αυτό το μέρος αποτελείται από δυο βήματα: ανίχνευση δέρματος και ανίχνευση προσώπου. Συγκεκριμένα η ανίχνευση ανθρωπίνου δέρματος χρησιμοποιείται ως φίλτραρισμα στα πιθανά πρόσωπα που έχουν εντοπισθεί με τη μέθοδο της ανίχνευσης προσώπου. Όποια περιοχή από αυτές που έχουν εντοπιστεί δεν καλύπτεται από ανθρώπινο δέρμα, τότε απορρίπτεται ως λανθασμένη ανίχνευση προσώπου.

2.1 Σχετική Εργασία στην Ανίχνευση Προσώπου

Μια απλή μέθοδος ανίχνευσης προσώπου είναι χρησιμοποιώντας εικόνες με ελεγχόμενο παρασκήνιο (background), όπως μονόχρωμο ή προκαθορισμένο και σταθερό. Το μειονέκτημα αυτής της μεθόδου είναι ότι η αναποτελεσματικότητά της σε περίπτωση αφαίρεσης του παρασκηνίου.

Μια δεύτερη μέθοδος είναι η εύρεση προσώπων με τη βοήθεια του χρώματος [Kar97]. Αρχικά, γίνεται ένα φίλτραρισμα της εικόνας ανιχνεύοντας περιοχές στις οποίες είναι πολύ πιθανό το χρώμα που υπάρχει να προέρχεται από ανθρώπινο δέρμα. Στη συνέχεια γίνεται ανίχνευση προσώπου με εξαγωγή πληροφοριών από τις περιοχές αυτές, κατάλληλων για να υποδείξουν την ύπαρξη προσώπου. Αυτό το αποτέλεσμα επιτυγχάνεται με ένα συνδυασμό κατωφλιαποίησης και μαθηματικής μορφολογίας, ικανό να αναγνωρίσει χαρακτηριστικά που δείχνουν την παρουσία προσώπου. Τα πειραματικά αποτελέσματα της μεθόδου αυτής δείχνουν εξαιρετική απόδοση στην περίπτωση που ένα πρόσωπο καταλαμβάνει μεγάλο ποσοστό της συνολικής εικόνας. Τα

μειονεκτήματα της μεθόδου είναι ότι υπάρχει μεγάλο ρίσκο λανθασμένης ανίχνευσης προσώπων, όταν αυτά καταλαμβάνουν πολύ μικρό χώρο στην εικόνα ή όταν υπάρχει μεγάλο μέρος δέρματος από το υπόλοιπο ανθρώπινο σώμα (για παράδειγμα, ένας γυμνός άνθρωπος να καταλαμβάνει το μεγαλύτερο μέρος της εικόνας). Φυσικά, ένα ακόμα μεγάλο μειονέκτημα είναι ότι η μέθοδος αυτή δεν μπορεί να λειτουργήσει για όλα τα ήδη χρώματος του ανθρώπινου δέρματος και κάτω από μεγάλη ποικιλία συνθηκών φωτισμού.

Μια ακόμη μεθοδολογία είναι η ανίχνευση προσώπου σε χρωματιστές εικόνες χρησιμοποιώντας Principal Component Analysis (PCA) [MeMu99]. Συγκεκριμένα, μια περιοχή εικόνας χαρακτηρίζεται ως πρόσωπο αν η απόστασή της από το μοντέλο προσώπου στο χώρο προσώπου που έχει δημιουργηθεί από την εκπαίδευση είναι μικρότερη από ένα επιλεγμένο κατώφλι. Το μειονέκτημα της μεθόδου αυτής είναι ότι αποτυγχάνει σημαντικά αν το πρόσωπο καταλαμβάνει μικρή περιοχή εικόνας.

Μια ακόμη μέθοδος ανίχνευσης προωπού είναι αυτή της Hausdorff Distance (HD) [JKF01]. Η απόσταση HD μεταξύ δυο σημείων χρησιμοποιείται στην Επεξεργασία Εικόνας σαν ένα μέτρο ομοιότητας μεταξύ ενός γενικευμένου μοντέλου προσώπου και πιθανά στιγμιότυπα αυτού του μοντέλου μέσα στην εικόνα. Ο ορισμός της απόστασης HD για δυο σημεία A και B είναι

$$H(A, B) = \max(h(A, B), h(B, A))$$

όπου

$$h(A, B) = \max_{a \in A} \left[\min_{b \in B} \|a - b\| \right] \quad \text{ή} \quad h(A, B) = \frac{1}{|A|} \sum_{a \in A} \min_{b \in B} \|a - b\|$$

Ο δεύτερος ορισμός της Hausdorff Distance με το μέσο όρο των σημείων είναι καταλληλότερος για προβλήματα Αναγνώρισης Προτύπων. Αν θεωρήσουμε ότι A και B είναι η εικόνα και το πρόσωπο αντίστοιχα, τότε ο στόχος είναι η εύρεση της παραμέτρου του μετασχηματισμού, έτσι ώστε να ελαχιστοποιηθεί η HD μεταξύ του μετασχηματισμένου μεοντέλου και του A . Συγκεκριμένα

$$d_{p^-} = \min_{p \in P} H(A, T_p(B))$$

όπου $T_p(B)$ είναι το μετασχηματισμένο μοντέλο. Η τιμή d_{p^-} είναι η τιμή απόστασης που ταιριάζει καλύτερα ως προς τη θέση και την κλίμακα. Ένα σημαντικό μειονέκτημα της μεθόδου αυτής είναι η δημιουργία ενός σωστού μοντέλου προσώπου $T_p(B)$, ώστε να καλύπτεται μεγάλη ποικιλία προσώπων.

Μια επιπρόσθετη μέθοδος ανίχνευσης προσώπου βασίζεται σε Νευρωνικά Δίκτυα [RBK98]. Συγκεκριμένα υπολογίζεται μια πυραμίδα εικόνων ώστε να γίνει ανίχνευση προσώπων σε διάφορες κλίμακες. Ένα παράθυρο σταθερού μεγέθους μετακινείται μέσα σε κάθε εικόνα της πυραμίδας και το περιεχόμενό του διορθώνεται για μη-ομοιόμορφο φωτισμό και υπόκειται στη διαδικασία της εξισορρόπησης ιστογράμματος (histogram equalization). Το αποτέλεσμα τροφοδοτείται σε κάποια παράλληλα νευρωνικά δίκτυα που επιτυγχάνουν την ανίχνευση προσώπου. Οι έξοδοι συνδυάζονται με το λογικό KAI (πύλη AND), ώστε να μειωθούν οι λανθασμένες ανιχνεύσεις. Το

μειονέκτημα του αλγορίθμου είναι η ταχύτητα, γεγονός που τον καθιστά ακατάλληλο για ανίχνευση προσώπου σε ακολουθία βίντεο.

Τέλος η πιο επιτυχημένη μέθοδος ανίχνευσης προσώπου προτάθηκε από τους Viola & Jones το 2001 [ViJo01a, ViJo01b]. Η βασική αρχή της μεθόδου είναι η μετακίνηση ενός παραθύρου, ικανού για ανίχνευση προσώπου, κατά μήκος μιας εικόνας εισόδου. Η θεωρία της Επεξεργασίας Εικόνας προτείνει την κλιμάκωση της εικόνας εισόδου με διαφορετικές κλίμακες και μεγέθη ώστε να εκτελεστεί ο ανιχνευτής σταθερού μεγέθους σε αυτές. Ωστόσο αυτή η διαδικασία είναι ιδιαίτερα προβληματική σε ταχύτητα, εξαιτίας του υπολογισμού των εικόνων διαφορετικού μεγέθους. Αντιθέτως, η μέθοδος των Viola-Jones προτείνει την κλιμάκωση του ανιχνευτή και την εφαρμογή του πολλές φορές στην εικόνα, κάθε φορά με διαφορετικό μέγεθος. Το σημαντικό είναι ότι ο ανιχνευτής που χρησιμοποιείται είναι ανεξάρτητος της κλίμακας και απαιτεί τον ίδιο αριθμό υπολογισμών για οποιοδήποτε μέγεθος. Τα βασικά συστατικά του ανιχνευτή είναι η *ολοκληρωτική εικόνα* (*integral image*), κάποια απλοϊκά ορθογώνια χαρακτηριστικά που μοιάζουν με τα *Haar wavelets* και ο αλγόριθμος *AdaBoost*. Ο αλγόριθμος που προκύπτει είναι ιδιαίτερα γρήγορος και αποτελεσματικός.

2.2 Προεπισκόπηση

Στο Μέρος I της διπλωματικής εργασίας αναλύονται δυο μέθοδοι ανίχνευσης ανθρώπινου δέρματος και η μέθοδος ανίχνευσης προσώπου των Viola-Jones με Haar wavelets χαρακτηριστικά. Κατόπιν, εξηγείται ο τρόπος που συνδυάζονται οι δυο μέθοδοι για ένα τελικό σύστημα ανίχνευσης προσώπου. Τέλος, παρουσιάζονται κάποια πειραματικά αποτελέσματα και εξηγείται πως η ανίχνευση προσώπου λειτουργεί ως διαδικασία αρχικοποίησης για το πρόβλημα της αναγνώρισης προσώπου.

Κεφάλαιο 3

Ανίχνευση Δέρματος

Σε αυτό το κεφάλαιο αναλύονται δυο μέθοδοι ανίχνευσης ανθρώπινου δέρματος.

3.1 Μοντελοποίηση Δέρματος

Ο στόχος της μοντελοποίησης δέρματος είναι η διαφοροποίηση μεταξύ των σημείων μιας εικόνας που αποτελούν δέρμα και αυτών που δεν αποτελούν δέρμα. Για να επιτευχθεί αυτό χρειάζεται ένα μετρικό σύστημα, το οποίο να μετράει την απόσταση μεταξύ του χρώματος ενός pixel και του χρωματικού τόνου του δέρματος.

3.1.1 Gaussian Κατανομή

Με βάση αυτά, το πιο σύνηθες παραμετρικό μοντέλο δέρματος είναι το Gaussian μοντέλο. Συγκεκριμένα η κατανομή του χρώματος του δέρματος μπορεί να μοντελοποιηθεί με μια ελλιπτική Gaussian συνάρτηση πυκνότητας πιθανότητας που ορίζεται ως

$$P(c|skin) = \frac{1}{2\pi |\Sigma_s|^{\frac{1}{2}}} e^{-\frac{1}{2}(c-\mu_s)^T \Sigma_s^{-1} (c-\mu_s)} \quad (3.1)$$

όπου c ένα διάνυσμα χρώματος. Το διάνυσμα μέσης τιμής μ_s και ο πίνακας συνδιακύμανσης Σ_s δίνονται από τις σχέσεις

$$\mu_s = \frac{1}{n} \sum_{j=1}^n c_j \quad \text{και} \quad \Sigma_s = \frac{1}{n-1} \sum_{j=1}^n (c_j - \mu_s)(c_j - \mu_s)^T$$

όπου n το πλήθος των δειγμάτων δέρματος c_j και $j = 1, \dots, n$. Η πιθανότητα $P(c|skin)$ μπορεί να χρησιμοποιηθεί απευθείας σαν ένα μέτρο που δείχνει κατά πόσο το χρώμα c μπορεί να θεωρηθεί δέρμα. Εναλλακτικά, το ίδιο αποτέλεσμα μπορεί να επέλθει με την απόσταση Mahalanobis από το χρωματικό διάνυσμα c στο διάνυσμα μέσης τιμής μ_s , δοθέντος του πίνακα συνδιακύμανσης Σ_s , δηλαδή

$$\lambda_s(c) = (c - \mu_s)^T \Sigma_s^{-1} (c - \mu_s)$$

3.1.2 Μίγμα Gaussian Κατανομών

Ένα μοντέλο μίγματος Gaussian κατανομών (Gaussian mixture model) έχει συνάρτηση πυκνότητας πιθανότητας

$$P(c|skin) = \sum_{i=1}^k \pi_i P_i(c|skin) \quad (3.2)$$

όπου k είναι ο αριθμός των κατανομών του μίγματος, π_i είναι τα βάρη με περιορισμό $\sum_{i=1}^k \pi_i = 1$ και $P_i(c|skin)$ είναι οι συναρτήσεις πυκνότητας πιθανότητας των Gaussians, καθεμιά με διαφορετική μέση τιμή και πίνακα συνδιακύμανσης. Η εκπαίδευση του μοντέλου γίνεται με την μέθοδο Expectation Maximization (EM), η οποία απαιτεί την εκ των προτέρων γνώση του αριθμού k των Gaussians. Οι λεπτομέρειες της εκπαίδευσης του μοντέλου μίξης Gaussian με τον αλγόριθμο EM αναλύονται στο [JoRe99]. Η ταξινόμηση με το μοντέλο αυτό γίνεται με σύγκριση της πιθανότητας $P(c|skin)$ με ένα επιλεγμένο από το χρήστη κατώφλι (threshold). Ακόμη τονίζεται ότι η επιλογή του αριθμού k των κατανομών k είναι πολύ σημαντική, καθώς πρέπει να γίνει η καλύτερη δυνατή εκαπίδευση των δεδομένων εισόδου, αποφεύγοντας τις αλληλοκαλύψεις. Οι τιμές του k κυμαίνονται στο διάστημα [2, 16]. Ενδεικτική τιμή είναι $k = 8$, με την οποία επιτυγχάνεται καλή ακρίβεια χωρίς μεγάλο κόστος υπολογισμού.

3.2 Χρωματικός Χώρος Δέρματος

Μια μέθοδος ανίχνευσης δέρματος είναι και με χρήση της τιμής χρώματος που έχουν τα pixels μιας εικόνας. Παρακάτω γίνεται μια συνοπτική παρουσίαση των διαφόρων χρωματικών χώρων, ώστε να επιλεγεί ο καταλληλότερος.

3.2.1 Παρουσίαση Χρωματικών Χώρων

Στις επόμενες παραγράφους παρουσιάζονται οι τρεις σημαντικότερες κατηγορίες χρωματικών χώρων.

RGB και Κανονικοποιημένος RGB Στον χρωματικό χώρο RGB, το χρώμα περιγράφεται από το συνδυασμό των χρωμάτων κόκκινου, πράσινου και μπλε. Είναι ο πιο διαδεδομένος χρωματικός χώρος για την επεξεργασία και αποθήκευση ψηφιακών αρχείων εικόνας. Ωστόσο, η υψήλη συσχέτιση μεταξύ των καναλιών, η έλλειψη αντιληπτικής ομοιομορφίας και η ανάμιξη των δεδομένων χρωματισμού και φωτεινότητας είναι οι λόγοι που καθιστούν τον χώρο RGB ακατάλληλο για ανάλυση χρώματος και αλγορίθμους αναγνώρισης βασισμένους στο χρώμα.

Ο κανονικοποιημένος χρωματικός χώρος RGB υπολογίζεται πολύ εύκολα από τον RGB με τις σχέσεις

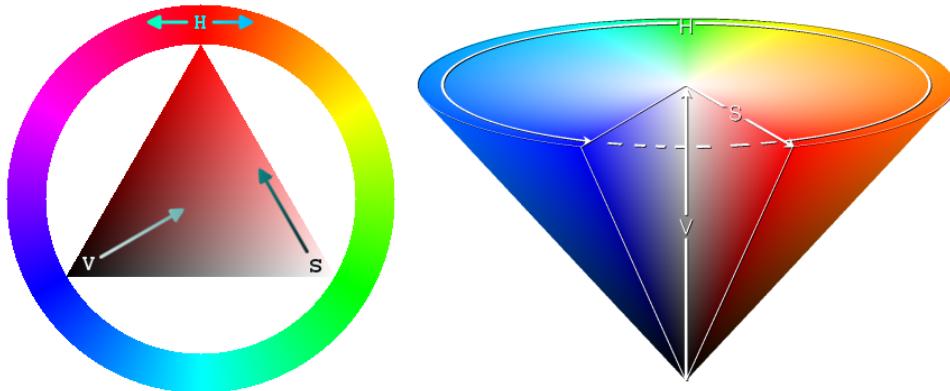
$$r = \frac{R}{R+G+B} , \quad g = \frac{G}{R+G+B} , \quad b = \frac{B}{R+G+B}$$

Δεδομένου ότι ισχύει

$$r + g + b = 1$$

η τρίτη μεταβλητή δεν περιέχει καμιά σημαντική πληροφορία, με αποτέλεσμα να μειώνονται οι διαστάσεις του χρωματικού χώρου. Ένα ακόμα πλεονέκτημα του χώρου αυτού είναι η ανεξαρτησία του στον προσανατολισμό της επιφάνειας ως προς την πηγή φωτός, ενώ είναι ανεξάρτητο και από τον περιβάλλοντα φωτισμό. Αυτά τα χαρακτηριστικά έχουν κάνει το συγκεκριμένο χρωματικό χώρο καταλληλότερο συγκριτικά με τον RGB.

HSI, HSV, HSL Τα αρχικά των χώρων αυτών σημαίνουν Hue, Saturation, Intensity (Value, Lightness), δηλαδή Απόχρωση, Κορεσμός, Ένταση (Τιμή, Φωτεινότητα). Είναι χρωματικοί χώροι που χρησιμοποιούνται όταν υπάρχει ανάγκη για έκφραση των ιδιοτήτων του χρώματος αριθμητικά. Η απόχρωση (H) ορίζει το κυρίαρχο χρώμα (π.χ. κόκκινο, κίτρινο, πράσινο, μωβ) σε μια περιοχή και ο κορεσμός (S) εκφράζει την πολυχρωμία μιας περιοχής αναλογικά με την φωτεινότητά της. Οι παράμετροι της έντασης (V), της τιμής (V) και της φωτεινότητας (L) ουσιαστικά εκφράζουν την φωτεινότητα του χρώματος. Πρακτικά, ο κορεσμός (S) εκφράζει την χρωματική καθαρότητα, που είναι ο βαθμός μίξης ενός καθαρού χρώματος με το άσπρο και αντίστοιχα η τιμή (V) εκφράζει το βαθμό μίξης ενός καθαρού χρώματος με το μαύρο. Το σύστημα συντεταγμένων του χρωματικού χώρου HSV είναι κυκλικό και τα χρώματα βρίσκονται μέσα σε έναν εξάγωνο κώνο, όπως φαίνεται και στο Σχήμα 3.1. Η τιμή του H κυμαίνεται μεταξύ 0° και 360° , ενώ οι τιμές των S και V μεταξύ 0 και 1.



Σχήμα 3.1: Χρωματικός χώρος HSV

Ο χρωματικός χώρος HSV περιγράφει ουσιαστικά τον τρόπο με τον οποίο βλέπει το ανθρώπινο μάτι. Το μεγάλο του πλεονέκτημα είναι ότι η ταυτότητα του χρώματος ουσιαστικά περιγράφεται από μια παράμετρο (Hue), ενώ οι άλλες δύο παράμετροι επηρεάζονται από τη φωτεινότητα, τη σκίαση και την τραχύτητα της επιφάνειας του αντικειμένου.

Η μετατροπή από κανονικοποιημένο RGB σε HSV, θεωρώντας ότι $r, g, b \in [0, 1]$, $\max = \max(r, g, b)$ και $\min = \min(r, g, b)$, δίνεται από τις εξισώσεις:

$$h = \begin{cases} 0 & , \text{αν } \max = \min \\ \left(60^\circ \times \frac{g - b}{\max - \min} + 360^\circ \right) \mod 360^\circ & , \text{αν } \max = r \\ \left(60^\circ \times \frac{b - r}{\max - \min} + 120^\circ \right) & , \text{αν } \max = g \\ \left(60^\circ \times \frac{r - g}{\max - \min} + 240^\circ \right) & , \text{αν } \max = b \end{cases}$$

$$s = \begin{cases} 0 & , \text{αν } \max = 0 \\ 1 - \frac{\min}{\max} & , \text{αλλιώς} \end{cases}$$

$$v = \max$$

Διαφορετικά η μετατροπή από τον χρωματικό χώρο RGB στον HSV γίνεται με τις σχέσεις

$$H = \arccos \frac{\frac{1}{2} [(R - G) + (R - B)]}{\sqrt{(R - G)^2 + (R - G)(G - B)}}$$

$$S = 1 - 3 \frac{R \cdot G \cdot B}{R + G + B}$$

$$V = \frac{1}{3} (R + G + B)$$

YCrCb Ο χρωματικός χώρος YCrCb είναι ένα καδικοποιημένο μη γραμμικό RGB σύμα, το οποίο χρησιμοποιείται συχνά για συμπίεση εικόνων. Το χρώμα αναπαρίσταται από τη μεταβλητή Y ως ένα άθροισμα με βάρη των τιμών RGB. Συγκεκριμένα ισχύει

$$Y = 0.299R + 0.587G + 0.114B$$

$$Cr = R - Y$$

$$Cb = B - Y$$

Ο απλός τρόπος μετασχηματισμού και η σαφής διάκριση μεταξύ φωτεινότητας και χρωματισμού καθιστούν τον χώρο YCrCb κατάλληλο για μοντελοποίηση του χρώματος του δέρματος.

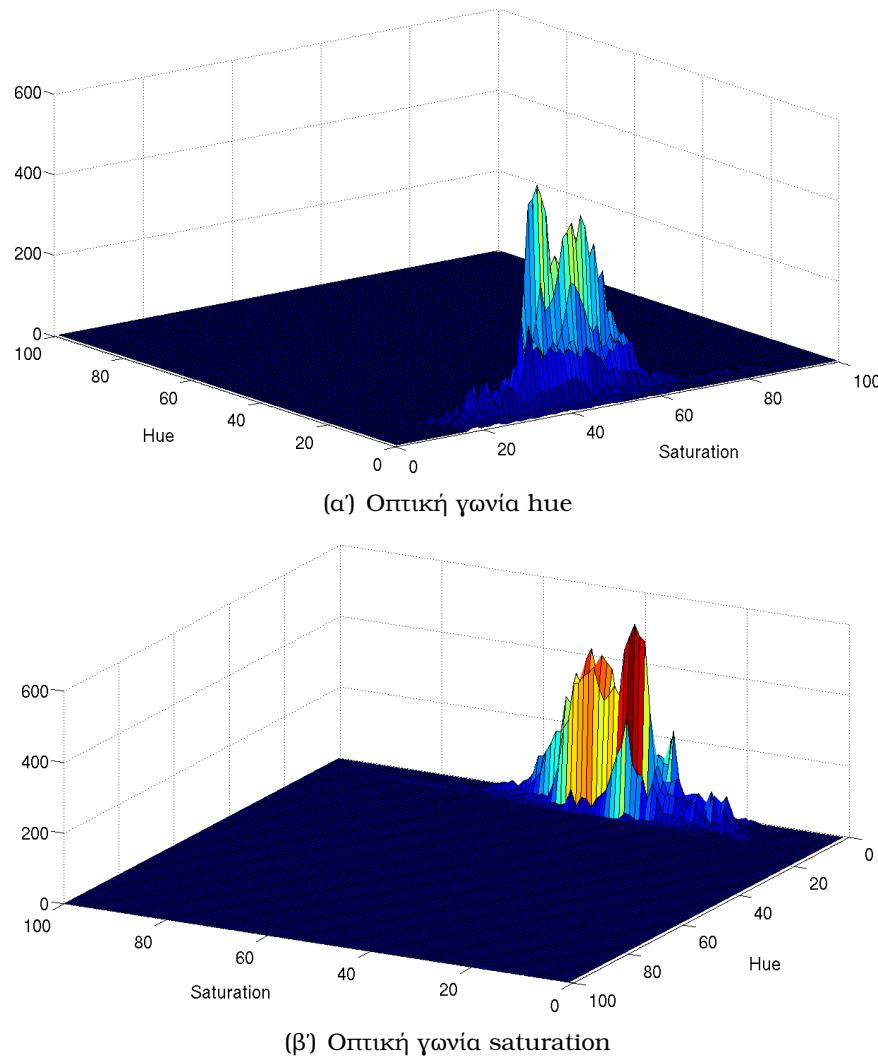
3.2.2 Επιλογή Χρωματικού Χώρου για Ανίχνευση Δέρματος

Αν και κάθε χρωματικός χώρος έχει τα πλεονεκτήματα και τα μειονεκτήματά του, τελικά για τη συγκεκριμένη διπλωματική εργασία επιλέχθηκε ο χώρος HSV. Θεωρήθηκε

μεγάλο πλεονέκτημα το γεγονός ότι η ταυτότητα του χρώματος ουσιαστικά αναπαρίσταται από ένα αριθμό, την τιμή της απόχρωσης (hue) και παραβλέφθηκε το πιθανό υπολογιστικό κόστος του μετασχηματισμού από τον χώρο RGB.

Για την εκπαίδευση του συστήματος, δόθηκαν κάποιες εικόνες προσώπων και οι συντεταγμένες των σημείων δέρματος, οι οποίες ορίστηκαν με το χέρι. Οι εικόνες εκπαίδευσης παρουσίαζαν όσο το δυνατόν μεγαλύτερη ποικιλία σε φωτισμό (φυσικός, τεχνητός), σε περιβάλλον (εσωτερικό, εξωτερικό), σε χρονική περίοδο (ήλιος, συννεφιά κτλ.) αλλά και ανθρώπινο δέρμα (χλωμό, σκούρο). Ακόμη να σημειωθεί ότι συμπεριλήφθηκαν μόνο φωτογραφίες λευκών Ευρωπαϊκών προσώπων και όχι άλλα είδη δέρματος.

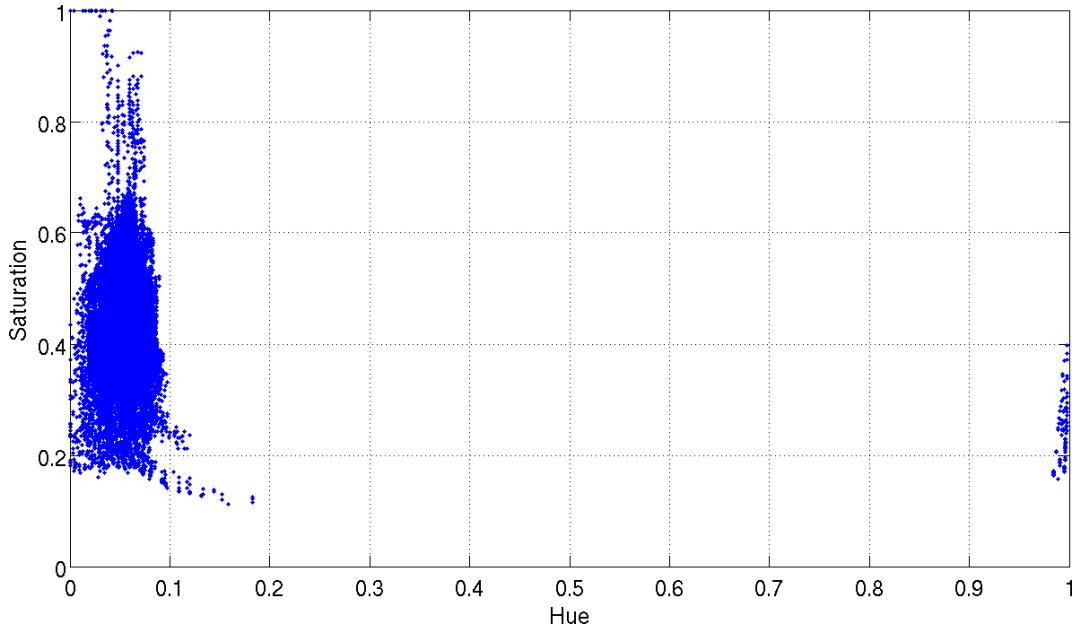
Στο Σχήμα 3.2 φαίνεται η διασπορά των τιμών των pixels δέρματος των εικόνων εκπαίδευσης στις παραμέτρους hue και saturation του χώρου HSV. Τα δύο διαγμάτα είναι σε τρεις διαστάσεις και απεικονίζουν αντίθετες όψεις.



Σχήμα 3.2: Τρισδιάστατη διασπορά χρώματος δέρματος στον χρωματικό χώρο HSV

Στο Σχήμα 3.3 φαίνεται η δισδιάστατη διασπορά του χρώματος δέρματος. Ουσιαστικά τα διαγράμματα αυτά βοηθούν στην επιλογή κατάλληλων ορίων μέσα στα οποία ένα pixel μιας εικόνας θεωρείται δέρμα. Έστερα από αρκετό πειρασματισμό ενδεικτικές συνθήκες για να είναι ένα χρώμα ανθρώπινο δέρμα είναι αν ισχύει

$$0 \leq H \leq 0.11 \text{ και } 0.2 \leq S \leq 0.7 \quad (3.3)$$



Σχήμα 3.3: Δισδιάστατη διασπορά χρώματος δέρματος στον χρωματικό χώρο HSV

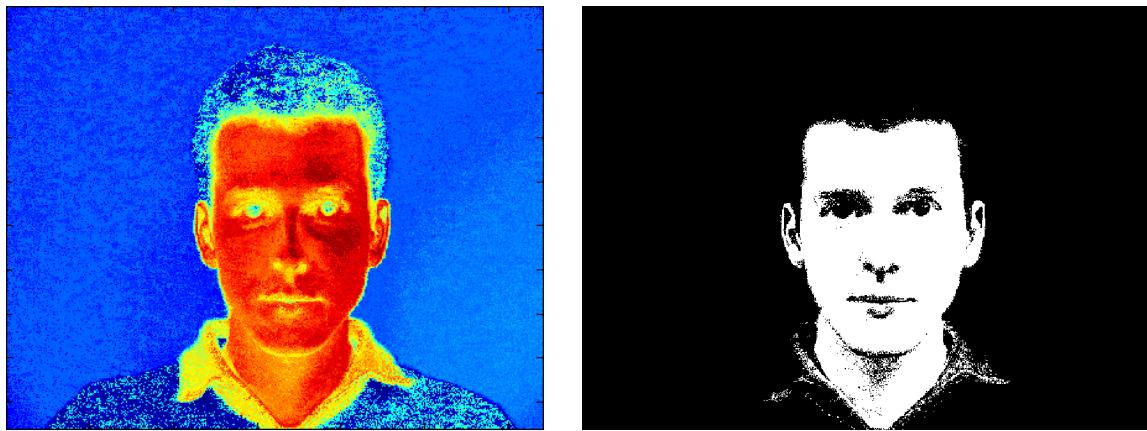
3.3 Πειραματικά Αποτελέσματα

Αρχικά οι δυο μέθοδοι που παρουσιάστηκαν, οι οποίες υλοποιήθηκαν σε MATLAB, εκτελούνται για όλες τις εικόνες της βάσης δεδομένων IMM [NLSS04], η οποία χρησιμοποιείται μετά για την εκπαίδευση του συστήματος αναγνώρισης προσώπου. Από τις δοκιμές αυτές βγαίνει το συμπέρασμα ότι ο πιθανοτικός αλγόριθμος έχει καλύτερα αποτελέσματα από την ανίχνευση δέρματος με το χρωματικό μοντέλο HSV. Συγκεκριμένα η πιθανοτική ανίχνευση βρίσκει σωστά όλα τα pixels δέρματος και μόνο αυτά, ενώ η ανίχνευση με βάση το χρώμα βρίσκει κάποια από τα pixels του προσώπου αλλά και κάποια που σίγουρα δεν αποτελούν δέρμα. Ένα παράδειγμα φαίνεται στο Σχήμα 3.4, όπου ο πιθανοτικός αλγόριθμος εκτελείται με τιμή κατωφλίου –1. Ωστόσο πρέπει να τονιστεί ότι ο φωτισμός στις εικόνες της βάσης δεδομένων είναι συγκεκριμένος, εννιαίος για όλες και ακατάλληλος για τη σωστή λειτουργία της μεθόδου HSV. Ακόμη το παρασκήνιο όλων των εικόνων είναι μονόχρωμο (πράσινο σκούρο), γεγονός που καθιστά την ανίχνευση του δέρματος πολύ πιο απλή.



(α) Αρχική εικόνα

(β) Ανίχνευση δέρματος με βάση το χρωματικό χώρο HSV



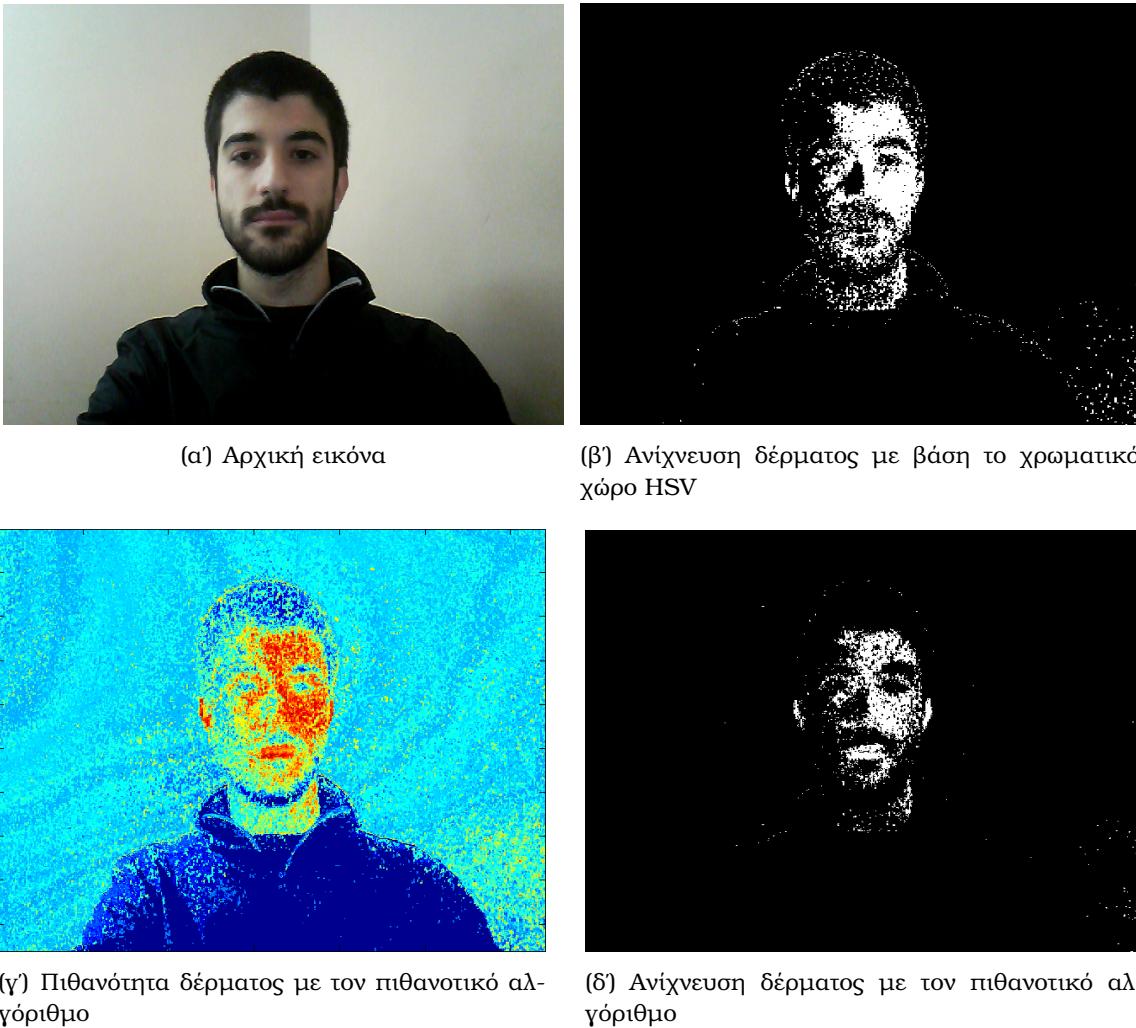
(γ) Πιθανότητα δέρματος με τον πιθανοτικό αλγόριθμο

(δ) Ανίχνευση δέρματος με τον πιθανοτικό αλγόριθμο

Σχήμα 3.4: Παράδειγμα ανίχνευσης δέρματος σε εικόνα της βάσης δεδομένων IMM

Δοκιμάζοντας τους δυο αλγόριθμους σε μεγάλη ποικιλία εικόνων, παρατηρήθηκε ότι ο καθένας έχει τα δικά του πλεονεκτήματα ανάλογα με τις συνθήκες φωτισμού κάτω από τις οποίες δημιουργήθηκε η εικόνα. Ο αλγόριθμος με βάση το χρωματικό χώρο HSV λειτουργεί καλύτερα όταν υπάρχει άφθονος φωτισμός, κυρίως φυσικός, ενώ ο πιθανοτικός αλγόριθμος λειτουργεί καλύτερα όταν οι συνθήκες φωτισμού δεν είναι καλές. Ένα παράδειγμα στο οποίο το σύστημα HSV φαίνεται να λειτουργεί καλύτερα από τον πιθανοτικό αλγόριθμο φαίνεται στο Σχήμα 3.5. Η αρχική εικόνα έχει τραβηγθεί από τη webcam του υπολογιστή.

Βέβαια ένα μεγάλο πλεονέκτημα που έχει ο πιθανοτικός αλγόριθμος είναι η δυνατότητα επιλογής της ευαισθησίας του που δίνεται στο χρήστη μέσω της παραμέτρου κατωφλίου. Οι τιμές της παραμέτρου αυτής κυμαίνονται στο διάστημα $[-15, 4]$, με ενδεικτική τιμή το -1 . Στο Σχήμα 3.6 φαίνεται η ανίχνευση δέρματος μιας εικόνας για διάφορες τιμές του κατωφλίου D_{thres} .



Σχήμα 3.5: Παράδειγμα ανίχνευσης δέρματος σε εικόνα από την webcam

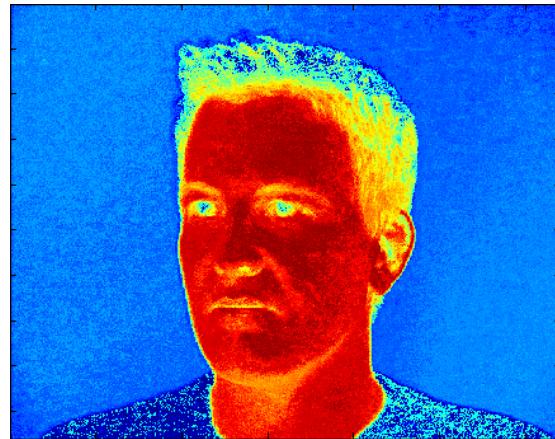
Τέλος τονίζεται ότι οι δυο αλγόριθμοι υλοποιήθηκαν ώστε να εκτελούνται σε πραγματικό χρόνο με είσοδο ακολουθία βίντεο που δίνεται από την webcam. Όπως αναμενόταν ο χρόνος εκτέλεσής τους ανά frame είναι ιδιαίτερα χαμηλός και φαίνεται στον Πίνακα 3.1. Ο μεγαλύτερος χρόνος της μεθόδου με βάση το χρώμα του δέρματος οφείλεται στην υπολογιστική πολυπλοκότητα της μετατροπής από RGB σε HSV.

<i>Μέθοδος</i>	<i>Χρόνος (sec)</i>
Αλγόριθμος με HSV	0.127
Πιθανοτικός αλγόριθμος	0.037

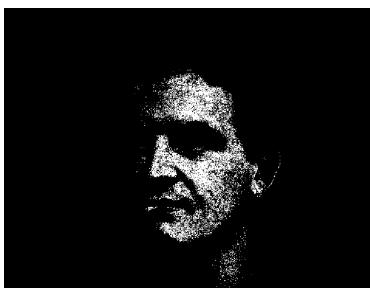
Πίνακας 3.1: Χρόνοι εκτέλεσης αλγορίθμων ανίχνευσης δέρματος ανά εικόνα



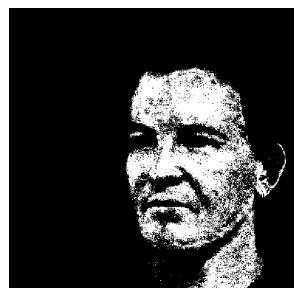
(α') Αρχική εικόνα



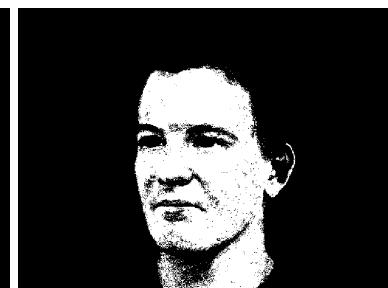
(β') Πιθανότητα δέρματος



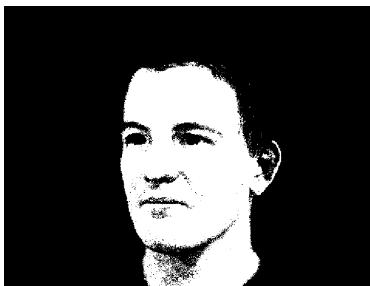
(γ') $D_{thres} = 4$



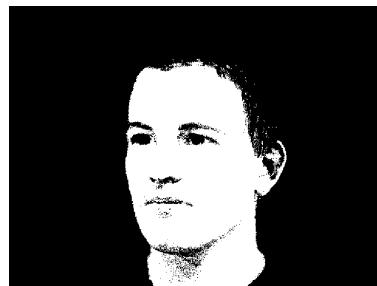
(δ') $D_{thres} = 3$



(ε') $D_{thres} = 2$



(f') $D_{thres} = 1$



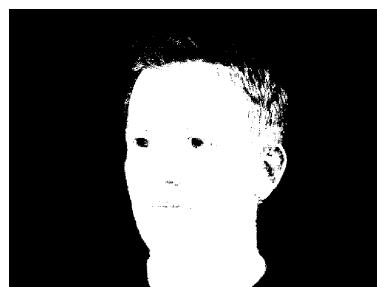
(ζ') $D_{thres} = 0$



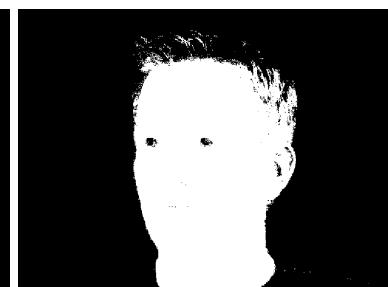
(η') $D_{thres} = -1$



(θ') $D_{thres} = -2$



(ι') $D_{thres} = -3$



(ια') $D_{thres} = -4$

Σχήμα 3.6: Ανίχνευση δέρματος με πιθανοτικό αλγόριθμο για διάφορες τιμές του κατωφλίου D_{thres}

Κεφάλαιο 4

Ανίχνευση Προσώπου με Χρήση Haar Χαρακτηριστικών και AdaBoost Αλγορίθμου

Η ανίχνευση προσώπου είναι πολύ σημαντική για τα στάδια της αναγνώρισης. Για το λόγο αυτό πρέπει να είναι σταθερή και αποτελεσματική. Επιλέχθηκε ο ανιχνευτής προσώπου που παρουσιάστηκε από τους Viola & Jones το 2001 [ViJo01a, ViJo01b], ο οποίος χρησιμοποιεί Haar χαρακτηριστικά και τον AdaBoost αλγόριθμο εκαπίδευσης. Η μέθοδος ουσιαστικά σαρώνει την εικόνα εισόδου σε όλες τις πιθανές θέσεις και κλίμακες και στη συνέχει ταξινομεί τα παράθυρα ως πρόσωπα ή μη-πρόσωπα.

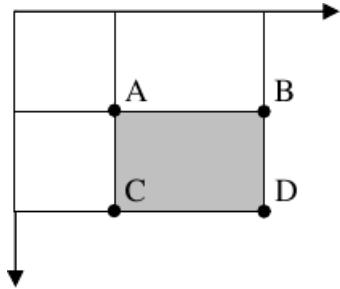
4.1 Ολοκληρωτική Εικόνα

Ο στόχος είναι η δημιουργία ενός ανιχνευτή προσώπου που θα είναι γρήγορος και ακριβής. Μεγάλο ρόλο στην ταχύτητα του αλγορίθμου παίζει η χρησιμοποίηση της ολοκληρωτικής εικόνας (integral image) που ορίζεται ως

$$I_{int}(x, y) = \sum_{x'=0}^x \sum_{y'=0}^y I(x', y') \quad (4.1)$$

Άρα η τιμή της ολοκληρωτικής εικόνας σε ένα σημείο (x, y) είναι το άθροισμα των τιμών όλων των σημείων πάνω και αριστερά από το σημείο αυτό, συμπεριλαμβανομένου του σημείου (x, y) . Η χρήση της ολοκληρωτικής εικόνας διευκολύνει τις υπολογιστικές διαδικασίες. Συγκεκριμένα, το άθροισμα των σημείων ενός ορθογώνιου της ολοκληρωτικής εικόνας μπορεί εύκολα να υπολογιστεί από τα τέσσερα ακρειανά του σημεία. Για το παράδειγμα του Σχήματος 4.1, όπου το ορθογώνιο ορίζεται από τα σημεία A , B , C και D , ισχύει

$$S_{ABCD} = I_{int}(D) - [I_{int}(B) + I_{int}(C)] + I_{int}(A)$$

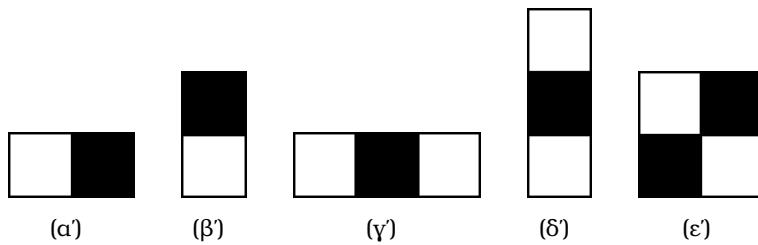


Σχήμα 4.1: Ολοκληρωτική Εικόνα

4.2 Χαρακτηριστικά Haar

Αν ο ανιχνευτής προσώπου σάρωνε κάθε ένα pixel ξεχωριστό, τότε το υπολογιστικό κόστος θα ήταν πολύ μεγάλο. Αντίθετα, η χρησιμοποίηση ενός συνόλου χαρακτηριστικών (feature set) των τιμών των pixels κάνει τον αλγόριθμο πολύ ταχύτερο. Έτσι, το σύνολο χαρακτηριστικών που χρησιμοποιείται χωρίζει την εικόνα σε ορθογώνιες περιοχές και υπολογίζει το άθροισμα των pixels των τετραγώνων μέσω της ολοκληρωτικής εικόνας (integral image). Μάλιστα αυτή η τακτική είναι ιδιαίτερα συμφέρουσα αν συνυπολογιστεί το γεγονός ότι τα χαρακτηριστικά περιέχουν μεγαλύτερη πληροφορία για την περιοχή σε σύγκριση με τα μεμονωμένα pixels.

Τα χαρακτηριστικά που πρότειναν οι Viola και Jones είναι ιδιαίτερα απλά ορθογώνια σχήματα, όπως φαίνεται και στο Σχήμα 4.2. Η τιμή του κάθε χαρακτηριστικού ορίζεται ως η διαφορά του αθροίσματος των pixels της λευκής περιοχής από αυτά της μάυρης περιοχής. Φυσικά η θέση και το μέγεθος των χαρακτηριστικών κάθε φορά διαφέρουν και εξαρτώνται από το παράθυρο που σαρώνει την εικόνα εισόδου.



Σχήμα 4.2: Viola-Jones Χαρακτηριστικά που μοιάζουν με Haar

Κάθε χαρακτηριστικό συνοδεύεται από κάποιες παραμέτρους. Για παράδειγμα το χαρακτηριστικό της εικόνας 4.2(γ') έχει τέσσερις παραμέτρους: τη θέση μέσα στο παράθυρο σάρωσης, το μέγεθος της άσπρης (ή θετικής) περιοχής, το μέγεθος της μάυρης (ή αρνητικής) περιοχής και το ύψος.

Έχει αποδειχθεί ότι μια καλή ανάλυση για τον ανιχνευτή είναι 24×24 pixels. Δίνοτας όλα τα πιθανά μεγέθη και όλες τις πιθανές θέσεις στα χαρακτηριστικά του Σχήματος 4.2, προκύπτει ένα σύνολο 160000 περίπου διαφορετικών χαρακτηριστικών. Τα χαρακτηριστικά αυτά, αν και απλοϊκά στη μορφή, επιτυγχάνουν εξαιρετικά

αποτελέσματα με πολύ μικρό υπολογιστικό κόστος.

4.3 Εξαγωγή Χαρακτηριστικών με τον Αλγόριθμο AdaBoost

Ο AdaBoost αλγόριθμος εκτελεί έναν αδύναμο ταξινομητή (weak classifier) πολλαπλές φορές και για κάθε εκτέλεση ανανεώνεται μια κατανομή από βάρη, η οποία υποδεικνύει το πόσο σημαντικά είναι τα δείγματα της βάσης δεδομένων για τη διαδικασία της ταξινόμησης. Το όνομά του είναι συντομογραφία του Adaptive Boosting.

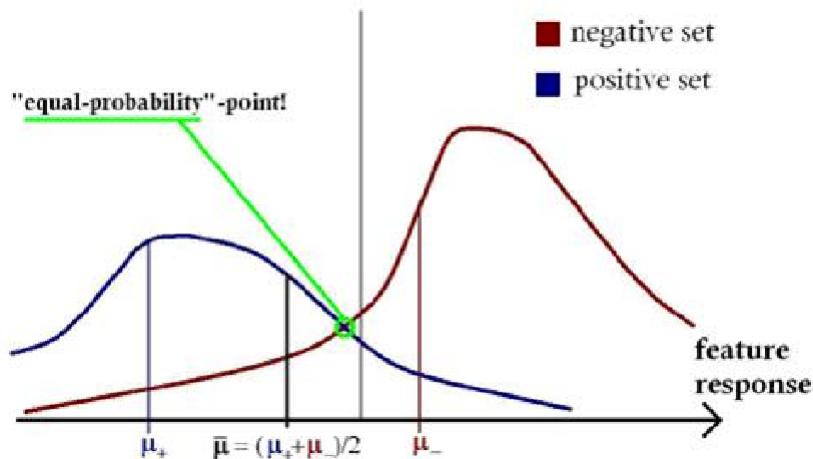
Αρχικά λοιπόν οι αδύναμοι ταξινομητές (weak classifiers) ορίζονται ως

$$h_i(x, f_i, p_i, \theta_i) = \begin{cases} 1, & \text{αν } p_i f_i(x) < p_i \theta_i \\ 0, & \text{αλλιώς} \end{cases} \quad (4.2)$$

όπου $f_i(x)$ είναι το διάνυσμα ενός χαρακτηριστικού υπολογισμένο στο παράθυρο x , p_i είναι η πιθανότητα και θ_i είναι το κατώφλι. Ο υπολογισμός των p_i και θ_i μπορεί να γίνει με δυο τρόπους. Και οι δυο βασίζονται στην εκτίμηση δυο κατανομών πιθανότητας από την εφαρμογή των τιμών των χαρακτηριστικών στα θετικά δείγματα (εικόνες με πρόσωπα) και στα αρνητικά δείγματα (εικόνες με μη-πρόσωπα). Έτσι, η τιμή κατωφλίου μπορεί να υπολογιστεί είτε με τη μέση τιμή των μέσων των δυο κατανομών είτε με το σημείο crossover, που αντιστοιχεί σε τιμή f_i τέτοια ώστε

$$p(f_i|non-face) = p(f_i|face) \quad (4.3)$$

Τελικά χρησιμοποιήθηκε το σημείο crossover και οι δυο κατανομές πιθανότητας φαίνονται στο Σχήμα 4.3.

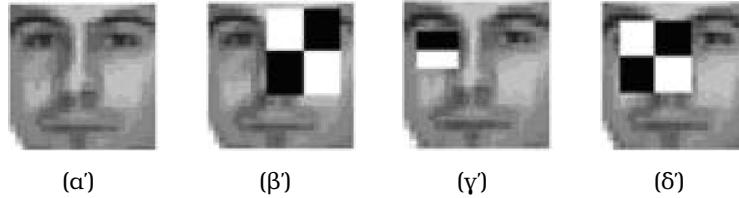


Σχήμα 4.3: Κατανομές πιθανότητας θετικών και αρνητικών δειγμάτων προσώπων

Επομένως γίνεται σαφές ότι ένα σύνολο χαρακτηριστικών ορίζει ένα αντίστοιχο σύνολο αδύναμων ταξινομητών. Το επόμενο βήμα είναι η χρήση του AdaBoost αλγορίθμου ώστε από το σύνολο των αδύναμων ταξινομητών, που πρακτικά είναι ενα

σύνολο δυαδικών τιμών, να σχηματιστεί ένας δυνατός ταξινομητής (strong classifier). Μάλιστα η διαδικασία του συνδυασμού αδύναμων ταξινομητών για τη δημιουργία ενός δυνατού ταξινομητή χρησιμοποιείται από την ανθρώπινη λογική κατά τη λήψη αποφάσεων. Για παράδειγμα η τελική απόφαση για την αγορά ενός προϊόντος βασίζεται στις απαντήσεις επιμέρους κριτηρίων, όπως η τιμή, η εγγύηση και η ποιότητά του.

Βασική παράμετρος στον AdaBoost αλγόριθμο είναι η χρήση κατανομών για βάρη. Αφού γίνει αρχικοποίηση της κατανομής των βαρών, σε κάθε επαναληπτικό βήμα η κατανομή ανανεώνεται αν η ταξινόμηση από τους αδύναμους ταξινομητές είναι ορθή. Το αποτέλεσμα είναι ότι στους αδύναμους ταξινομητές που καταφέρνουν να ταξινομήσουν δύσκολες εικόνες, δίνεται μεγαλύτερη τιμή βάρους στον τελικό δυνατό ταξινομητή. Στο Σχήμα 4.4 φαίνονται κάποια από τα ορθογώνια χαρακτηριστικά που επιλέχθηκαν από τον αλγόριθμο AdaBoost. Τα βήματα της μεθόδου φαίνονται στον αλγόριθμο 1.



Σχήμα 4.4: Χαρακτηριστικά επιλεγμένα από τον αλγόριθμο AdaBoost

Αλγόριθμος 1 Αλγόριθμος AdaBoost

Input: Εικόνες εκπαίδευσης (x_1, \dots, x_n) και οι αντίστοιχες ετικέτες (y_1, \dots, y_n) , όπου $y_i \in \{0, 1\}$ για αρνητικά και θετικά δείγματα αντίστοιχα. Ο αριθμός των αρνητικών δειγμάτων είναι m και αυτός των θετικών δειγμάτων $l = n - m$

- 1: Αρχικοποίηση των n βαρών ως $(2m)^{-1}$ για $y_i = 0$ και $(2l)^{-1}$ για $y_i = 1$
 - 2: **for** $t = 1, \dots, T$ **do**
 - 3: Κανονικοποίηση των βαρών $w_{t,i} \leftarrow \frac{w_{t,i}}{\sum_{j=1}^n w_{t,j}}$
 - 4: Επιλογή του καλύτερου αδύναμου ταξινομητή ως προς το σφάλμα με βάρη $\epsilon_t = \min_{f,p,\theta} \sum_i w_i |h(x_i, f, p, \theta) - y_i|$
 - 5: Ορισμός της $h_t(x) = h(x, f_t, p_t, \theta_t)$, όπου f_t , p_t και θ_t οι τιμές των μεταβλητών που ελαχιστοποιούν το ϵ_t
 - 6: Ανανέωση των βαρών $w_{t+1,i} = w_{t,i} \beta_t^{1-e_i}$ όπου $e_i = 0$ αν το δείγμα x_i ταξινομείται σωστά, $e_i = 1$ αν το δείγμα x_i ταξινομείται λανθασμένα και $\beta_t = \frac{\epsilon_t}{1 - \epsilon_t}$
 - 7: **end for**
 - 8: Ο δυνατός ταξινομητής με $\alpha_t = \log \frac{1}{\beta_t}$ ορίζεται ως

$$h_x = \begin{cases} 1, & \text{αν } \sum_{t=1}^T \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^T \alpha_t \\ 0, & \text{αλλιώς} \end{cases}$$
-

4.4 Ταξινομητής Cascade

Όπως έχει αναφερθεί, η βασική αρχή της μεθόδου ανίσχνευσης προσώπου των Viola-Jones είναι η πολλαπλή σάρωση της εικόνας εισόδου με τον ανιχνευτή, κάθε φορά με διαφορετικό μέγεθος. Ακόμα και αν μια εικόνα περιλαμβάνει ένα ή περισσότερα πρόσωπα, ο μεγαλύτερος αριθμός των παραθύρων υπολογισμού θα έδινε αρνητική απάντηση, δηλαδή παρουσία μη-πρόσωπων. Για το λόγο αυτό ακολουθείται μια διαφορετική λογική, κατά την οποία *αντί o αλγόριθμος να βρίσκει πρόσωπα, μπορεί να απορρίπτει μη-πρόσωπα.* Η λογική αυτή βασίζεται στο γεγονός ότι η διαδικασία της απόρριψης μη-πρόσωπων είναι ταχύτερη από τη διαδικασία εύρεσης προσώπων. Επομένως, γίνεται ξεκάθαρο ότι η χρήση ενός δυνατού ταξινομητή είναι αναποτελεσματική, αφού η χρονική πολυπλοκότητα παραμένει ίδια οποιαδήποτε και αν είναι η είσοδος. Για το λόγο αυτό οι Viola-Jones εισήγαγαν την έννοια του cascade ταξινομητή (ακολουθιακός ταξινομητής).

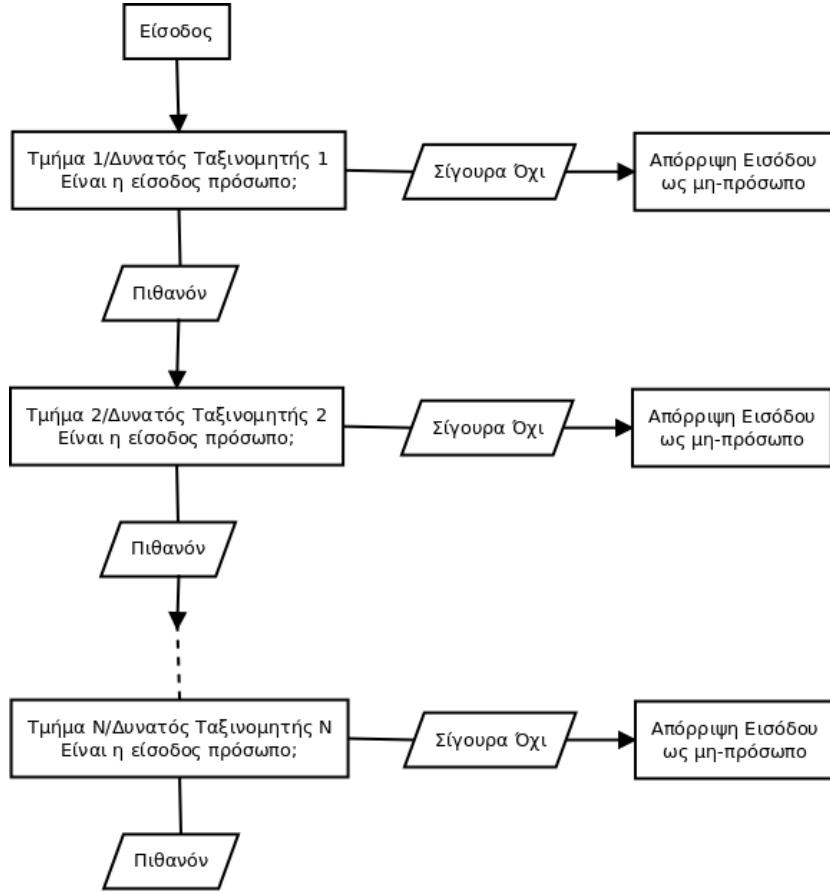
Ο ταξινομητής cascade αποτελείται από σειριακά τμήματα, καθένα από τα οποία είναι ένας δυνατός ταξινομητής. Το κάθε τμήμα καθορίζει κατά πόσο ένα παράθυρο εισόδου είναι σίγουρα μη-πρόσωπο ή πιθανό πρόσωπο. Αν το παράθυρο ταξινομηθεί ως μη-πρόσωπο από κάποιο τμήμα (δυνατό ταξινομητή), τότε απορρίπτεται. Αν όμως ένα παράθυρο ταξινομηθεί ως πιθανό πρόσωπο, τότε συνεχίζει στο επόμενο τμήμα του ταξινομητή cascade. Προφανώς, όσο περισσότερα τμήματα καταφέρει να περάσει ένα πράθυρο χωρίς να απορριφθεί, τόσο μεγαλύτερη είναι η πιθανότητα να περιλαμβάνει τελικά πρόσωπο. Η διαδικασία αυτή αναπαρίσταται στο Σχήμα 4.5.

Ένα σημαντικό πλεονέκτημα του ταξινομητή cascade είναι η μείωση των false-positive εικόνων (λανθασμένα θετικών). Με τη χρήση ενός δυνατού ταξινομητή, η μείωση αυτή επιτυγχάνεται με αποδοχή των false-negative εικόνων ώστε να διορθωθεί η αναλογία με τις false-positive. Ωστόσο αυτή η τακτική δεν είναι σωστή. Με τη χρήση όμως ενός ταξινομητή πολλών σταδίων, οι false-positive εικόνες δεν αποτελούν πρόβλημα καθώς αναμένεται να απορριφθούν σε κάποιο στάδιο και να έχουν μείνει πολύ λίγες στο τελικό στάδιο.

Οι Viola-Jones ονομάζουν τον cascade ταξινομητή και attentional cascade, δηλαδή ταξινομητή κατευθυνόμενης προσοχής. Αυτό, με βάση και την ανάλυση που προηγήθηκε, σημαίνει ότι δίνεται μεγαλύτερη υπολογιστική ισχύς σε περιοχές της εικόνας που υπάρχει μεγάλη πιθανότητα να περιέχουν ένα ή περισσότερα πρόσωπα. Στον αλγόριθμο 2 φαίνονται τα βήματα του cascade ταξινομητή. Ο αλγόριθμος απαιτεί σαν είσοδο τον ελάχιστο βαθμό αποδεκτής ανίχνευσης minimum acceptable detection rate d και τον μέγιστο βαθμό false-positive εικόνων f . Ακόμη τονίζεται ότι κατά την εκπαίδευση ενός τμήματος (δυνατού ταξινομητή) n , τα αρνητικά δείγματα πρέπει να είναι false-negative δείγματα που προέρχονται από το τμήμα $n - 1$.

4.5 Πειραματικά Αποτελέσματα

Για την εκπαίδευση του ανιχνευτή προσώπου χρησιμοποιήθηκε η βάση δεδομένων LFW [HRB07]. Αυτή είναι μια βάση δεδομένων του University of Massachusetts που



Σχήμα 4.5: Ταξινομητής Cascade

περιέχει περισσότερες από 13000 εικόνες προσώπου που έχουν βρεθεί στο διαδίκτυο. Όλες οι εικόνες της βάσης έχουν προέλθει από ανίχνευση με τη μέθοδο Viola-Jones.

Η ανίχνευση προσώπου υλοποιήθηκε σε μορφή MEX αρχείου στο MATLAB για να επιτευχθεί μεγαλύτερη ταχύτητα. Το πρόγραμμα δέχεται σαν είσοδο μια εικόνα οποιουδήποτε μεγέθους και επιστρέφει έναν πίνακα fd . Ο πίνακας αυτός έχει διαστάσεις $N_f \times 3$, όπου N_f είναι ο αριθμός των προσώπων που ανιχνεύτηκαν στην εικόνα και οι 3 αριθμοί της κάθε γραμμής προσδιορίζουν τις συντεταγμένες του κέντρου του προσώπου που βρέθηκε (x, y) και το πλάτος του αντίστοιχα. Αν ανιχνευτούν πάνω από ένα πρόσωπο, τότε ο πίνακας fd ταξινομείται σε φθίνουσα σειρά με βάση το πλάτος του προσώπου, δηλαδή ως προς την τρίτη στήλη. Προφανώς αν δεν ανιχνευτεί κανένα πρόσωπο, ο πίνακας που επιστρέφεται είναι κενός. Τέλος, με βάση τις κλίμακες που εφαρμόζονται στον ανιχνευτή, θα πρέπει η απόσταση των δυο ματιών να είναι τουλάχιστον 20 pixels ώστε το πρόσωπο να ανιχνευτεί.

Η εικόνα εισόδου του αλγορίθμου πρέπει να είναι κλίμακας του γκρι (grayscale). Αυτή η επιλογή έγινε για μείωση των δεδομένων που επεξεργάζονται και κατ' επέκταση μείωση του χρονικού κόστους εκτέλεσης. Ακόμη, ο αλγόριθμος απαιτεί σαν είσοδο την τιμή κατωφλίου θ_{V-J} η οποία παίζει πολύ σημαντικό ρόλο στην αποτελεσματικότητά του. Αρχικά, ο αλγόριθμος εκτελείται για τη βάση δεδομένων IMM [NLSS04], η οποία

Αλγόριθμος 2 Ταξινομητής Cascade

Input: Βαθμός false-positive f , βαθμός ανίχνευσης d και στόχος βαθμού false-positive F_{target} . Σύνολο P θετικών δειγμάτων (προσώπων) και σύνολο N αρνητικών δειγμάτων (μη-προσώπων)

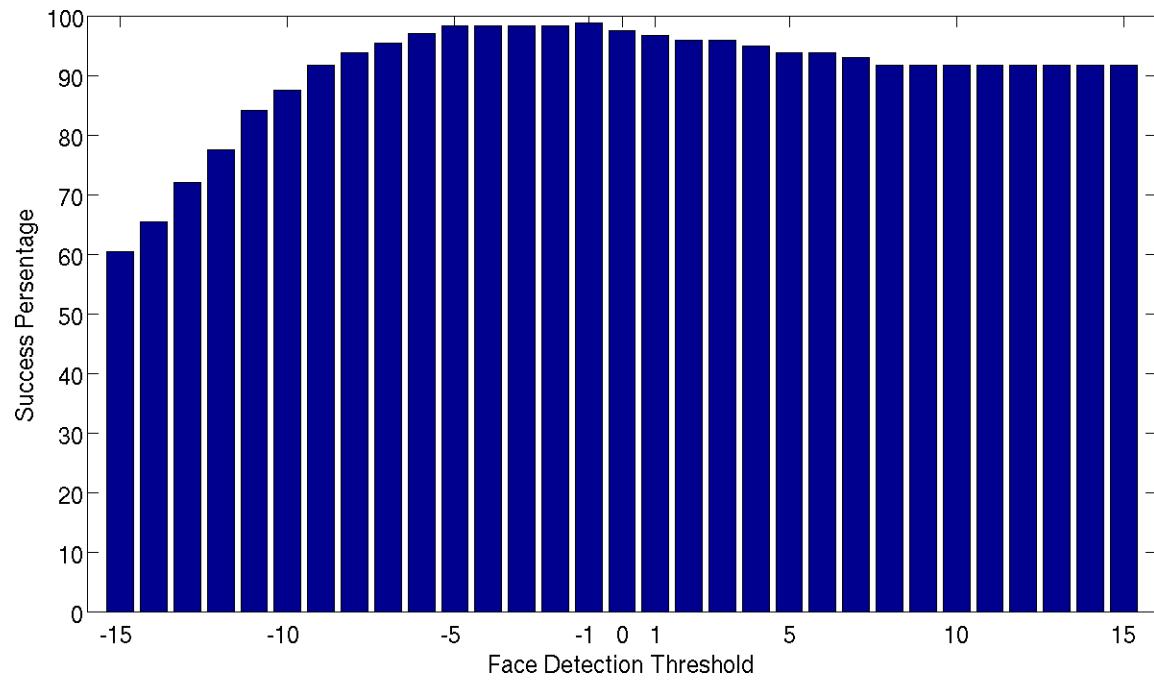
- 1: Αρχικοποίηση $F_0 = 1$, $D_0 = 1$, $i = 0$
- 2: **while** $F_i > F_{target}$ και $n_i < N$ **do**
- 3: $i = i + 1$, $n_i = 0$ και $F_i = F_{i-1}$
- 4: **while** $F_t > F_{i-1}$ **do**
- 5: $n_i = n_i + 1$
- 6: Εκτέλεση αλγορίθμου AdaBoost με P , N και τα n_i χαρακτηριστικά
- 7: Υπολογισμός του τρέχοντος ταξινομητή για το σύνολο που προέκυψε ώστε να καθοριστούν τα F_i και D_i
- 8: Μείωση της τιμής κατωφλίου του i -οστού ταξινομητή ώστε ο βαθμός ανίχνευσης να γίνει τουλάχιστον $d \cdot D_{i-1}$
- 9: **end while**
- 10: $N = 0$
- 11: **if** $F_i > F_{target}$ **then**
- 12: Υπολογισμός του τρέχοντος ανιχνευτή cascade για τις εικόνες μη-προσώπου και τοποθέτηση τυχόντων λανθασμένων ανιχνεύσεων στο σύνολο N
- 13: **end if**
- 14: **end while**

χρησιμοποιείται στο επόμενο μέρος για την εκπαίδευση του συστήματος αναγνώρισης προσώπου. Στο Σχήμα 4.6 φαίνεται το διάγραμμα του ποσοστού επιτυχίας του αλγορίθμου σε όλες τις εικόνες της βάσης δεδομένων για διάφορες τιμές κατωφλίου θ_{V-J} . Η λειτουργία του αλγορίθμου κρίνεται εξαιρετική. Συγκεκριμένα τα καλύτερα αποτελέσματα επιτυγχάνονται για τιμή κατωφλίου στο διάστημα

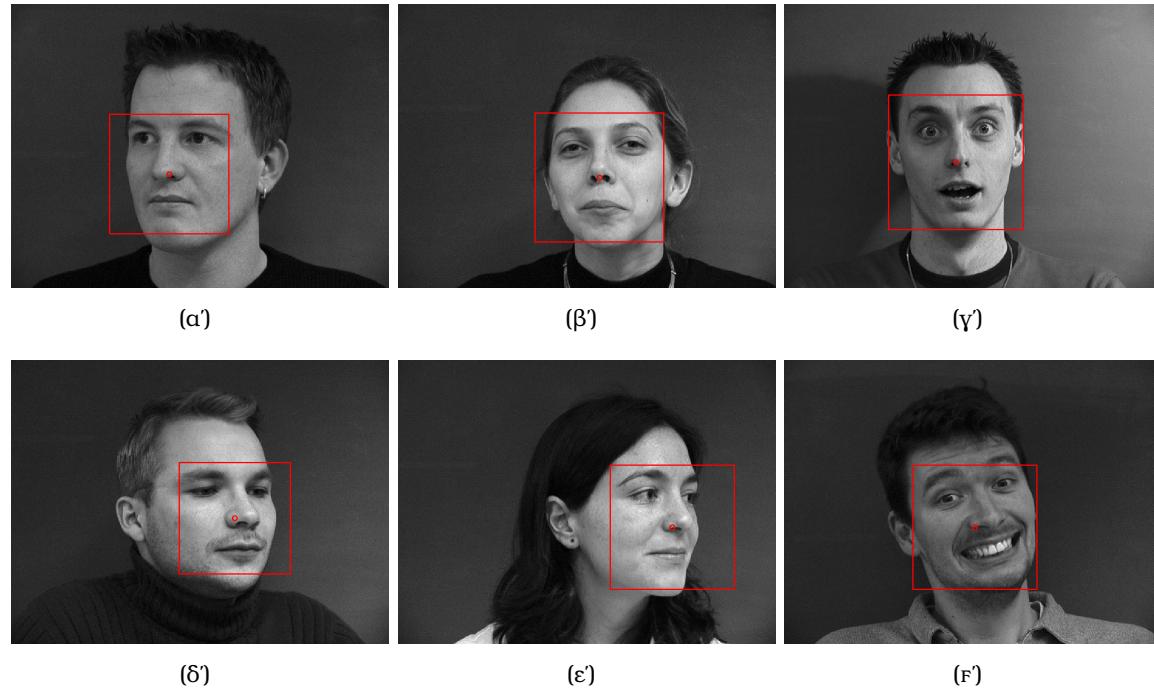
$$\theta_{V-J} \in [-6, 3]$$

με καταλληλότερη τιμή $\theta_{V-J} = -1$, στην οποία οι αποτυχίες εμφανίζονται σε 3 εικόνες από τις 240 συνολικά της βάσης δεδομένων. Τονίζεται ότι ως αποτυχία θεωρείται η μη ανίχνευση κανενός προσώπου, η ανίχνευση ενός λανθασμένου προσώπου, αλλά και η ανίχνευση πολλών προσώπων, δεδομένου ότι κάθε εικόνα της βάσης δεδομένων περιέχει ακριβώς ένα πρόσωπο.

Στο Σχήμα 4.7 φαίνονται έξι παραδείγματα εκτέλεσης του αλγορίθμου σε εικόνες της βάσης δεδομένων IMM. Ωστόσο οι εικόνες αυτές είναι σχετικά "εύκολες" για τον ανιχνευτή, καθώς περιέχουν ένα πρόσωπο κοντινό στην κάμερα και μονόχρωμο παρασκήνιο. Στο Σχήμα 4.8 φαίνονται παραδείγματα εικόνων με περισσότερα πρόσωπα και πολύπλοκο παρασκήνιο. Κάθε φορά είναι στην ευχέρεια του χρήστη η καταλληλότερη επιλογή της τιμής κατωφλίου. Η περίπτωση των εικόνων με πολύπλοκο παρασκήνιο, στην οποία υπάρχει μεγάλος κίνδυνος να ανιχνευτούν αρκετά λανθασμένα πρόσωπα, αντιμετωπίζεται στο επόμενο κεφάλαιο με τη συνδυασμένη χρήση του ανιχνευτή προσώπου με τον ανιχνευτή δέρματος.



Σχήμα 4.6: Ποσοστό επιτυχίας του ανιχνευτή προσώπου στη βάση δεδομένων IMM για διάφορες τιμές κατωφλίου



Σχήμα 4.7: Παραδείγματα ανίχνευσης προσώπου στη βάση δεδομένων IMM



Σχήμα 4.8: Παραδείγματα ανίχνευσης προσώπου σε εικόνες με πολλά πρόσωπα ή/και πολύπλοκο παρασκήνιο

Η ταχύτητα εκτέλεσης του αλγορίθμου ανά εικόνα εισόδου διάστασης 640×480 είναι $0.024sec = 24msec$. Ενώ σε περιπτώσεις εικόνων με μεγάλο αριθμό προσώπων (περίπου 30-40) ή πολύπλοκο παρασκήνιο ο χρόνος γίνεται $0.11sec = 110msec$.

Ο αλγόριθμος είναι έτσι δομημένος ώστε να γίνεται σωστή ανίχνευση προσώπου ακόμα και όταν μέρος αυτού είναι κρυμμένο. Η εξαίρεση είναι όταν υπάρχει εμπόδιο μπροστά στα μάτια, καθώς τα Haar χαρακτηριστικά που χρησιμοποιήθηκαν βασίζον-

ται πολύ στην περιοχή των ματιών. Ωστόσο αν υπάρχει εμπόδιο μπροστά στο στόμα, η ανίχνευση γίνεται φυσιολογικά. Ένα τέτοιο παράδειγμα φαίνεται στο Σχήμα 4.9.



Σχήμα 4.9: Παραδείγματα ανίχνευσης προσώπου με εμπόδια

Κεφάλαιο 5

Συνδυασμός Ανιχνευτή Προσώπου και Ανιχνευτή Δέρματος

Στο κεφάλαιο αυτό παρουσιάζεται ένα τελικό σύστημα ανίχνευσης προσώπου που συνδυάζει τον ανιχνευτή δέρματος του κεφαλαίου 3 με τον ανιχνευτή προσώπου Viola-Jones του κεφαλαίου 4.

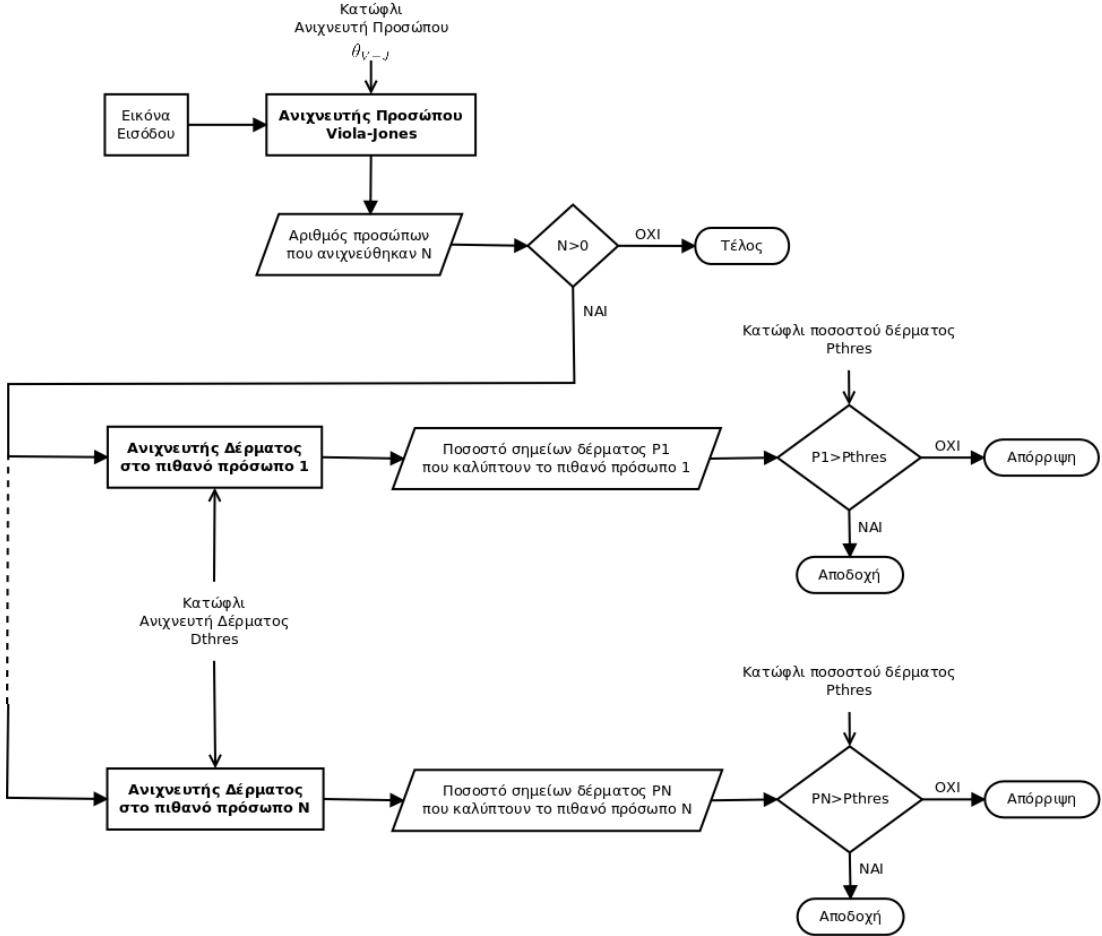
5.1 Τελικό Σύστημα Ανίχνευσης Προσώπου

Όπως εξηγήθηκε στο κεφάλαιο 4, η λειτουργία του ανιχνευτή προσώπου Viola-Jones είναι πολύ αποτελεσματική σε περιπτώσεις εικόνων με απλοϊκό παρασκήνιο (background). Ωστόσο, όταν το παρασκήνιο γίνεται πολύπλοκο, αποτελούμενο από πολλά χρώματα και αντικείμενα διαφόρων σχημάτων και μεγεθών, τότε ο ανιχνευτής προσώπου δεν είναι πολύ σταθερός. Συγκεκριμένα, ανάλογα με την τιμή κατωφλίου θ_{V-J} που επιλέγει ο χρήστης μπορεί να γίνεται σωστή ανίχνευση του προσώπου αλλά και ταυτόχρονη λανθασμένη ανίχνευση διαφόρων άλλων αντικειμένων ή να αποφεύγεται η λανθασμένη ανίχνευση αντικειμένων που αποτελούν μη-πρόσωπα με κίνδυνο να μην ανιχνεύεται ούτε το ίδιο το πρόσωπο.

Για να αντιμετωπιστεί το πρόβλημα αυτό προτείνεται η συνδυασμένη χρήση του ανιχνευτή προσώπου με τον ανιχνευτή δέρματος. Ο συνδυασμός αυτός γίνεται με τέτοιο τρόπο ώστε το τελικό σύστημα να εκμεταλλεύεται τα πλεονεκτήματα της κάθε μεθόδου.

Έτσι από τον ανιχνευτή προσώπου είναι απαραίτητος ο εντοπισμός όλων των προσώπων που περιέχει η εικόνα ακόμα και αν αυτό απαιτεί τη λανθασμένη ανίχνευση άλλων αντικειμένων ως πρόσωπα. Από τον ανιχνευτή δέρματος θέλουμε να αποφύγουμε την λανθασμένη ανίχνευση σημείων του παρασκηνίου ως δέρμα, ενώ είμαστε σίγουροι ότι εντοπίζει κάποια από τα σημεία του πραγματικού προσώπου της εικόνας ως δέρμα. Το τελικό σύστημα ανίχνευσης προσώπου αποτελείται από τον ανιχνευτή προσώπου Viola-Jones, στην έξοδο του οποίου εφαρμόζεται ο ανιχνευτής δέρματος. Στο Σχήμα 5.1 φαίνεται το διάγραμμα ροής του συστήματος.

Το τελικό σύστημα ανίχνευσης προσώπου ουσιαστικά εκτελεί τον ανιχνευτή δέρματος σε κάθε περιοχή στην οποία ο ανιχνευτής προσώπου Viola-Jones θεωρεί ότι



Σχήμα 5.1: Συνολικό σύστημα ανιχνευτή προσώπου

υπάρχει πρόσωπο. Με τον τρόπο αυτό ο ανιχνευτής δέρματος εκτελείται σε πολύ μικρά παράθυρα της εικόνας εισόδου με αποτέλεσμα να μην αυξάνει το χρονικό κόστος εκτέλεσης του συνολικού συστήματος. Τα βήματα του τελικού συστήματος ανίχνευσης προσώπου φαίνονται στον αλγόριθμο 3. Χρησιμοποιείται ο πιθανοτικός αλγόριθμος ανίχνευσης δέρματος και όχι ο χρωματικός με HSV.

5.2 Πειραματικά Αποτελέσματα

Το χρονικό κόστος του τελικού συνδυασμένου αλγορίθμου ανά εικόνα εισόδου είναι περίπου $0.06sec = 60msec$. Αυτό μεταφράζεται σε $17frames/sec$, το οποίο είναι ιδιαίτερα ικανοποιητικό δεδομένου ότι το τελικό σύστημα αναγνώρισης εκφράσεων προσώπου λειτουργεί σε πραγματικό χρόνο με είσοδο από βίντεο. Βέβαια τονίζεται ότι η χρησιμότητά του είναι μόνο για εικόνες με πολύπλοκο παρασκήνιο, καθώς σε οποιαδήποτε άλλη περίπτωση ο ανιχνευτής προσώπου Viola-Jones είναι αρκετός.

Το τελικό σύστημα δέχεται τρεις παραμέτρους: το κατώφλι του ανιχνευτή προ-

Αλγόριθμος 3 Τελικό σύστημα ανίχνευσης προσώπου με συνδυασμό ανιχνευτή προσώπου Viola-Jones και ανιχνευτή δέρματος

Input: Εικόνα εισόδου I , κατώφλι ανιχνευτή Viola-Jones θ_{V-J} , κατώφλι ανιχνευτή δέρματος D_{thres} και ελάχιστο ποσοστό των σημείων μιας περιοχής που αποτελούν δέρμα P_{thres}

- 1: Εκτέλεση ανιχνευτή προσώπου Viola-Jones με εισόδους I και θ_{V-J} . Η συνάρτηση επιστρέφει τον αριθμό προσώπων που ανιχνεύθηκαν N , τις συντεταγμένες του κέντρου του κάθε προσώπου και το πλάτος του παραθύρου
- 2: **if** $N > 0$ **then**
- 3: **for** όλα τα παράθυρα πιθανού προσώπου $n = 1, \dots, N$ **do**
- 4: Εκτέλεση πιθανοτικού ανιχνευτή δέρματος με είσοδο το παράθυρο n και το κατώφλι D_{thres}
- 5: Υπολογισμός του ποσοστού P_n της συνολικής επιφάνειας του παραθύρου που καλύπτεται από σημεία δέρματος
- 6: **if** $P_n > P_{thres}$ **then**
- 7: Αποδοχή του παραθύρου n ως πρόσωπο
- 8: **else**
- 9: Απόρριψη του παραθύρου n ως μη-πρόσωπο
- 10: **end if**
- 11: **end for**
- 12: **else**
- 13: Δεν βρέθηκε κανένα πρόσωπο στην εικόνα I
- 14: **end if**

σώπου Viola-Jones θ_{V-J} , το κατώφλι του πιθανοτικού ανιχνευτή δέρματος D_{thres} και το ελάχιστο ποσοστό κάλυψης του παραθύρου από δέρμα P_{thres} . Αρχικά η τιμή του θ_{V-J} πρέπει να οριστεί αρκετά μεγάλη, με βάση και το Σχήμα 4.6, ώστε να είναι σίγουρο ότι το πραγματικό πρόσωπο είναι μέσα στις πιθανές περιοχές προσώπου και δεν έχει απορριφθεί. Δεν αποτελεί πρόβλημα ο μεγάλος αριθμός λανθασμένων περιοχών (περιοχών που δεν περιέχουν πρόσωπο), διότι αναμένεται να απορριφθούν εύκολα από τον ανιχνευτή δέρματος. Ενδεικτικές τιμές του κατωφλίου του ανιχνευτή προσώπου είναι

$$\theta_{V-J} \in [3, 12]$$

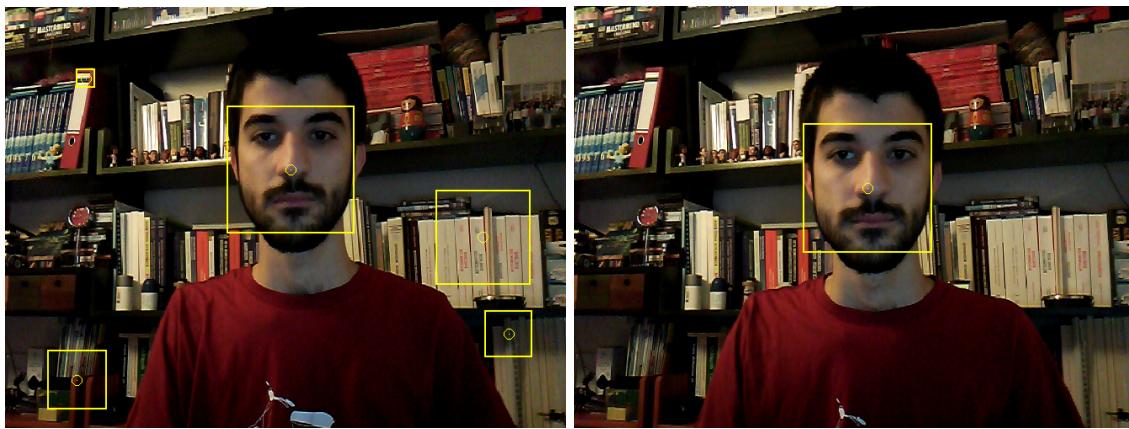
Οι ενδεικτικές τιμές του κατωφλίου του ανιχνευτή δέρματος παραμένουν ίδιες, δηλαδή

$$D_{thres} \in [-15, 4]$$

με καταλληλότερη την $D_{thres} = -1$. Όσον αφορά στην παράμετρο P_{thres} , κατόπιν αρκετών δοκιμών προέκυψε

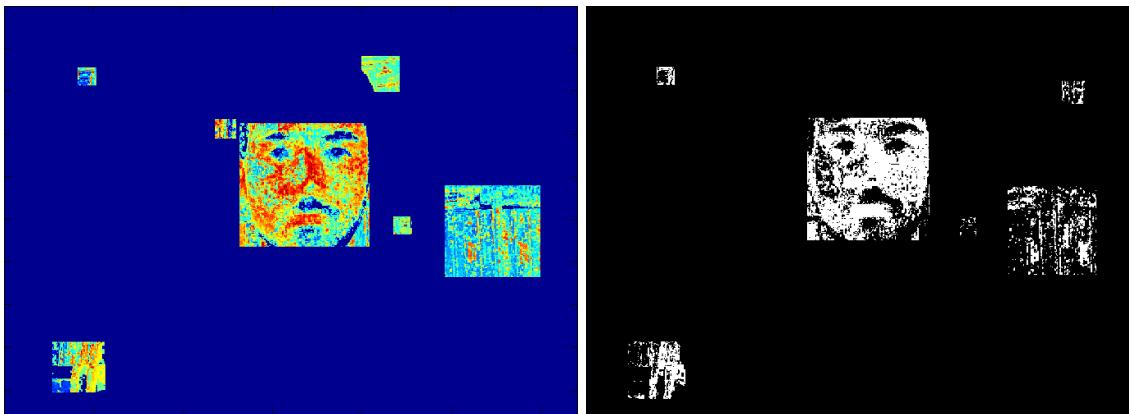
$$P_{thres} \in [30, 60]$$

με καταλληλότερη τιμή την $P_{thres} = 50\%$. Στο Σχήμα 5.2 φαίνεται ένα παράδειγμα εκτέλεσης του συνδυασμένου αλγορίθμου ανιχνευτή προσώπου με όλα τα ενδιάμεσα στάδια σε μια εικόνα με πολύπλοκο και πολύχρωμο παρασκήνιο.



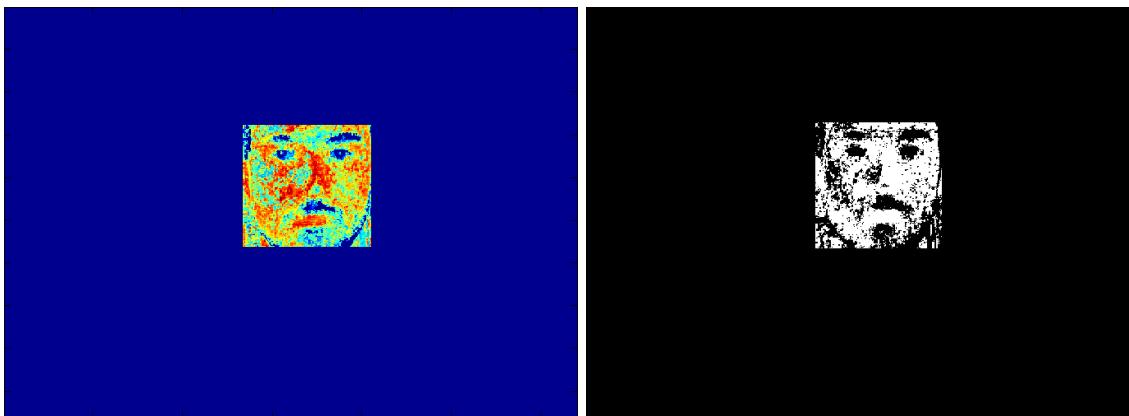
(α') Ανίχνευση προσώπου χωρίς ανίχνευση δέρματος

(β') Τελική ανίχνευση προσώπου



(γ') Πιθανότητα δέρματος σε όλα τα παράθυρα πιθανού προσώπου

(δ') Ανίχνευση δέρματος σε όλα τα παράθυρα πιθανού προσώπου



(ε') Παράθυρο πιθανότητας δέρματος μετά την εφαρμογή του ποσοσού κάλυψης

(φ') Παράθυρο ανίχνευσης δέρματος μετά την εφαρμογή του ποσοσού κάλυψης

Σχήμα 5.2: Παράδειγμα συνδυασμού ανιχνευτή προσώπου και ανιχνευτή δέρματος

Μέρος II

**Οπτική Μοντελοποίηση Προσώπου
με Active Appearance Models σε
Πραγματικό Χρόνο**

Κεφάλαιο 6

Εισαγωγή

Τα Active Appearance Models (AAM) είναι μια μέθοδος για το ταίριασμα ενός στατιστικού μοντέλου του σχήματος και της υφής ενός αντικειμένου σε μια νέα εικόνα. Η μεταβλητότητα του σχήματος και της υφής παράγεται από τη διαδικασία της εκπαίδευσης σε ένα κατάλληλο σύνολο εικόνων. Η χρήση της μεθόδου Principal Component Analysis (PCA) στα δεδομένα του σχήματος και της υφής οδηγεί στη δημιουργία ενός παραμετροποιημένου μοντέλου που περιλαμβάνει πλήρως όλη την πληροφορία των εικόνων εκπαίδευσης.

Η μέθοδος ταιριάσματος ενός AAM μοντέλου σε μια νέα εικόνα προσώπου είναι ένα μη-γραμμικό πρόβλημα βελτιστοποίησης. Συγκεκριμένα, ο στόχος είναι η ελαχιστοποίηση της διαφοράς μεταξύ της υφής που δημιουργήθηκε από το μοντέλο υφής και της υφής της εικόνας εισόδου που καλύπτεται από το μοντέλο σχήματος.

6.1 Προεπισκόπηση

Στο Μέρος II της εργασίας αναλύεται η δημιουργία και εκπαίδευση ενός μοντέλου Active Appearance Model (AAM). Συγκεκριμένα, εξηγείται η δημιουργία του μοντέλου σχήματος προσώπου και του μοντέλου υφής προσώπου. Επίσης, εξηγείται η χρήση του μετασχηματισμού ομοιότητας για την μοντελοποίηση της πόζας του προσώπου, καθώς και η χρήση ενός ομοπαραλληλικού (affine) μετασχηματισμού για το φωτισμό της υφής του προσώπου.

Στη συνέχεια εξηγούνται οι προσαρμογές που προτείνουν οι Papadopoulos και Mairagkos [PaMa08]. Αρχικά, γίνεται μια προσαρμογή στον πίνακα ταιριάσματος AAM που αντισταθμίζει την απόκλιση της συνθετικής υφής του μοντέλου από τη μέση υφή. Το χαρακτηριστικό αυτής της προσαρμογής είναι ότι εισάγεται μια παράμετρος που αντιστοιχεί στο βαθμό του αλγορίθμου η οποία ρυθμίζει την ακρίβεια έναντι της υπολογιστικής πολυπλοκότητας. Μάλιστα αποδεικνύεται ότι οι αλγόριθμοι Project-Out και Simultaneous που προτάθηκαν από τους Baker και Matthews [MaBa04, GMB05] αντιστοιχούν ουσιαστικά στην ελάχιστη (μηδενική) και μέγιστη τιμή της παραμέτρου του βαθμού του αλγορίθμου. Μια δεύτερη προσαρμογή που αναλύεται είναι η ανανέωση του μέσου προτύπου υφής. Αποδεικνύεται ότι η διαδικασία αυτή είναι ιδιαίτερα μικρής χρονικής πολυπλοκότητας και βελτιώνει πολύ τη λειτουργία του αλγορίθμου με

μικρό επιλεγμένο βαθμό. Η τρίτη και σημαντικότερη προσαρμογή που προτείνεται είναι η ενσωμάτωση στη διαδικασία ταιριάσματος prior περιορισμών στις παραμέτρους του μοντέλου. Αυτή η προσαρμογή μετατρέπει τον αντίστροφο-συνθετικό (inverse-compositional) αλγόριθμο σε ευθύ-προσθετικό (forwards-additive) που βελτιώνει την αποτελεσματικότητα και την σταθερότητα της διαδικασίας ταιριάσματος των AAM.

Όλα τα παραπάνω συνοδεύονται από την απαραίτητη μαθηματική ανάλυση αλλά και από αναλυτικά σχήματα και παραδείγματα. Στην τελευταία παράγραφο του Μέρους II παρουσιάζονται με λεπτομέρεια τα πειραματικά αποτελέσματα.

Κεφάλαιο 7

Εκπαίδευση Μοντέλου ΑΑΜ

Στο κεφάλαιο αυτό περιγράφεται η δημιουργία ενός μοντέλου σχήματος προσώπου και ενός μοντέλου υφής, με τέτοιο τρόπο ώστε να περιέχουν όλη τη μεταβλητότητα της βάσης εικόνων. Κατόπιν, εξηγείται πως δημιουργείται το ανεξάρτητο μοντέλο που συνδυάζει τη μεταβλητότητα των δυο προηγούμενων. Σημειώνεται ότι η βάση εικόνων που χρησιμοποιείται για τα παραδείγματα που ακολουθούν είναι η IMM [NLSS04] του Technical University of Denmark (DTU).

7.1 Μοντέλο Σχήματος Προσώπου

Το σχήμα ορίζεται ως:

“... όλη η γεωμετρική πληροφορία που απομένει όταν αφαιρεθούν από ένα αντικείμενο οι επιδράσεις περιστροφής, κλιμάκωσης και θέσης.”
— David George Kendall (μαθηματικός)

Σε αυτή την παράγραφο περιγράφεται η κατασκευή ενός μοντέλου σχήματος προσώπου. Το σχήμα περιγράφεται από ένα δίανυσμα συντεταγμένων των *landmark* σημείων. Όλα τα σχήματα της βάσης εικόνων πρέπει να ευθυγραμμιστούν σε ένα κοινό πλαίσιο αναφοράς με τη μέθοδο General Procrustes Analysis (GPA). Στη συνέχεια με χρήση της μεθόδου Principal Component Analysis (PCA) μελετάται η μεταβλητότητα των ευθυγραμμισμένων σχημάτων ώστε να δημιουργηθούν οι τελικές παράμετροι που καθορίζουν τις μεταβολές του σχήματος.

7.1.1 Landmark Σημεία

Το σχήμα προσώπου αναπαρίσταται από ένα σύνολο L landmark σημείων ορισμένα σε δυο διαστάσεις. Η βάση δεδομένων που χρησιμοποιείται ορίζει το σχήμα προσώπου με $L = 58$ landmark σημεία, τα οποία είναι δια χειρός ορισμένα για καθεμιά από τις 240 εικόνες που περιέχει. Ένα παράδειγμα φαίνεται στο Σχήμα 7.1.

Από μαθηματικής άποψης λοιπόν το σχήμα αναπαρίσταται από ένα διάνυσμα μήκους $2L$. Συγκεκριμένα ένα σχήμα με L landmark σημεία $\{(x_i, y_i) : i = 1, \dots, L\}$

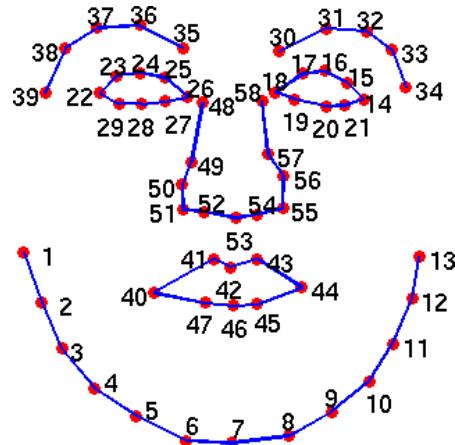
ορίζεται με το $2L$ διάνυσμα

$$\mathbf{s} = [x_1, y_1, x_2, y_2, \dots, x_{L-1}, y_{L-1}, x_L, y_L]^T \quad (7.1)$$



Σχήμα 7.1: Εικόνα προσώπου που το σχήμα έχει ορισθεί με 58 landmark σημεία

Ωστόσο, το διάνυσμα σχήματος δεν περιέχει καμία πληροφορία σχετικά με τον τρόπο που ενώνονται τα L αυτά σημεία. Τα σημεία είναι αριθμημένα με τέτοια σειρά ώστε να αναπαριστούν σημαντικές θέσεις του προσώπου και χωρίζοντάς τα σε υποσύνολα να οριοθετούν τα σημαντικά του τμήματα (μάτια, στόμα, μύτη, φρύδια, σαγών). Επομένως χρησιμοποιείται ένα κοινός τρόπος ένωσης των landmark σημείων όπως φαίνεται στο Σχήμα 7.2.



Σχήμα 7.2: Αριθμηση landmark σημείων

Όλα τα διανύσματα σχήματος s_i των εικόνων προς εκπαίδευση της βάσης δεδομένων περιέχουν ένα μετασχηματισμό. Σε δισδιάστατες εικόνες ο μετασχηματισμός αυτός είναι Μετασχηματισμός Ομοιότητας, δηλαδή περιστροφή, κλιμάκωση και μετακίνηση. Για να επιτευχθεί η δημιουργία ενός μοντέλου σχήματος προσώπου με στατιστική εγκυρότητα που να περιγράφει της σχέσεις σχήματος μεταξύ των εικόνων, θα πρέπει να αφαιρεθεί οποιαδήποτε μεταβλητότητα οφείλεται στον Μετασχηματισμό Ομοιότητας. Αυτό επιτυγχάνεται με την ευθυγράμμιση των σχημάτων σε ένα κοινό σύστημα συντεταγμένων, όπως περιγράφεται στην επόμενη παράγραφο.

7.1.2 Ευθυγράμμιση Σχημάτων

Η διαδικασία της ευθυγράμμισης δυο σχημάτων περιλαμβάνει την εύρεση των παραμέτρων του Μετασχηματισμού Ομοιότητας (κλιμάκωση, περιστροφή και μετατόπιση) ώστε να μειωθεί βέλτιστα η διαφορά μεταξύ των δυο σχημάτων βάσει ενός τρόπου μέτρησης. Η επίλυση του συγκεκριμένου προβλήματος γίνεται με την μέθοδο *Procrustes Analysis* [Gow75].

Η μέθοδος αυτή ευθυγραμμίζει δυο σχήματα που έχουν τον ίδιο αριθμό landmark σημείων με ένα-προς-ένα αντιστοιχίες σημείων. Η ευθυγράμμιση περισσότερων από δυο σχημάτων βασίζεται στην ευθυγράμμιση ζεύγων σχημάτων, όπου έχουμε επιλέξει το ένα από αυτά να είναι το σχήμα αναφοράς. Τελικά η διαδικασία ευθυγράμμισης καταφέρνει να αφαιρέσει την Ευκλείδια πληροφορία Ομοιότητας από το σύνολο εικόνων και να αφήσει την ποικιλία εκφράσεων, ταυτότητας και πόζας για περαιτέρω στατιστική ανάλυση.

7.1.2.1 Ευθυγράμμιση Δυο Σχημάτων (Procrustes Analysis)

Έστω ότι έχουμε δυο διανύσματα σχημάτων \mathbf{x}_1 και \mathbf{x}_2 . Η ευθυγράμμιση των δυο σχημάτων ισοδυναμεί με την εύρεση των παραμέτρων του μετασχηματισμού ομοιότητας \mathbf{T} , δηλαδή του ποσοστού κλιμάκωσης s , της γωνίας περιστροφής θ και του διαστήματος μετατόπισης στους δυο άξονες (t_x, t_y) , έτσι ώστε όταν εφαρμοστεί ο μετασχηματισμός στο σχήμα \mathbf{x}_1 να ευθυγραμμιστεί βέλτιστα με το \mathbf{x}_2 . Το κριτήριο είναι η ελαχιστοποίηση της Procrustes απόστασης

$$D_{\text{Procrustes}}(\mathbf{x}_1, \mathbf{x}_2) = \sqrt{\sum_{i=1}^L (x_{1,i} - x_{2,i})^2 + (y_{1,i} - y_{2,i})^2} \quad (7.2)$$

ως προς τις παραμέτρους s , θ και (t_x, t_y) . Αρχικά ευθυγραμμίζονται τα κεντροειδή $\bar{\mathbf{x}}_1$ και $\bar{\mathbf{x}}_2$ των δυο σχημάτων με τις σχέσεις

$$\mathbf{x}_{1c} = \mathbf{x}_1 - \bar{\mathbf{x}}_1, \quad \mathbf{x}_{2c} = \mathbf{x}_2 - \bar{\mathbf{x}}_2$$

όπου τελικά οι παράμετροι (t_x, t_y) δίνονται από το $\bar{\mathbf{x}}_2$. Στη συνέχεια η παράμετρος της κλίμακας s κανονικοποιείται εφαρμόζοντας έναν Ισομορφικό μετασχηματισμό

$$\hat{\mathbf{x}}_1 = \frac{\mathbf{x}_{1c}}{\|\mathbf{x}_{1c}\|}, \quad \hat{\mathbf{x}}_2 = \frac{\mathbf{x}_{2c}}{\|\mathbf{x}_{2c}\|}$$

Όσον αφορά στην παράμετρο περιστροφής θ , ορίζουμε τη μήτρα περιστροφής

$$\mathbf{R} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$$

και αναπαριστούμε τα ευθυγραμμισμένα ως προς την κλίμακα και την μετατόπιση σχήματα με τη μορφή διανυσμάτων $\mathbf{X}_1' = [\hat{x}_1 | \hat{y}_1]_{L \times 2}$ και $\mathbf{X}_2' = [\hat{x}_2 | \hat{y}_2]_{L \times 2}$. Έτσι η διαφορά \mathbf{E} μετά την περιστροφή του \mathbf{X}_2' γίνεται

$$\mathbf{E} = \mathbf{R}\mathbf{X}_2' - \mathbf{X}_1'$$

Όμως δεδομένου ότι το ίχνος (trace) του πίνακα $\mathbf{E}\mathbf{E}^T$ ισούται με την Procrustes απόσταση της εξίσωσης (7.2)

$$tr(\mathbf{E}\mathbf{E}^T) = D_{Procrustes}(\mathbf{x}_1, \mathbf{x}_2) \quad (7.3)$$

η ελαχιστοποίηση της Procrustes απόστασης είναι ισοδύναμο πρόβλημα με την ελαχιστοποίηση του $tr(\mathbf{E}\mathbf{E}^T)$ ως προς την περιστροφή θ . Με εφαρμογή της μεθόδου Singular Value Decomposition (SVD) στον πίνακα $\mathbf{X}_2'\mathbf{X}_1'$

$$SVD(\mathbf{X}_2'\mathbf{X}_1') = \mathbf{U}\mathbf{S}\mathbf{V}^T$$

βγαίνει το συμπέρασμα ότι η βέλτιστη μήτρα περιστροφής που ευθυγραμμίζει το \mathbf{X}_2' με το \mathbf{X}_1' ελαχιστοποιώντας την εξίσωση (7.3) είναι

$$\mathbf{R} = \mathbf{U}\mathbf{V}^T$$

Η μέθοδος Procrustes Analysis συνοψίζεται στον αλγόριθμο 4.

Αλγόριθμος 4 Procrustes Analysis

1: Υπολογισμός κεντροειδών:

$$(\bar{x}_1, \bar{y}_1) = \left(\frac{1}{L} \sum_{i=1}^L x_{1i}, \frac{1}{L} \sum_{i=1}^L y_{1i} \right) \text{ και } (\bar{x}_2, \bar{y}_2) = \left(\frac{1}{L} \sum_{i=1}^L x_{2i}, \frac{1}{L} \sum_{i=1}^L y_{2i} \right)$$

2: Ευθυγράμμιση κεντροειδών:

$$(x_{1c}, y_{1c}) \rightarrow (x_1 - \bar{x}_1, y_1 - \bar{y}_1) \text{ και } (x_{2c}, y_{2c}) \rightarrow (x_2 - \bar{x}_2, y_2 - \bar{y}_2)$$

3: Κανονικοποίηση σχημάτων με ισομορφική κλίμακα: $\hat{\mathbf{x}}_1 = \frac{\mathbf{x}_{1c}}{\|\mathbf{x}_{1c}\|}$ και $\hat{\mathbf{x}}_2 = \frac{\mathbf{x}_{2c}}{\|\mathbf{x}_{2c}\|}$

4: Δημιουργία πινάκων: $\mathbf{X}_1' = [\hat{x}_1 | \hat{y}_1]_{L \times 2}$ και $\mathbf{X}_2' = [\hat{x}_2 | \hat{y}_2]_{L \times 2}$

5: Εφαρμογή $SVD(\mathbf{X}_2'\mathbf{X}_1') = \mathbf{U}\mathbf{S}\mathbf{V}^T$

6: Η βέλτιστη μήτρα περιστροφής δίνεται από: $\mathbf{R} = \mathbf{U}\mathbf{V}^T$

7.1.2.2 Ευθυγράμμιση Πολλαπλών Σχημάτων (Generalized Procrustes Analysis)

Η μέθοδος Generalized Procrustes Analysis (GPA) [Gow75] αποτελείται από διαδοχικές εκτελέσεις της Procrustes Analysis σε ζεύγη σχημάτων χρησιμοποιώντας ένα σχήμα αναφοράς και ευθυγραμμίζοντας τα υπόλοιπα σε αυτό. Στη δική μας περίπτωση το σχήμα αναφοράς είναι το μέσο σχήμα. Αρχικά, οποιοδήποτε σχήμα μπορεί να επιλεγεί ως το αρχικό μέσο σχήμα. Μετά την εκτέλεση του αλγορίθμου 4 για όλα τα σχήματα (M) υπολογίζεται μια νέα εκτίμηση του μέσου σχήματος και η διαδικασία εκτελείται και πάλι. Οι επαναλήψεις σταματάνε όταν δεν υπάρχουν μεγάλες αλλαγές στο μέσο σχήμα μεταξύ διαδοχικών εκτελέσεων. Η μέθοδος GPA περιγράφεται στον αλγόριθμο 5.

Η μέθοδος GPA αρκεί να εκτελεστεί για δυο επαναλήψεις, καθώς στις επόμενες η μεταβολή του μέσου σχήματος είναι από ελάχιστη έως μηδαμινή. Στο Σχήμα 7.3 φαίνεται το σύνολο των landmark σημείων όλων των εικόνων της βάσης δεδομένων πριν και μετά τη διαδικασία της ευθυγράμμισης καθώς και το τελικό μέσο σχήμα προσώπου.

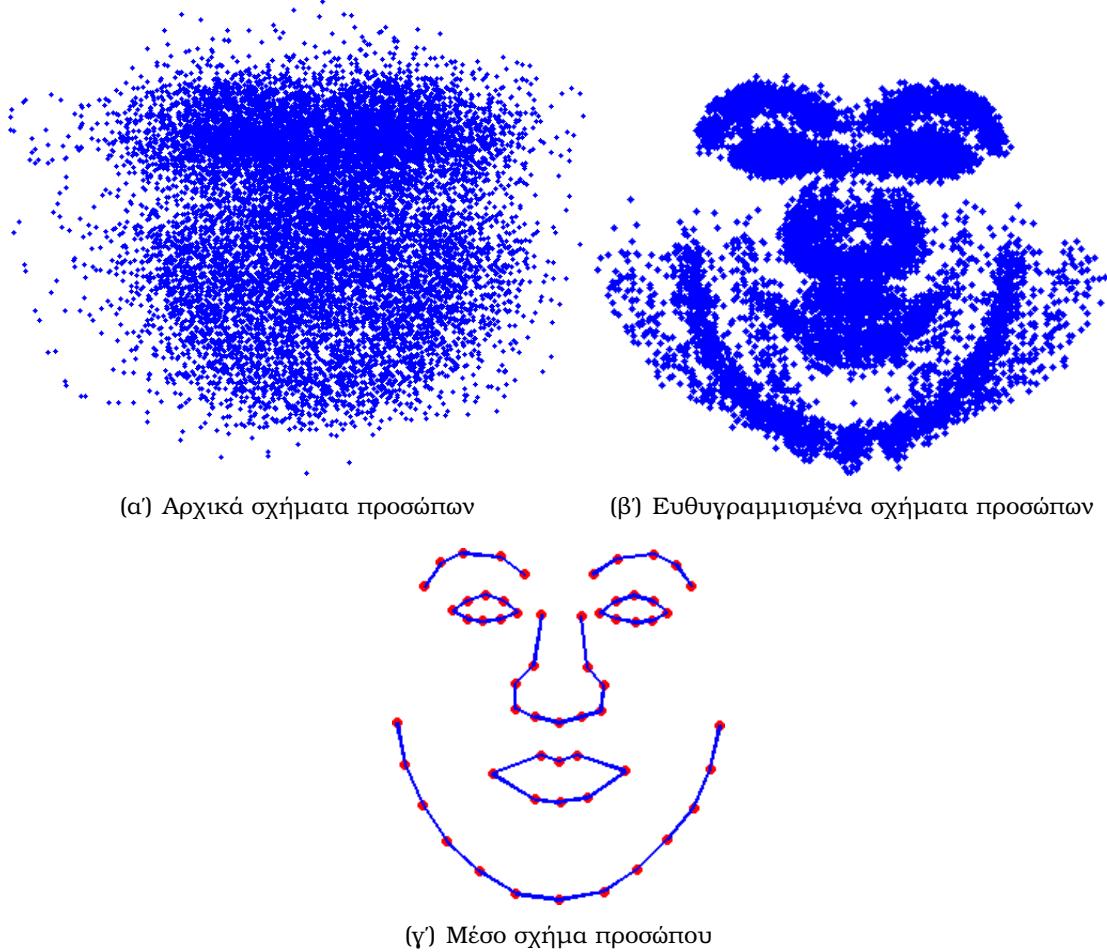
Αλγόριθμος 5 Generalized Procrustes Analysis (GPA)

- 1: Επιλογή του πρώτου σχήματος σαν την αρχική εκτίμηση του μέσου σχήματος:
 $\bar{s}_0 = s_1$
 - 2: $k = 0$
 - 3: **repeat**
 - 4: **for** $i = 1, \dots, M$ **do**
 - 5: ProcrustesAnalysis(s_i, \bar{s})
 - 6: **end for**
 - 7: $k = k + 1$
 - 8: Νέα εκτίμηση του μέσου σχήματος: $\bar{s}_k = \frac{1}{M} \sum_{i=1}^M s_i$
 - 9: **until** σύγκλιση μέσου σχήματος $\bar{s}_k \approx \bar{s}_{k-1}$
 - 10: Τελική εκτίμηση μέσου σχήματος προσώπου: $s_0 = \bar{s}_k$
-

7.1.3 Principal Component Analysis (PCA)

Η Principal Component Analysis (PCA) [Jol89] είναι μια στατιστική μέθοδος μείωσης των διαστάσεων των δεδομένων εισόδου. Συγκεκριμένα, η μέθοδος PCA περιστρέφει όλο το πλήθος των δισδιάστατων δεδομένων των σχημάτων εκπαίδευσης με σκοπό τη μεγιστοποίηση της μεταβλητότητας και επιστρέφει τους κύριους άξονες των δεδομένων. Με μαθηματική ορολογία ορίζεται ως ένας γραμμικός ορθογώνιος μετασχηματισμός που προβάλλει τα δεδομένα σε ένα νέο σύστημα συντεταγμένων ορισμένο από τον άξονα της μεταβλητότητας των δεδομένων.

Η δική μας βάση δεδομένων αποτελείται από M διανύσματα σχήματος $s_i : i = 1, \dots, M$, όπου κάθε σχήμα s_i είναι ένα διάνυσμα μήκους $2L$ ορισμένο σύμφωνα με τη σχέση (7.1). Σημειώνεται ότι το πλήθος των δειγμάτων πρέπει να είναι μεγαλύτερο από το μήκος τους ($M > 2L$).



Σχήμα 7.3: Ευθυγράμμιση σχημάτων με Generalized Procrustes Analysis

Η μέθοδος PCA εκτελείται ως ανάλυση ιδιοδιανυσμάτων του πίνακα συνδιακύμανσης Σ_s των εικόνων εκπαίδευσης

$$\Sigma_s = \frac{1}{M-1} \mathbf{S} \mathbf{S}^T = \frac{1}{M-1} \sum_{i=1}^M (\mathbf{s}_i - \mathbf{s}_0) (\mathbf{s}_i - \mathbf{s}_0)^T \quad (7.4)$$

όπου \mathbf{S} ένας $2L \times M$ πίνακας

$$\mathbf{S} = [\mathbf{s}_1 - \mathbf{s}_0, \mathbf{s}_2 - \mathbf{s}_0, \dots, \mathbf{s}_M - \mathbf{s}_0]$$

και \mathbf{s}_0 το μέσο σχήμα προσώπου όπως υπολογίστηκε από την Generalized Procrustes Analysis

$$\mathbf{s}_0 = \frac{1}{M} \sum_{i=1}^M \mathbf{s}_i$$

Ο πίνακας συνδιακύμανσης (7.4) έχει τελικά διαστάσεις $2L \times 2L$. Η ανάλυση ιδιοδιανυσμάτων στον πίνακα συνδιακύμανσης δίνει ένα διαγώνιο πίνακα Λ_l με τις

ιδιοτιμές λ_i στην κύρια διαγώνιο και τον πίνακα Φ_l του οποίου οι στήλες του είναι τα ιδιοδιανύσματα ϕ_i . Τα ιδιοδιανύσματα ϕ_i και οι αντίστοιχες ιδιοτιμές του πίνακα συνδιακύμανσης ταξινομούνται έτσι ώστε $\lambda_i \geq \lambda_{i+1}$. Το ιδιοδιάνυσμα που αντιστοιχεί στη μεγαλύτερη ιδιοτιμή αναπαριστά τον άξονα με τη μεγαλύτερη μεταβλητότητα. Η δεύτερη ιδιοτιμή αντιστοιχεί στην αμέσως μεγαλύτερη μεταβλητότητα σε άξονα που είναι ορθογώνιος ως προς τον πρώτο. Αντίστοιχα, τα επόμενα ιδιοδιανύσματα αντιστοιχούν σε άξονες ορθογώνιους με τους προηγούμενους σύμφωνα με τη φθίνουσα μεταβλητότητα των δεδομένων εισόδου. Κάποια ιδιοδιανύσματα αντιστοιχούν σε τόσο μικρό ποσοστό της συνολικής μεταβλητότητας που μπορούν να αγνοηθούν ακόμα και ως θόρυβος.

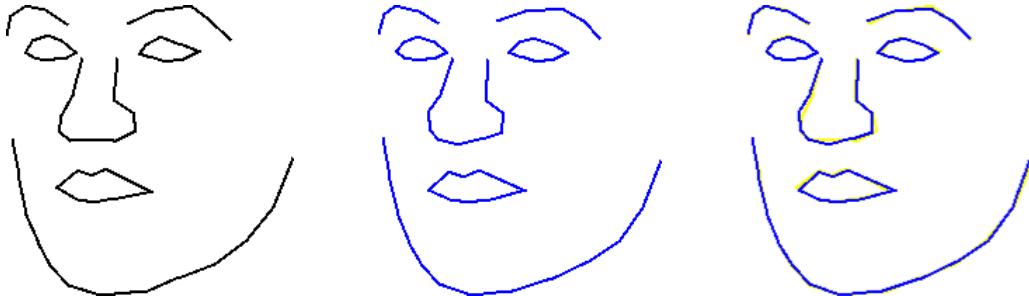
Κρατώντας τις n πιο σημαντικές ιδιοτιμές και τα αντίστοιχα ιδιοδιανύσματα, οποιοδήποτε σχήμα προσώπου από τη βάση εικόνων μπορεί να προσεγγιστεί με μια παραμόρφωση του μέσου σχήματος προσώπου με έναν γραμμικό συνδυασμό των n ιδιοδιανυσμάτων

$$\mathbf{s} \approx \mathbf{s}_0 + \Phi \mathbf{p} \quad (7.5)$$

όπου $\Phi = (\phi_1 | \phi_2 | \dots | \phi_n)$ είναι ορθογώνιος πίνακας και \mathbf{p} είναι ένα διάνυσμα μήκους n που μπορεί να υπολογιστεί ως

$$\mathbf{p} = \Phi^{-1}(\mathbf{s} - \mathbf{s}_0) = \Phi^T(\mathbf{s} - \mathbf{s}_0) \quad (7.6)$$

Ένα τέτοιο παράδειγμα φαίνεται στο Σχήμα 7.4.



(α') Σχήμα προσώπου s ορισμένο από τη βάση δεδομένων

(β') Προσέγγιση αρχικού σχήματος προσώπου με παραμόρφωση του μέσου σχήματος $\mathbf{s}_0 + \Phi \mathbf{p}$ με τους παράγοντες \mathbf{p}

(γ') Τα σχήματα προσώπων των εικόνων 7.4(a) και 7.4(b) ευθυγραμμισμένα

Σχήμα 7.4: Προσέγγιση σχήματος προσώπου ορισμένο από τη βάση δεδομένων με παραμόρφωση του μέσου σχήματος με ένα γραμμικό συνδυασμό ιδιοδιανυσμάτων του μοντέλου σχήματος

7.1.3.1 Αριθμός Αξόνων Μεταβλητότητας

Ο αριθμός των αξόνων μεταβλητότητας (number of modes of variation) που κρατάμε, n , επιλέγεται κατά τέτοιο τρόπο από το χρήστη ώστε να μπορεί να διατηρήσει όσο το

δυνατόν μεγαλύτερο ποσοστό της συνολικής μεταβλητότητας χωρίς όμως ο αριθμός των αξόνων να είναι κοντά στον αρχικό αριθμό διαστάσεων $2L$. Κάθε ιδιοτιμή λ_i δίνει τη μεταβλητότητα των δεδομένων στην κατεύθυνση του αντίστοιχου ιδιοδιανύσματος ϕ_i . Επομένως, η συνολική μεταβλητότητα των δεδομένων δίνεται από το άθροισμα όλων των ιδιοτιμών των M εικόνων εκπαίδευσης

$$V = \sum_{i=1}^M \lambda_i$$

Ο χρήστης επιλέγει τις n σημαντικότερες ιδιοτιμές έτσι ώστε

$$\sum_{i=1}^n \lambda_i \geq \alpha V \quad (7.7)$$

όπου α είναι ένα ποσοστό της συνολικής μεταβλητότητας με ενδεικτικές τιμές στο διάστημα μεταξύ 90% και 95%.

7.1.4 Σύνθεση Σχημάτων Προσώπου

Χάρη στη μέθοδο PCA κατασκευάστηκε ένα μοντέλο σχήματος που περιέχει όλη την πληροφορία μεταβλητότητας της βάσης εικόνων. Με αυτό το μοντέλο μπορούν να δημιουργηθούν νέα στιγμιότυπα σχημάτων προσώπου που μοιάζουν με αυτά της βάσης δεδομένων.

Σημειώνεται ότι στη συνέχεια θα χρησιμοποιούνται οι συμβολισμοί του Πίνακα 7.1 για το μοντέλο σχήματος. Με βάση τους συμβολισμούς αυτούς, η σύνθεση ενός νέου σχήματος προσώπου s_p γίνεται με τη σχέση

$$s_p = s_0 + \sum_{i=1}^n p_i s_i \quad (7.8)$$

δηλαδή γίνεται με παραμόρφωση του μέσου σχήματος s_0 χρησιμοποιώντας ένα γραμμικό συνδυασμό ιδιοδιανυσμάτων με κατάλληλα βάρη. Τα βάρη ουσιαστικά είναι οι παράμετροι σχήματος του διανύσματος p .

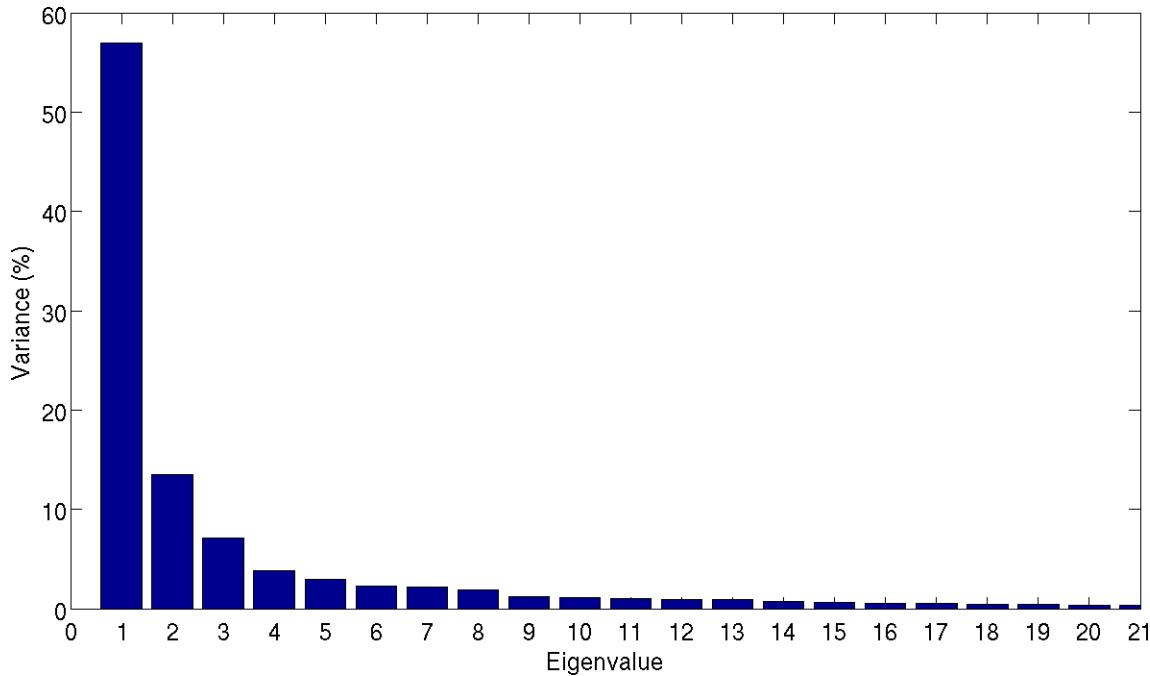
Iδιοδιανύσματα	s_i
Πλήθος ιδιοδιανυσμάτων (modes of variation)	n
Πλήθος landmark σημείων	L
Παράμετροι σχήματος	$p_{n \times 1}$
Στιγμιότυπο σχήματος προσώπου	s_p

Πίνακας 7.1: Συμβολισμοί μοντέλου σχήματος προσώπου

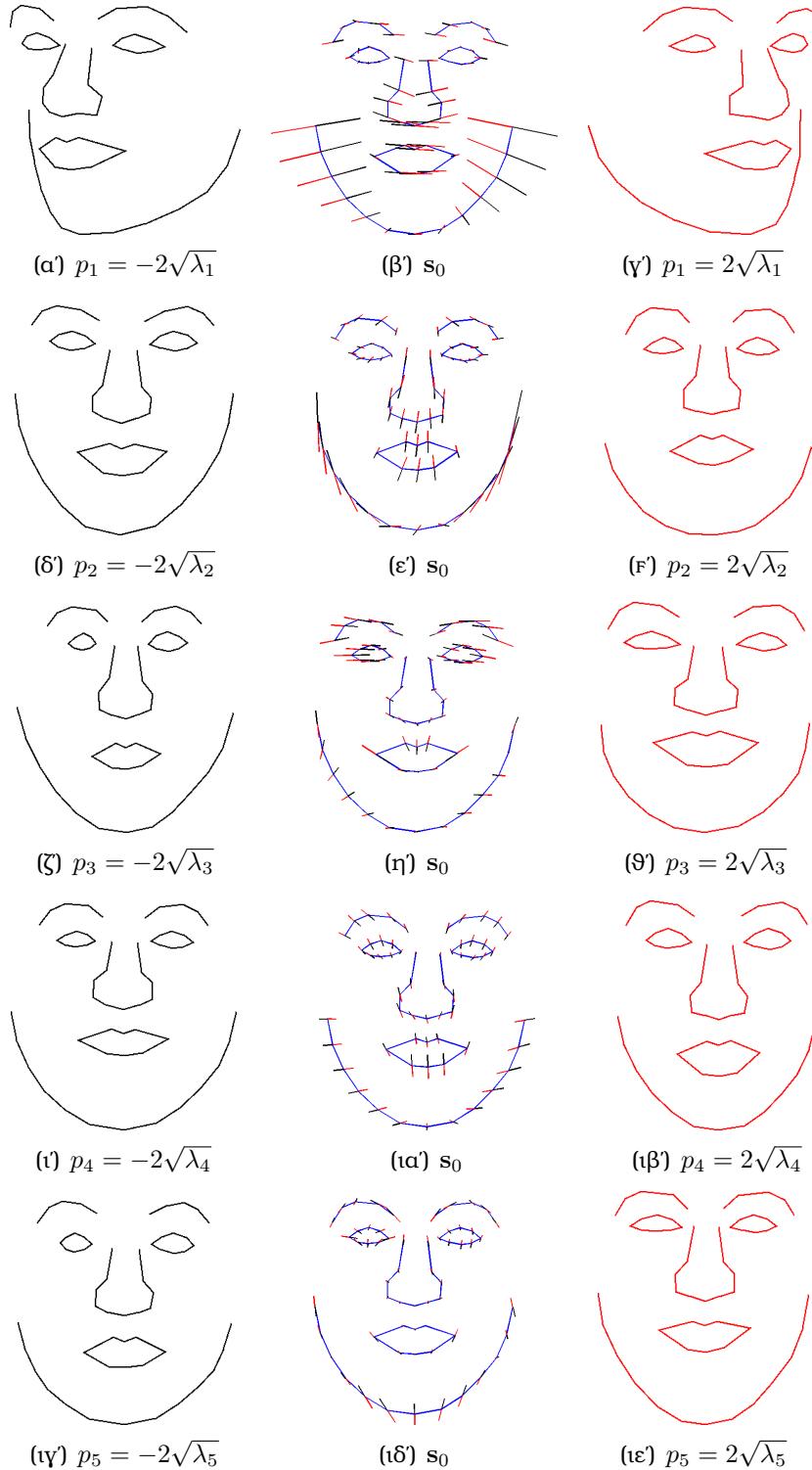
Ωστόσο, για να αποφύγουμε την υπερβολική παραμόρφωση του μέσου σχήματος και την αλλοίωση των χαρακτηριστικών του ανθρώπινου προσώπου πρέπει να επιλεγούν κατάλληλες παράμετροι p_i . Αν υποθέσουμε ότι οι παράμετροι που περιγράφουν

τα σχήματα των εικόνων εκπαίδευσης είναι ανεξάρτητες και ακολουθούν Gaussian κατανομή, τότε είναι ασφαλής η επιλογή παραμέτρων στο διάστημα $[-3\sqrt{\lambda_i}, 3\sqrt{\lambda_i}]$. Κρατώντας το 95% της συνολικής μεταβλητότητας των σχημάτων της βάσης δεδομένων προκύπτουν $n = 21$ άξονες μεταβλητότητας (modes of variation) και κατ' επέκταση 21 παράμετροι σχήματος p_i .

Στο Σχήμα 7.5 φαίνεται ένα ραβδόγραμμα που απεικονίζει το ποσοστό της συνολικής μεταβλητότητας, που αντιπροσωπεύει κάθε ιδιοτιμή και το αντίστοιχο ιδιοδιάνυσμά της s_i . Στο Σχήμα 7.6 φαίνονται οι παραμορφώσεις που προκαλούν στο μέσο σχήμα προσώπου τα πέντε πιο σημαντικά ιδιοδιανύσματα. Η μεσαία στήλη απεικονίζει το μέσο σχήμα προσώπου με βέλη που δείχνουν της κατεύθυνση μετακίνησης των landmark σημείων και οι αριστερή και δεξιά στήλες δείχνουν την παραμόρφωση που προκαλείται για $p_i = \pm 2\sqrt{\lambda_i}$, $i = 1, \dots, n$. Από τα δυο αυτά σχήματα γίνεται προφανές ότι το πρώτο ιδιοδιάνυσμα καταλαμβάνει το 57% της συνολικής μεταβλητότητας γιατί αντιπροσωπεύει την περιστροφή του κεφαλιού ως προς τον κατακόρυφο άξονα.



Σχήμα 7.5: Ποσοστό μεταβλητότητας που αντιστοιχεί σε κάθε ιδιοτιμή του μοντέλου σχήματος προσώπου



Σχήμα 7.6: Τα πρώτα πέντε πιο σημαντικά ιδιοδιανύσματα (modes of variation) του μοντέλου σχήματος προσώπου

7.2 Μοντέλο Υφής Προσώπου

Σε αυτή την παράγραφο περιγράφεται η κατασκευή ενός μοντέλου υφής προσώπου. Ουσιαστικά ο όρος υφή προσώπου αναφέρεται στο δέρμα που καλύπτει το κάθε ανθρώπινο πρόσωπο και τα πιθανά χαρακτηριστικά που διαθέτει (γένια, σκούρο/χλωμό, σημάδια κτλ.). Σε αντιστοιχία με τη διαδικασία ευθυγράμμισης των σχημάτων στην περίπτωση του μοντέλου σχήματος προσώπου, το μοντέλο υφής απαιτεί την ευθυγράμμιση όλων των δειγμάτων υφής σε ένα κοινό πλαίσιο αναφοράς. Επιλέγοντας σαν πλαίσιο αναφοράς το μέσο σχήμα προσώπου, οι υφές προσώπου της βάσης εικόνων χαρτογραφούνται σε αυτό με τέτοιο τρόπο ώστε τα landmark σημεία τους να ταυτίζονται. Συγκεκριμένα, εφαρμόζεται η μέθοδος τριγωνισμού Delaunay στα landmark σημεία του μέσου σχήματος s_0 ώστε να δημιουργηθούν τρίγωνα που θα επιτρέψουν την χαρτογράφηση των τιμών όλων των pixels μέσω βαρυκεντρικών συντεταγμένων (Piecewise Affine Warping). Στη συνέχεια, εφαρμόζεται η τεχνική της φωτομετρικής κανονικοποίησης (photometric normalization) σε όλα τα χαρτογραφημένα δείγματα υφής, ώστε με χρήση της Principal Component Analysis (PCA) να δημιουργηθεί το μοντέλο υφής.

7.2.1 Ορισμός Υφής Προσώπου

Η υφή του προσώπου καθορίζεται από τις τιμές έντασης όλων των σημείων (pixels) εσωτερικά του σχήματος προσώπου. Επομένως για N τέτοια pixels δειγματολειψίας, η υφή του προσώπου αναπαρίσταται από το διάνυσμα μήκους N

$$\mathbf{g} = [g_1, g_2, \dots, g_{N-1}, g_N]^T \quad (7.9)$$

Σε αυτό το σημείο πρέπει να ξεκαθαριστεί ότι στα επόμενα κεφάλαια της αναγνώρισης προσώπου και της αναγνώρισης των εκφράσεων προσώπου χρησιμοποιείται το μοντέλο υφής του προσώπου σε κλίμακα χρωμάτων του γκρι (grayscale) και όχι στο χρωματικό μοντέλο RGB. Βέβαια αυτή η επιλογή μπορεί εύκολα να αλλάξει, ωστόσο θεωρήθηκε περιττό το να διαχειρίζεται το τελικό σύστημα τον τριπλάσιο όγκο δεδομένων καθώς αν το μοντέλο ήταν σε RGB το μήκος των διανυσμάτων υφής g θα ήταν $3N$. Μάλιστα τονίζεται ότι το grayscale μοντέλο υφής εκπαιδεύτηκε μετατρέποντας όλες τις εικόνες της βάσης δεδομένων σε grayscale και όχι μετατρέποντας τη μέση υφή προσώπου και τα ιδιοδιανύσματα του RGB μοντέλου σε grayscale.

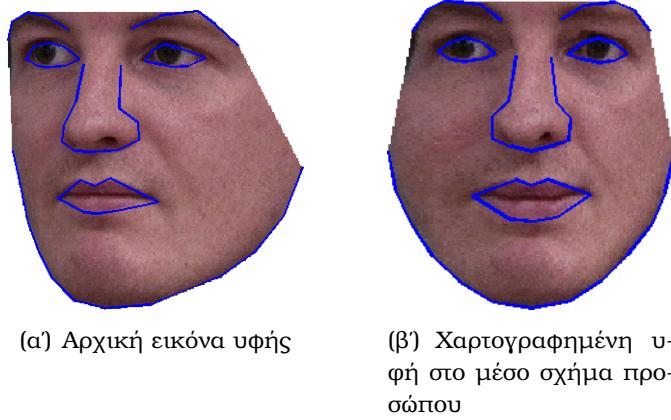
Απαραίτητο βήμα για τη δημιουργία του μοντέλου είναι η χαρτογράφηση όλων των εικόνων εκπαίδευσης στο μέσο σχήμα προσώπου (πλαίσιο αναφοράς) με τέτοιο τρόπο ώστε να υπάρχει αντιστοιχία μεταξύ των landmark σημείων του αρχικού σχήματος και του μέσου σχήματος. Η διαδικασία αυτή καταφέρνει να αφαιρέσει τις διαφορές μεταξύ των διανυσμάτων υφής λόγω των αλλαγών σχήματος, δημιουργώντας ένα κοινό πλαίσιο αναφοράς υφής.

7.2.2 Χαρτογράφηση Διανυσμάτων Υφής Προσώπου

Κατά τη διαδικασία της χαρτογράφησης αλλάζει η χωρική διάταξη των σημείων της αρχικής εικόνας ώστε να ταιριάζει με το σχήμα αναφοράς. Τυπικά αυτή η μετατροπή

από την αρχική στην τελική εικόνα συμβολίζεται με $\mathbf{I} \in \mathcal{R}^2 \rightarrow \mathbf{I}' \in \mathcal{R}^2$ όπου πρακτικά τα landmark σημεία της αρχικής εικόνας $\{\mathbf{x}_1, \dots, \mathbf{x}_L\}$ αντιστοιχούν στα landmark σημεία της τελικής εικόνας $\{\mathbf{x}'_1, \dots, \mathbf{x}'_L\}$.

Η χαρτογράφηση των pixels της εικόνας γίνεται ακολουθώντας αντίστροφη διαδικασία: αντιστοιχίζονται τα σημεία $\{\mathbf{x}'_1, \dots, \mathbf{x}'_L\}$ στα $\{\mathbf{x}_1, \dots, \mathbf{x}_L\}$ και κατόπιν αντιμετωπίζονται τα όποια προβλήματα με "κενά" που πιθανόν να προκύπτουν. Η διαδικασία αυτή γίνεται με τη μέθοδο Piecewise Affine Warp (συνάρτηση ομοπαραλληλικής στρέβλωσης με τμηματική αντιστοιχία). Συγκεκριμένα, χωρίζουμε το μέσο σχήμα προσώπου (convex hull) σε τρίγωνα με τη μέθοδο του Delaunay τριγωνισμού και βρίσκουμε την αντιστοιχία του κάθε σημείου της εικόνας χρησιμοποιώντας βαρυκεκτρικές συντεταγμένες. Το αποτέλεσμα της χαρτογράφησης της υφής μιας εικόνας φαίνεται στο Σχήμα 7.7.

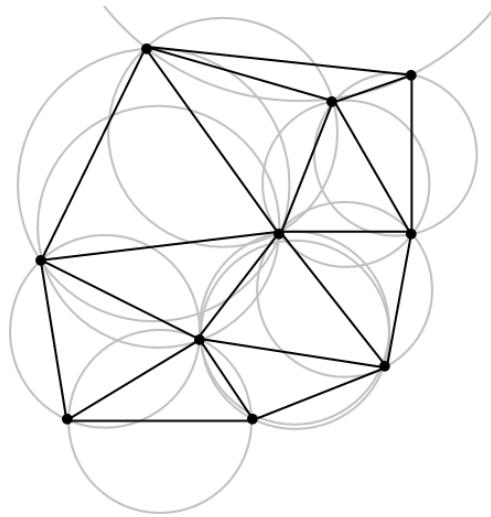


Σχήμα 7.7: Παράδειγμα χαρτογράφησης εικόνας υφής στο μέσο σχήμα προσώπου

7.2.2.1 Delaunay Τριγωνισμός

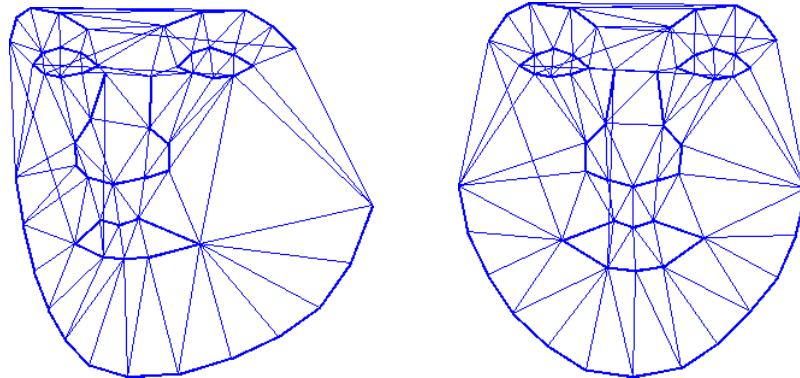
Στην υπολογιστική γεωμετρία ο *Delaunay τριγωνισμός* (Boris Delaunay, 1934) για ένα σύνολο σημείων P στο επίπεδο είναι ο τριγωνισμός $DT(P)$ έτσι ώστε κανένα σημείο του συνόλου P να μην ανήκει εντός του περιγεγραμμένου κύκλου οποιουδήποτε τριγώνου του $DT(P)$. Ο περιγεγραμμένος κύκλος ενός τριγώνου ορίζεται ως ο κύκλος που περνά και από τις τρεις κορυφές του τριγώνου. Ένα παράδειγμα δέκα τυχαίων σημείων φαίνεται στο Σχήμα 7.8, όπου απεικονίζεται το αποτέλεσμα του τριγωνισμού μαζί με τους περιγεγραμμένους κύκλους. Ο Delaunay τριγωνισμός μεγιστοποιεί την ελάχιστη γωνία όλων των γωνιών των τριγώνων του τριγωνισμού και τείνει να αποφεύγει τη δημιουργία τριγώνων με μικρό εμβαδόν.

Για ένα κοίλο σύνολο σημείων ο Delaunay τριγωνισμός δημιουργεί τρίγωνα προσθέτοντας σημεία που δεν ανήκουν στο αρχικό σύνολο σε κατάλληλες θέσεις. Σε αυτές τις περιπτώσεις χρησιμοποιείται Delaunay τριγωνισμός με περιορισμούς ή ο αλγόριθμος Thin Plate Splines. Στο Σχήμα 7.9 φαίνεται το αποτέλεσμα του Delaunay τριγωνισμού σε ένα αρχικό σχήμα προσώπου και στο μέσο σχήμα προσώπου.



Σχήμα 7.8: Τριγωνισμός Delaunay

Ειδικά ο τριγωνισμός του Σχήματος 7.9(β) είναι πολύ σημαντικός καθώς χρησιμοποιείται κατά κόρον στη μέθοδο του Piecewise Affine Warp ως πλαίσιο αναφοράς.



(α') Τριγωνισμός Delaunay στο σχήμα προσώπου της εικόνας 7.7(a)

(β') Τριγωνισμός Delaunay στο μέσο σχήμα προσώπου

Σχήμα 7.9: Τριγωνισμός Delaunay σε σχήμα προσώπου

7.2.2.2 Piecewise Affine Warp

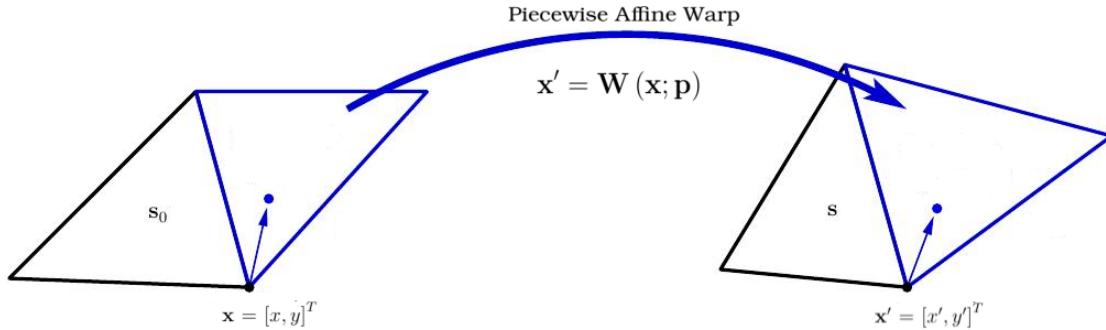
Με τον όρο Piecewise Affine Warp ορίζουμε μια συνάρτηση που δοθέντος ενός σημείου $\mathbf{x} = [x, y]^T$ στο εσωτερικό ενός σχήματος βρίσκει τις συντεταγμένες $\mathbf{x}' = [x', y']^T$ του αντίστοιχου σημείου στο εσωτερικό ενός σχήματος προορισμού. Είναι μια συνάρτηση ομοπαραλληλικής στρέβλωσης με αντιστοιχία σημείων που προϋποθέτει ότι ο αριθμός

των landmark σημείων που ελέγχουν το αρχικό σχήμα και το σχήμα προορισμού είναι ίδιος.

Ακολουθώντας τους συμβολισμούς που χρησιμοποιούνται στα [PaMa08, BGM04, MaBa04] από μαθηματικής άποψης η συνάρτηση Piecewise Affine Warp ορίζεται ως

$$\mathbf{x}' = \mathbf{W}(\mathbf{x}; \mathbf{p}) \quad (7.10)$$

όπου \mathbf{x} οι συντεταγμένες του σημείου στο αρχικό σχήμα, \mathbf{x}' οι συντεταγμένες του σημείου στο σχήμα προορισμού και \mathbf{p} οι παράμετροι σχήματος που μας δίνουν το σχήμα προορισμού σύμφωνα με τη σχέση (7.8). Η διαδικασία αυτή απεικονίζεται στο Σχήμα 7.10.



Σχήμα 7.10: Συνάρτηση piecewise affine warp

Ο υπολογισμός των συντεταγμένων στο σχήμα προορισμού γίνεται με χρήση των βαρυκεντρικών συντεταγμένων. Οι βαρυκεντρικές συντεταγμένες είναι ένα είδος ομογενών συντεταγμένων όπου η θέση ενός σημείου καθορίζεται από το κέντρο μάζας (ή κέντρο βάρους) των κορυφών ενός τριγώνου. Στη συνέχεια παρουσιάζονται δύο ορισμοί της συνάρτησης Piecewise Affine Warp.

Πρώτος ορισμός συνάρτησης \mathbf{W} Ισχύει η σχέση

$$\mathbf{x}' = \mathbf{W}(\mathbf{x}; \mathbf{p}) = \alpha \mathbf{x}'_1 + \beta \mathbf{x}'_2 + \gamma \mathbf{x}'_3 = \begin{bmatrix} \alpha x'_1 + \beta x'_2 + \gamma x'_3 \\ \alpha y'_1 + \beta y'_2 + \gamma y'_3 \end{bmatrix} \quad (7.11)$$

όπου $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ είναι οι κορυφές του τριγώνου του αρχικού σχήματος, $\mathbf{x}'_1, \mathbf{x}'_2, \mathbf{x}'_3$ είναι οι κορυφές του τριγώνου του σχήματος προορισμού, \mathbf{x} είναι το σημείο που ανήκει στο τρίγωνο ($\mathbf{x}_1 \mathbf{x}_2 \mathbf{x}_3$) και \mathbf{x}' είναι το σημείο προορισμού του \mathbf{x} που ανήκει στο τρίγωνο ($\mathbf{x}'_1 \mathbf{x}'_2 \mathbf{x}'_3$). Οι βαρυκεντρικές συντεταγμένες της σχέσης (7.11) υπολογίζονται από τις σχέσεις

$$\begin{aligned} d &= (x_2 - x_1)(y_3 - y_1) - (x_3 - x_1)(y_2 - y_1) = \\ &= x_1(y_2 - y_3) + x_2(y_3 - y_1) + x_3(y_1 - y_2) \\ \alpha &= 1 - \beta - \gamma \\ \beta &= \frac{(x_1 - x_3)(y - y_3) - (x - x_3)(y_1 - y_3)}{d} \\ \gamma &= \frac{(x_2 - x_1)(y - y_1) - (x - x_1)(y_2 - y_1)}{d} \end{aligned} \quad (7.12)$$

Επομένως οι συντεταγμένες α, β, γ εξαρτώνται από τις κορυφές του αρχικού τριγώνου $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ και από το αρχικό σημείο \mathbf{x} . Για τις βαρυκεντρικές συντεταγμένες σε αντιστοιχία με τη σχέση (7.11) ισχύει και

$$\mathbf{x} = \alpha \mathbf{x}_1 + \beta \mathbf{x}_2 + \gamma \mathbf{x}_3 = \begin{bmatrix} \alpha x_1 + \beta x_2 + \gamma x_3 \\ \alpha y_1 + \beta y_2 + \gamma y_3 \end{bmatrix} \quad (7.13)$$

Δεύτερος ορισμός συνάρτησης \mathbf{W} Ισχύει η σχέση

$$\mathbf{x}' = \mathbf{W}(\mathbf{x}; \mathbf{p}) = \begin{bmatrix} \alpha_1 + \alpha_2 x + \alpha_3 y \\ \alpha_4 + \alpha_5 x + \alpha_6 y \end{bmatrix} \quad (7.14)$$

όπου οι παράμετροι α_i εξαρτώνται μόνο από τις κορυφές των δυο τριγώνων $\alpha_i = \alpha_i(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}'_1, \mathbf{x}'_2, \mathbf{x}'_3)$, $\forall i = 1, \dots, 6$. Συγκεκριμένα οι παράμετροι δίνονται από τις σχέσεις

$$\begin{aligned} d &= (x_2 - x_1)(y_3 - y_1) - (x_3 - x_1)(y_2 - y_1) = \\ &= x_1(y_2 - y_3) + x_2(y_3 - y_1) + x_3(y_1 - y_2) \\ \alpha_1 &= \frac{x'_1(x_2y_3 - x_3y_2) + x'_2(x_3y_1 - x_1y_3) + x'_3(x_1y_2 - x_2y_1)}{d} \\ \alpha_2 &= \frac{x'_1(y_2 - y_3) + x'_2(y_3 - y_1) + x'_3(y_1 - y_2)}{d} \\ \alpha_3 &= \frac{x'_1(x_3 - x_2) + x'_2(x_1 - x_3) + x'_3(x_2 - x_1)}{d} \\ \alpha_4 &= \frac{y'_1(x_2y_3 - x_3y_2) + y'_2(x_3y_1 - x_1y_3) + y'_3(x_1y_2 - x_2y_1)}{d} \\ \alpha_5 &= \frac{y'_1(y_2 - y_3) + y'_2(y_3 - y_1) + y'_3(y_1 - y_2)}{d} \\ \alpha_6 &= \frac{y'_1(x_3 - x_2) + y'_2(x_1 - x_3) + y'_3(x_2 - x_1)}{d} \end{aligned} \quad (7.15)$$

Οι δυο ορισμοί της συνάρτησης $\mathbf{W}(\mathbf{x}; \mathbf{p})$ που δόθηκαν με τις εξισώσεις (7.11) και (7.14) είναι ισοδύναμοι. Η επιλογή ενός από τους δυο τρόπους εξαρτάται από τον σκοπό της χρήσης της συνάρτησης με κριτήριο τη μείωση της υπολογιστικής πολυπλοκότητας, καθώς οι παράμετροι της καθεμιάς εξαρτώνται από διαφορετικά πράγματα.

Για τις βαρυκεντρικές συντεταγμένες των εξισώσεων (7.12) ισχύει η ιδότητα $\alpha + \beta + \gamma = 1$. Ένα σημείο \mathbf{x} ανήκει σε ένα τρίγωνο αν για τις βαρυκεντρικές συντεταγμένες του τριγώνου αυτού ισχύει η σχέση

$$0 \leq \alpha, \beta, \gamma \leq 1 \quad (7.16)$$

Ο ψευδοκώδικας της συνάρτησης Piecewise Affine Warp φαίνεται στον αλγόριθμο 6. Για να υλοποιηθεί το βήμα 6 του αλγορίθμου 6 θα πρέπει να υπάρχει πλήρης

Αλγόριθμος 6 Συνάρτηση Piecewise Affine Warp

- 1: Έστω $\{\mathbf{x}_1, \dots, \mathbf{x}_L\}$ τα landmark σημεία του αρχικού σχήματος
 - 2: Έστω $\{\mathbf{x}'_1, \dots, \mathbf{x}'_L\}$ τα landmark σημεία του σχήματος προορισμού
 - 3: Εφαρμόζουμε Delaunay τριγωνισμό στο σχήμα $\{\mathbf{x}_1, \dots, \mathbf{x}_L\}$
 - 4: **for** όλα τα σημεία $\mathbf{x} = [x, y]^T$ που ανήκουν στο σχήμα $\{\mathbf{x}_1, \dots, \mathbf{x}_L\}$ **do**
 - 5: Βρες το τρίγωνο t στο οποίο ανήκει το σημείο \mathbf{x} από τη σχέση (7.16)
 $0 \leq \alpha_t, \beta_t, \gamma_t \leq 1 \Rightarrow \mathbf{x} \in t$
 - 6: Βρες το αντίστοιχο τρίγωνο t' του σχήματος προορισμού $\{\mathbf{x}'_1, \dots, \mathbf{x}'_L\}$ θεωρώντας ότι οι θέσεις των landmark σημείων των δυο σχημάτων παρουσιάζουν αντίστοιχια
 - 7: Υπολόγισε τις βαρυκεντρικές συντεταγμένες με τις σχέσεις (7.12) ή (7.15)
 - 8: Υπολόγισε τη θέση του \mathbf{x}' μέσα στο t' από τη σχέση (7.11) ή (7.14)
 - 9: **end for**
-

αντίστοιχία μεταξύ των θέσεων των landmark σημείων όλων των σχημάτων προσώπου. Πρακτικά αυτό σημαίνει ότι στα διανύσματα των σχημάτων τα σημεία πρέπει να έχουν τις σωστές θέσεις και να ακολουθούν τη μέθοδο αρίθμησης του Σχήματος 7.3(γ'). Έτσι, το τρίγωνο t που έχει βρεθεί στο αρχικό σχήμα εισόδου είναι σίγουρο ότι καλύπτει την αντίστοιχη περιοχή με αυτή του σχήματος t' στο μέσο σχήμα προσώπου.

7.2.2.3 Δειγματοληψία Εικόνας με Διγραμμική Παρεμβολή

Έχοντας τις αντίστοιχίες όλων των σημείων στο εσωτερικό μεταξύ δυο σχημάτων προσώπου, η χαρτογράφηση των διανυσμάτων υφής προσώπου γίνεται με δειγματοληψία (αντιγραφή) της έντασης των σημείων από την αρχική εικόνα της βάσης δεδομένων στα σημεία του σχήματος του μέσου προσώπου. Αν I είναι η αρχική εικόνα και I' η χαρτογραφημένη εικόνα, τότε η διαδικασία της δειγματοληψίας τυπικά εκφράζεται από τη σχέση

$$I'(\mathbf{x}') = I(\mathbf{x}) \Rightarrow I'(\mathbf{W}(\mathbf{x}; \mathbf{p})) = I(\mathbf{x})$$

Πρακτικά, όμως, η δειγματοληψίας γίνεται με αντίστροφο τρόπο

$$I'(\mathbf{x}) = I(\mathbf{x}') \Rightarrow I'(\mathbf{x}) = I(\mathbf{W}(\mathbf{x}; \mathbf{p})) \quad (7.17)$$

δηλαδή η συνάρτηση \mathbf{W} εφαρμόζεται με αρχικό σχήμα το μέσο σχήμα προσώπου και σχήμα προορισμού το σχήμα της εικόνας της βάσης δεδομένων.

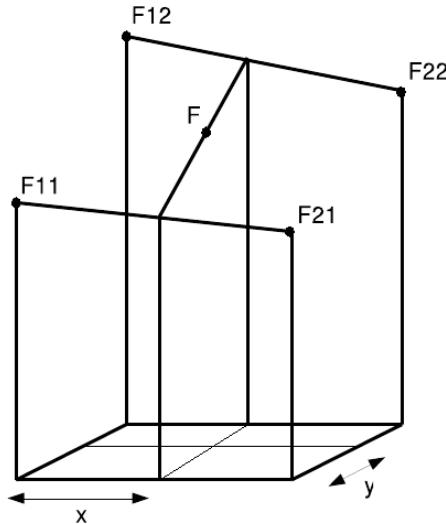
Για να αποφευχθούν τυχόν κενά στην εικόνα υφής που δημιουργείται, χρησιμοποιείται η τεχνική της διγραμμικής παρεμβολής (bilinear interpolation correction).

Συγκεκριμένα για μια συνάρτηση F , η τιμή της παρεμβολής στο σημείο (x, y) δίνεται από τη σχέση

$$F(x, y) \approx F_{11}(1-x)(1-y) + F_{21}x(1-y) + F_{12}(1-x)y + F_{22}xy \quad (7.18)$$

και σε μορφή πινάκων

$$F(x, y) \approx \begin{bmatrix} 1-x & x \end{bmatrix} \begin{bmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{bmatrix} \begin{bmatrix} 1-y \\ y \end{bmatrix} \quad (7.19)$$



Σχήμα 7.11: Διγραμμική παρεμβολή

όπου F_{11} , F_{12} , F_{21} και F_{22} οι γείτονες του σημείου F όπως φαίνεται στο Σχήμα 7.11.

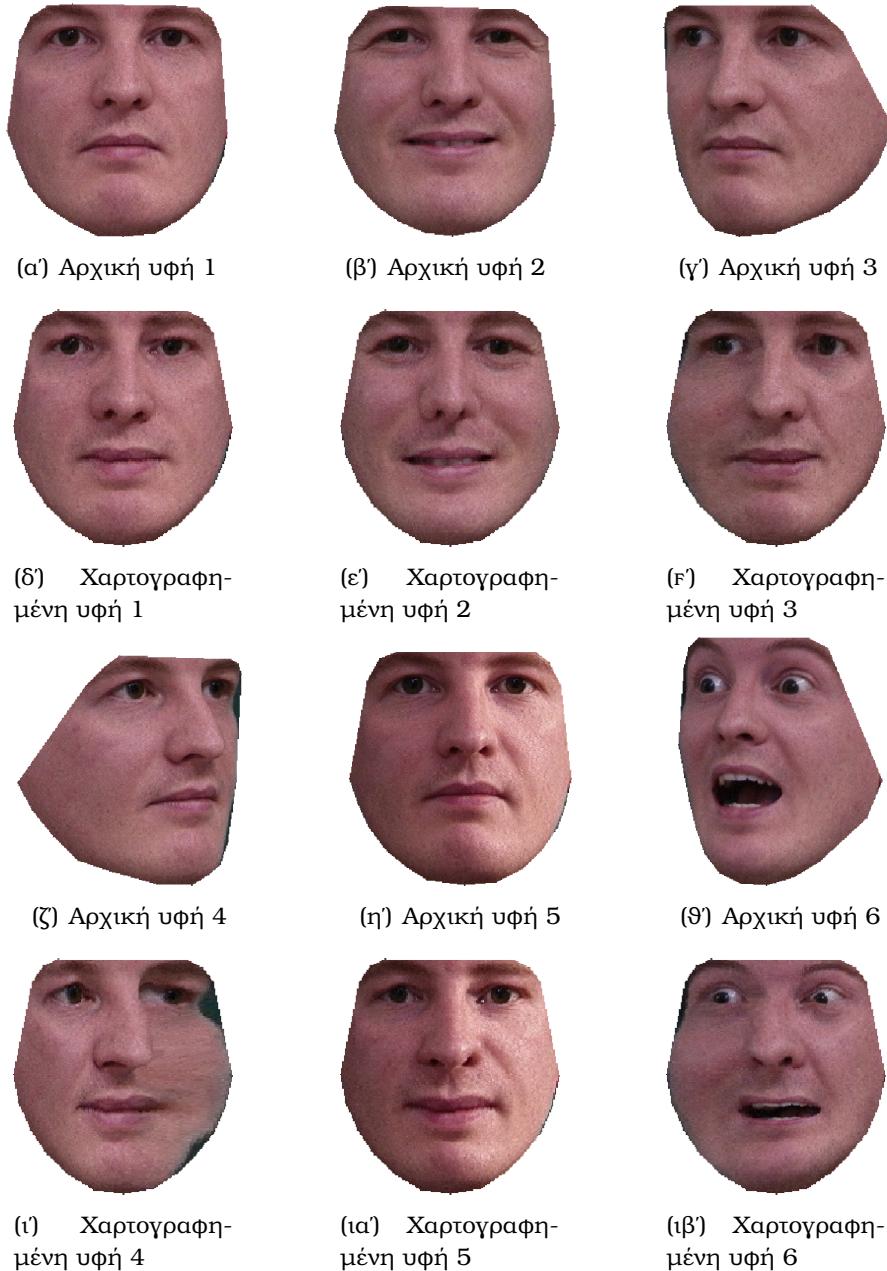
Στον αλγόριθμο 7 φαίνεται η τελική διαδικασία που ακολουθείται για τη χαρτογράφηση των διανυμάτων υφής της βάσης δεδομένων στο μέσο σχήμα προσώπου.

Αλγόριθμος 7 Χαρτογράφηση διανυμάτων υφής στο μέσο σχήμα προσώπου

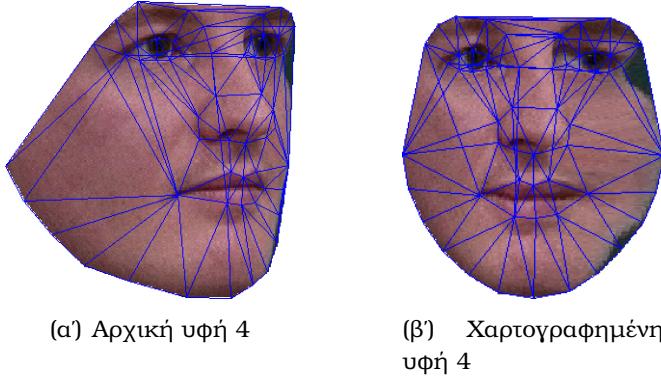
- 1: Έστω s_0 το μέσο σχήμα προσώπου
 - 2: Εφαρμόζουμε Delaunay τριγωνισμό στο μέσο σχήμα προσώπου s_0
 - 3: **for** όλες τις εικόνες της βάσης δεδομένων $i = 1, \dots, M$ **do**
 - 4: Έστω s_i το σχήμα προσώπου της επιλεγμένης εικόνας I_i
 - 5: **for** όλα τα σημεία $x = [x, y]^T$ που ανήκουν στο μέσο σχήμα προσώπου s_0 **do**
 - 6: Βρες το τρίγωνο t στο οποίο ανήκει το σημείο x από τη σχέση (7.16)
 $0 \leq \alpha_t, \beta_t, \gamma_t \leq 1 \Rightarrow x \in t$
 - 7: Βρες το αντίστοιχο τρίγωνο t' του σχήματος προορισμού s_i θεωρώντας ότι οι θέσεις των landmark σημείων των δυο σχημάτων παρουσιάζουν αντιστοιχία
 - 8: Υπολόγισε τις βαρυκεντρικές συντεταγμένες με τις σχέσεις (7.12) ή (7.15)
 - 9: Υπολόγισε τη θέση του x' μέσα στο t' από τη σχέση (7.11) ή (7.14)
 - 10: Αντέγραψε την τιμή έντασης της αρχικής εικόνας στην εικόνα μέσου προσώπου σύμφωνα με τη σχέση (7.17): $I'_i(x) = I_i(x')$
 - 11: **end for**
 - 12: **end for**
-

Στο Σχήμα 7.12 φαίνεται το αποτέλεσμα της χαρτογράφησης (piecewise affine warp και δειγματοληψία) των έξι εικόνων ενός ανθρώπου από τη βάση δεδομένων IMM. Οι εικόνες παρουσιάζουν διαφορετική πόζα και εκφράσεις. Παρατηρούμε ότι η χαρτογράφηση στις εικόνες 1, 2 και 5, όπου η πόζα είναι πλήρως μετωπική είναι σωστή, καθώς η αντιστοιχία των τριγώνων Delaunay είναι απόλυτη. Ωστόσο, σε

εικόνες όπου το πρόσωπο είναι στραμμένο δεξιά ή αριστερά, με αντιπροσωπευτικό παράδειγμα την εικόνα 4, η χαρτογράφηση αποτυχάνει. Οι πιθανοί λόγοι είναι είτε η μη ακριβής τοποθέτηση των landmark σημείων με αποτέλεσμα στο αρχικό διάνυσμα υφής να υπάρχουν pixels από το φόντο της εικόνας είτε η αποτυχία του Delaunay τριγώνισμού που τοποθετεί πρόσθετα σημεία ώστε να αποφύγει τη δημιουργία τριγώνων πολύ μικρού εμβαδού. Η εικόνα 4 με τον τριγωνισμό της φαίνεται και στο Σχήμα 7.13.



Σχήμα 7.12: Παραδείγματα χαρτογράφησης υφής με διάφορες πόζες και εκφράσεις



Σχήμα 7.13: Παράδειγμα αποτυχημένης χαρτογράφησης με Delaunay τριγωνισμό

7.2.3 Φωτομετρική Ομαλοποίηση

Όπως κατά τη δημιουργία του μοντέλου σχήματος προσώπου αφαιρέθηκε από όλα τα σχήματα προσώπου η πόζα λόγω των παραμέτρων του μετασχηματισμού ομοιότητας, κατά αντίστοιχο τρόπο και για τη δημιουργία του μοντέλου υφής προσώπου πρέπει να αφαιρεθούν οι επιδράσεις του φωτισμού. Οι διαφορές στον φωτισμό μεταξύ των εικόνων μπορεί να οφείλονται στη χρήση διαφορετικής φωτογραφικής μηχανής, διαφορετικού φιλμ, στις καιρικές συνθήκες και στην ώρα της ημέρας που τραβήχθηκαν οι εικόνες. Η τεχνική που χρησιμοποιείται είναι της φωτομετρικής ομαλοποίησης (photometric normalization) όπως αναλύεται στο [CETO1].

Αν \mathbf{g} είναι το αρχικό διάνυσμα υφής, τότε το ομαλοποιημένο χωρίς τις επιδράσεις φωτεινότητας δίνεται από τη σχέση

$$\mathbf{g}_{norm} = \frac{\mathbf{g} - \beta \mathbf{1}}{\alpha} \quad (7.20)$$

όπου $\mathbf{1}$ είναι το μοναδιαίο διάνυσμα $\mathbf{1} = [1, 1, \dots, 1, 1]^T$. Ουσιαστικά κάθε διάνυσμα υφής \mathbf{g} μάκους N ευθυγραμμίζεται με το μέσο διάνυσμα υφής προσώπου \mathbf{g}_0 , αντισταθμίζοντάς το ώστε να έχει μηδενική μέση τιμή

$$\mathbf{g}_{zm} = \mathbf{g}_0 - \hat{g} \mathbf{1}, \text{ με } \hat{g} = \frac{1}{N} \sum_{i=1}^N g_i = \frac{1}{N} (\mathbf{g} \cdot \mathbf{1})$$

και εφαρμόζοντας κατάλληλη κλίμακα ώστε να έχει μοναδιαία διακύμανση

$$\mathbf{g} = \frac{1}{\sigma} \mathbf{g}_{zm}, \text{ με } \sigma^2 = \frac{1}{N} \sum_{i=1}^N (g_i - \hat{g})^2$$

Επομένως, οι παράμετροι α και β της εξίσωσης (7.20) υπολογίζονται ως

$$\alpha = \mathbf{g} \cdot \mathbf{g}_0 \text{ και } \beta = \frac{\mathbf{g} \cdot \mathbf{1}}{N} \quad (7.21)$$

Αλγόριθμος 8 Φωτομετρική Ομαλοποίηση

- 1: Επιλογή του πρώτου διανύσματος υφής σαν την αρχική εκτίμηση της μέσης υφής:
 $\bar{g}_0 = g_1$
 - 2: $k = 0$
 - 3: **repeat**
 - 4: **for** $i = 1, \dots, M$ **do**
 - 5: Υπολογισμός $\alpha = g_i \cdot \bar{g}$ και $\beta = \frac{g_i \cdot 1}{N}$
 - 6: Ομαλοποίηση της υφής g_i με την εξίσωση 7.20
 - 7: **end for**
 - 8: $k = k + 1$
 - 9: Νέα εκτίμηση της μέσης υφής: $\bar{g}_k = \frac{1}{M} \sum_{i=1}^M g_i$
 - 10: **until** σύγκλιση μέσης υφής $\bar{g}_k \approx \bar{g}_{k-1}$
 - 11: Τελική εκτίμηση μέσης υφής προσώπου: $g_0 = \bar{g}_k$
-

Στον αλγόριθμο 8 φαίνονται τα βήματα που απαιτούνται για τη φωτομετρική ομαλοποίηση των M εικόνων της βάσης δεδομένων και τον τελικό υπολογισμό της μέσης υφής προσώπου g_0 . Η διαδικασία της ομαλοποίησης ουσιαστικά μετασχηματίζει τα διανύσματα υφής ώστε να ισχύει $g_{norm} \cdot 1 = 0$ και $|g_{norm}| = 1$.

Στην περίπτωση που στο μοντέλο υφής προσώπου χρησιμοποιούσαμε το RGB χρωματικό σύστημα και όχι grayscale εικόνες, η καλύτερη μέθοδος ομαλοποίησης θα ήταν η μέθοδος ισοστάθμισης ιστογράμματος (histogram equalization) ανεξάρτητα σε κάθε χρωματικό κανάλι.

7.2.4 Principal Component Analysis (PCA) Χαμηλής Μνήμης

Στη μέθοδο Principal Component Analysis, όπως αναφέρθηκε και στην παράγραφο 7.1.3, όταν ο αριθμός των διαστάσεων N είναι μεγαλύτερος από τον αριθμό των δειγμάτων M , τότε οδηγούμαστε στον υπολογισμό ενός πίνακα συνδιακύμανσης τεραστίων διαστάσεων ($N \times N$). Ειδικά στην περίπτωση μας ο αριθμός των εικόνων είναι $M = 240$ και ο αριθμός των διαστάσεων (αριθμός pixels μέσα στο μέσο σχήμα προσώπου) είναι $N = 22477$. Ωστόσο, είναι δυνατόν να εκτελεστεί ανάλυση ιδιοδιανυσμάτων χρησιμοποιώντας έναν πολύ μικρότερο πίνακα μεγέθους $M \times M$, όπως περιγράφεται στο [CET01].

Αρχικά δημιουργούμε τον πίνακα G που στις στήλες του έχει τη διαφορά του κάθε διανύσματος υφής της βάσης δεδομένων με τη μέση υφή

$$G = [g_1 - g_0 \ g_2 - g_0 \ \dots \ g_M - g_0]$$

Ο πίνακας G έχει διαστάσεις $N \times M$. Από αυτόν προκύπτει ο πίνακας συνδιακύμανσης

$$C_{N \times N} = \frac{1}{M-1} G G^T \quad (7.22)$$

Ωστόσο, για να αποφευχθεί το τεράστιο μέγεθος του, θεωρείται ο πίνακας

$$C'_{M \times M} = \frac{1}{M-1} G^T G \quad (7.23)$$

Σύμφωνα με το Θεώρημα Eckart-Young μπορεί να αποδειχθεί ότι οι μη-μηδενικές ιδιοτιμές του πίνακα C' είναι ίδιες με τις ιδιοτιμές του πίνακα C

$$\Delta_C = \Delta_{C'}$$

και για τα ιδιοδιανύσματά τους ισχύει

$$\Phi_C = G\Phi_{C'}$$

Με άλλα λόγια, αν ϕ_i και δ_i είναι αντίστοιχα τα ιδιοδιανύσματα και οι ιδιοτιμές του πίνακα C' με $i = 1, \dots, M$ και $\delta_i > \delta_{i+1}$, τότε μπορεί να αποδειχθεί ότι τα πρώτα M ιδιοδιανύσματα του $G\phi_i$ είναι ιδιοδιανύσματα του C με αντίστοιχες ιδιοτιμές δ_i . Τα υπόλοιπα $N - M$ ιδιοδιανύσματα του C έχουν όλα μηδενικές ιδιοτιμές. Τονίζεται ότι ο $G\phi_i$ δεν έχει απαραίτητα κανονικοποιημένες στήλες.

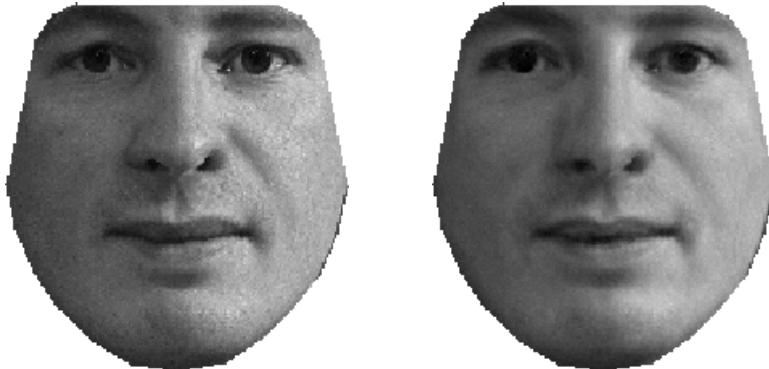
Κρατώντας τις m πιο σημαντικές ιδιοτιμές και τα αντίστοιχα ιδιοδιανύσματα, ο-ποιοδήποτε διάνυσμα υφής προσώπου από τη βάση εικόνων μπορεί να προσεγγιστεί με μια παραμόρφωση της μέσης υφής προσώπου με έναν γραμμικό συνδυασμό των m ιδιοδιανυσμάτων

$$g \approx g_0 + \Phi \lambda \quad (7.24)$$

όπου $\Phi = (\phi_1 | \phi_2 | \dots | \phi_m)$ είναι ορθογώνιος πίνακας και λ είναι ένα διάνυσμα μήκους m που μπορεί να υπολογιστεί ως

$$\lambda = \Phi^{-1} (g - g_0) = \Phi^T (g - g_0) \quad (7.25)$$

Ένα τέτοιο παράδειγμα φαίνεται στο Σχήμα 7.14.



(a') Υφή προσώπου g από τη βάση δεδομένων

(b') Προσέγγιση αρχικής υφής προσώπου με παραμόρφωση της μέσης υφής $g_0 + \Phi \lambda$ με τους παράγοντες λ

Σχήμα 7.14: Προσέγγιση υφής προσώπου ορισμένη από τη βάση δεδομένων με παραμόρφωση της μέσης υφής με ένα γραμμικό συνδυασμό ιδιοδιανυσμάτων του μοντέλου υφής

7.2.5 Σύνθεση Υφής Προσώπου

Χάρη στη μέθοδο PCA κατασκευάστηκε ένα μοντέλο υφής που περιέχει όλη την πληροφορία μεταβλητότητας της βάσης εικόνων. Με αυτό το μοντέλο μπορούν να δημιουργηθούν νέα σπιγμιότυπα υφής προσώπου που μοιάζουν με αυτά της βάσης δεδομένων.

Σημειώνεται ότι στη συνέχεια θα χρησιμοποιούνται οι συμβολισμοί του Πίνακα 7.2 για το μοντέλο υφής. Με βάση τους συμβολισμούς αυτούς, η σύνθεση ενός νέου διανύσματος υφής προσώπου A_{λ} γίνεται με τη σχέση

$$A_{\lambda} = A_0 + \sum_{i=1}^m \lambda_i A_i \quad (7.26)$$

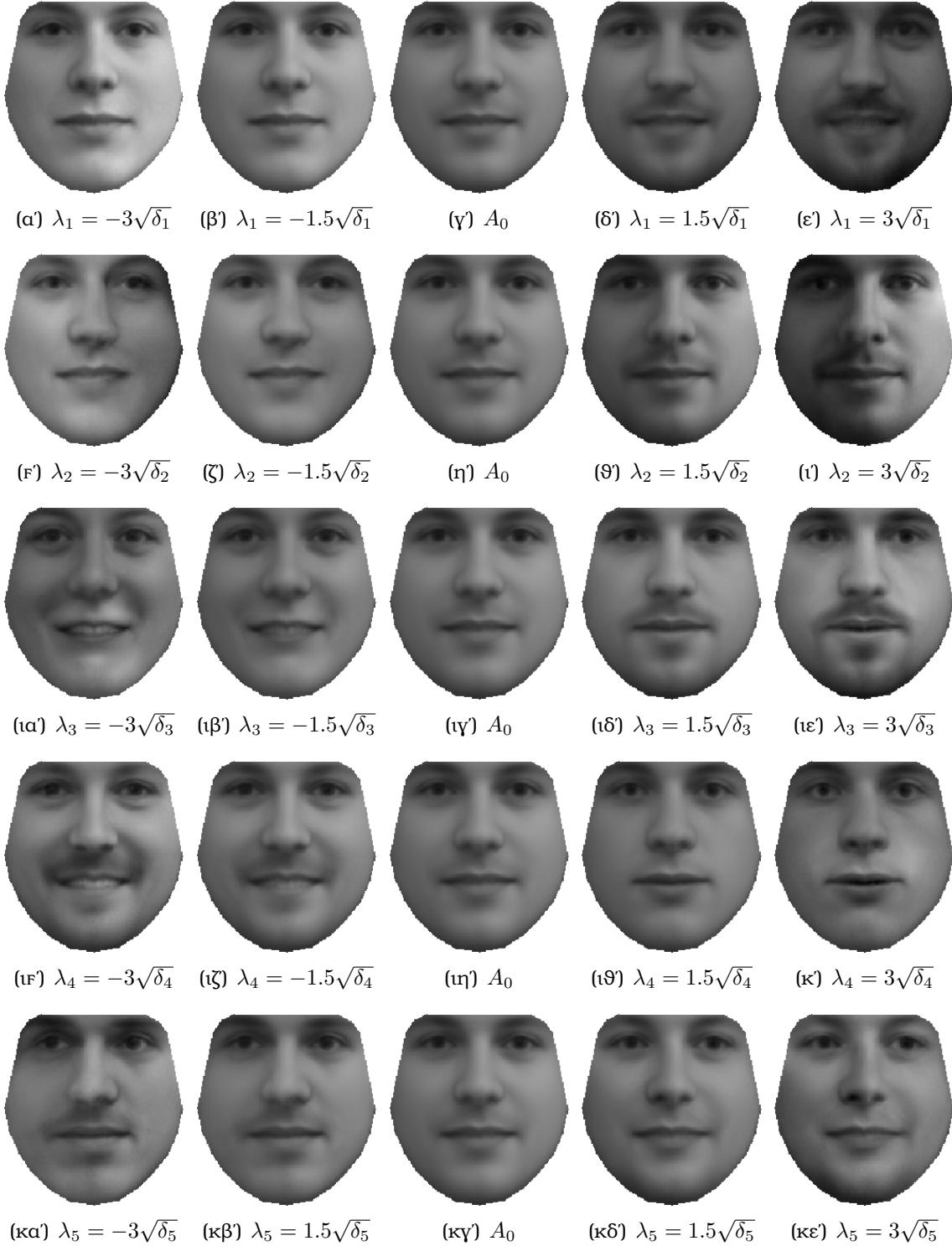
δηλαδή γίνεται με παραμόρφωση της μέσης υφής A_0 χρησιμοποιώντας ένα γραμμικό συνδυασμό ιδιοδιανυσμάτων με κατάλληλα βάρη. Τα βάρη ουσιαστικά είναι οι παράμετροι υφής του διανύσματος λ .

Ιδιοδιανύσματα	A_i
Πλήθος ιδιοδιανυσμάτων (modes of variation)	m
Πλήθος pixels μέσου σχήματος	N
Παράμετροι υφής	$\lambda_{m \times 1}$
Σπιγμιότυπο υφής προσώπου	A_{λ}

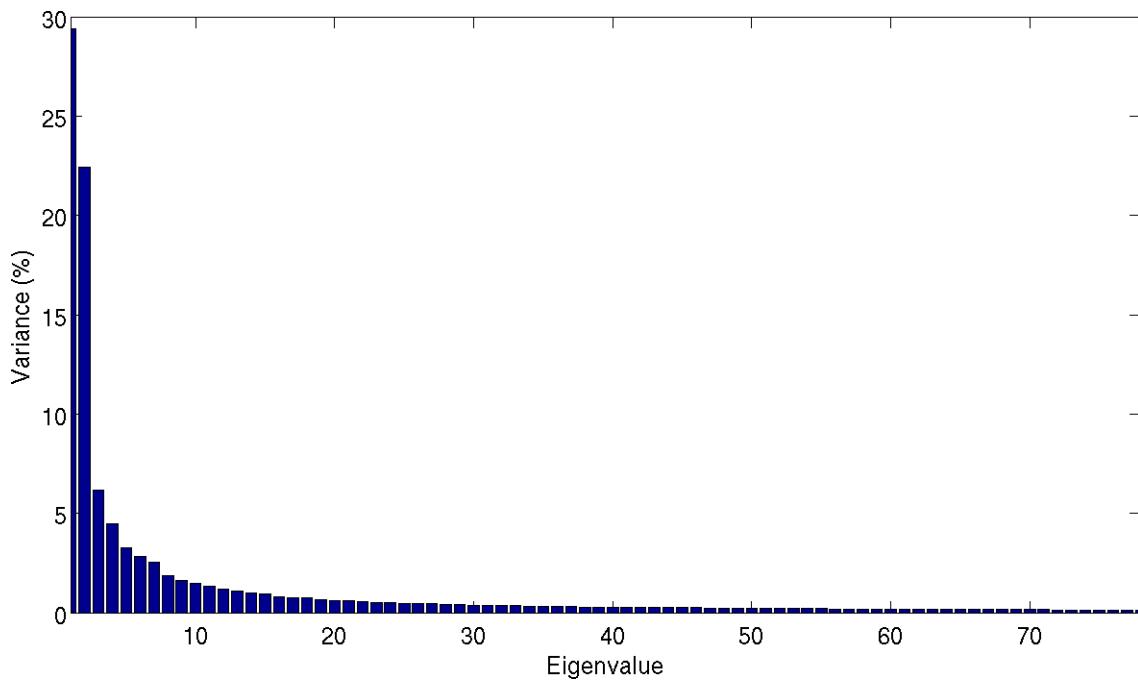
Πίνακας 7.2: Συμβολισμοί μοντέλου υφής προσώπου

Όπως και στην περίπτωση του μοντέλου σχήματος, για να αποφύγουμε την υπερβολική παραμόρφωση της μέσης υφής προσώπου και την αλλοίωση των χαρακτηριστικών του δέρματος του ανθρώπινου προσώπου πρέπει να επιλεγούν κατάλληλες παράμετροι λ_i . Ασφαλής επιλογή παραμέτρων υφής είναι στο διάστημα $[-3\sqrt{\delta_i}, 3\sqrt{\delta_i}]$. Κρατώντας το 92% της συνολικής μεταβλητότητας των διανυσμάτων υφής της βάσης δεδομένων προκύπτουν $m = 78$ άξονες μεταβλητότητας (modes of variation) και κατ' επέκταση 78 παράμετροι υφής λ_i .

Στο Σχήμα 7.16 φαίνεται ένα ραβδόγραμμα που απεικονίζει το ποσοστό της συνολικής μεταβλητότητας που αντιπροσωπεύει κάθε ιδιοτιμή και το αντίστοιχο ιδιοδιάνυσμά της A_i . Στο Σχήμα 7.15 φαίνονται οι παραμορφώσεις που προκαλούν στη μέση υφή προσώπου τα πέντε πιο σημαντικά ιδιοδιανύσματα. Η μεσαία στήλη απεικονίζει τη μέση υφή προσώπου και οι αριστερή και δεξιά στήλες δείχνουν της παραμόρφωση που προκαλείται για $\lambda_i = \pm 1.5\sqrt{\delta_i}, \pm 3\sqrt{\delta_i}, i = 1, \dots, m$. Υπενθυμίζεται ότι το μοντέλο υφής προσώπου εκπαιδεύτηκε για grayscale εικόνες.



Σχήμα 7.15: Τα πρώτα πέντε πιο σημαντικά ιδιοδιανύσματα (modes of variation) του μοντέλου υφής προσώπου



Σχήμα 7.16: Ποσοστό μεταβλητότητας που αντιστοιχεί σε κάθε ιδιοτιμή του μοντέλου υφής προσώπου

7.3 Ανεξάρτητο Μοντέλο Προσώπου

Σε αυτή την παράγραφο περιγράφεται η ένωση των στατιστικών μοντέλων σχήματος και υφής προσώπου. Ακόμη παρουσιάζεται ο Μετασχηματισμός Ομοιότητας που δημιουργεί τέσσερις επιπλέον παραμέτρους στο μοντέλο σχήματος και ο Ομοπαραλληλικός Μετασχηματισμός Υφής που δημιουργεί δυο επιπλέον παραμέτρους στο μοντέλο υφής.

Αρχικά οι Cootes, Edwards και Taylor [CET01] πρότειναν την εφαρμογή ενός τρίτου Principal Component Analysis που δημιουργεί ένα μοναδικό μοντέλο (combined model) το οποίο ελέγχει και το σχήμα και την υφή με ένα διάνυσμα παραμέτρων. Τα πλεονεκτήματα αυτού του συνδυασμένου μοντέλου είναι ότι χρειάζεται λιγότερες παραμέτρους για να δημιουργήσει ένα στιγμιότυπο προσώπου και είναι περισσότερο συμπαγές. Ωστόσο, δεν μπορεί να χρησιμοποιηθεί σε πολλούς αλγόριθμους ταιριάσματος του μοντέλου.

Οι Matthews και Baker [MaBa04] χρησιμοποιούν την *ανεξάρτητη* ένωση των μοντέλων σχήματος και υφής προσώπου. Συγκεκριμένα, για να δημιουργηθεί ένα στιγμιότυπο προσώπου, πρώτα φτιάχνονται το στιγμιότυπο σχήματος και το στιγμιότυπο υφής ανεξάρτητα, χρησιμοποιώντας δυο διαφορετικά διανύσματα παραμέτρων. Η ένωση γίνεται εφαρμόζοντας τη συνάρτηση στρέβλωσης (warp) στο στιγμιότυπο υφής ώστε να ταιριάξει με το στιγμιότυπο σχήματος. Το πλεονέκτημα του ανεξάρτητου μοντέλου (independent model) είναι ότι δίνει την ελευθερία της επιλογής οποιουδήποτε αλγορίθμου ταιριάσματος.

Στην εργασία αυτή χρησιμοποιείται το ανεξάρτητο μοντέλο προσώπου, ώστε να εφαρμοστούν οι βελτιώσεις στον αλγόριθμο ταιριάσματος που προτείνονται από τους Παπανδρέου και Μαραγκό [PaMa08].

7.3.1 Μοντέλο Σχήματος Προσώπου και Πόζα Προσώπου

Σε αυτή την παράγραφο αναλύεται ο Μετασχηματισμός Ομοιότητας και η επιρροή του στο μοντέλο σχήματος προσώπου και στη συνάρτηση Piecewise Affine Warp.

7.3.1.1 Μετασχηματισμός Ομοιότητας

Ο Μετασχηματισμός Ομοιότητας ορίζει την πόζα του προσώπου και ουσιαστικά δημιουργεί τέσσερις επιπλέον παραμέτρους. Σε μορφή πίνακα ορίζεται ως

$$S_t(x) = \begin{pmatrix} 1+t_1 & -t_2 \\ t_2 & 1+t_1 \end{pmatrix} x + \begin{pmatrix} t_3 \\ t_4 \end{pmatrix} = \begin{bmatrix} (1+t_1)x - t_2y + t_3 \\ t_2x + (1+t_1)y + t_4 \end{bmatrix} \quad (7.27)$$

όπου οι παράμετροι $t_{1:4} = [t_1, t_2, t_3, t_4]^T$ έχουν τις ακόλουθες ερμηνείες

$$\left. \begin{array}{l} \text{Κλίμακα: } s \\ \text{Περιστροφή: } \theta \end{array} \right\} \Rightarrow \begin{array}{l} t_1 = s \cos \theta - 1 \\ t_2 = s \sin \theta \end{array}$$

Μετακίνηση στον x: t_3
Μετακίνηση στον y: t_4

Θεωρώντας τα διανύσματα

$$\begin{aligned}\mathbf{s}_1^* &= (x_1^0, y_1^0, \dots, x_L^0, y_L^0)^T \\ \mathbf{s}_2^* &= (-y_1^0, x_1^0, \dots, -y_L^0, x_L^0)^T \\ \mathbf{s}_3^* &= (1, 0, \dots, 1, 0)^T \\ \mathbf{s}_4^* &= (0, 1, \dots, 0, 1)^T\end{aligned}\tag{7.28}$$

τότε ο γραμμικός συνδυασμός των διανυσμάτων αυτών είναι ακριβώς ίσος με το σύνολο των δισδιάστατων μετασχηματισμών ομοιότητας

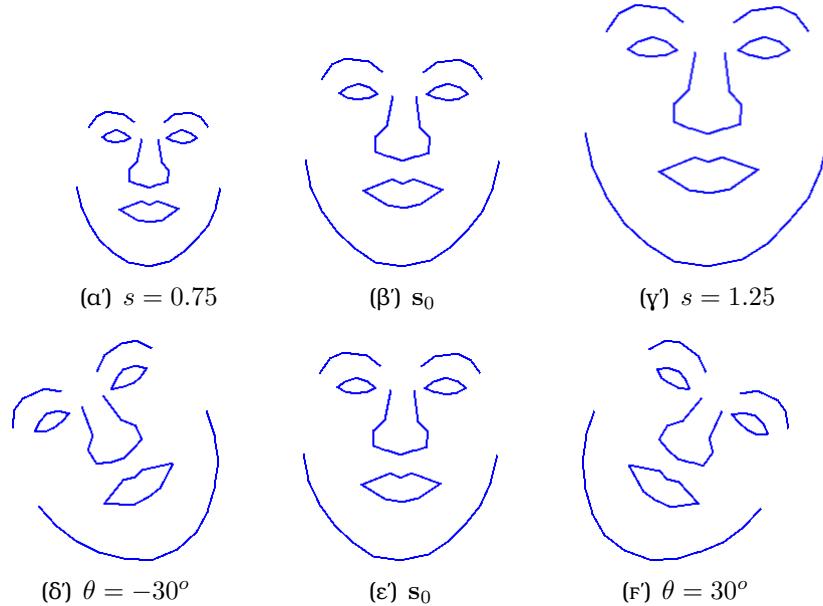
$$\mathbf{S}_t(\mathbf{x}) = \mathbf{s}_0 + \sum_{i=1}^4 t_i \mathbf{s}_i^* \tag{7.29}$$

όπου $(x_j^0, y_j^0), j = 1, \dots, L$ είναι το j -οστό στοιχείο του μέσου σχήματος προσώπου \mathbf{s}_0 .

Να σημειωθεί ότι σε κάποια βήματα του αλγορίθμου ταιριάσματος απαιτούνται τα διανύσματα \mathbf{s}_i^* , $i = 1, \dots, 4$ σε ορθοκανονικοποιημένη μορφή. Επομένως οι παράμετροι ομοιότητας πρέπει να σταθμιστούν με τον αντίστροφο του πολλαπλασιαστή ορθοκανονικοποίησης, ο οποίος είναι

$$m = [(\mathbf{S}_{orth}^*)^T \cdot \mathbf{S}^*]^{-1}$$

Ακόμη τονίζεται ότι για $\mathbf{t} = [0, 0, 0, 0]^T$ ο μετασχηματισμός δεν προκαλεί καμία επίδραση στο σχήμα εισόδου. Στο Σχήμα 7.17 φαίνεται η επίδραση του μετασχηματισμού ομοιότητας στο μέσω σχήμα προσώπου. Η πρώτη γραμμή εικόνων δείχνει την κλίμακα και η δεύτερη την περιστροφή.



Σχήμα 7.17: Παραδείγματα μετασχηματισμού ομοιότητας

7.3.1.2 Σύνθεση Συνάρτησης Warp και Μετασχηματισμού Ομοιότητας

Με τις τέσσερις επιπλέον παραμέτρους t του μετασχηματισμού Ομοιότητας που ορίστηκαν στην παράγραφο 7.3.1.1, η μορφή της συνάρτησης Piecewise Affine Warp, όπως ορίστηκε στην παράγραφο 7.2.2.2 θα πρέπει να αλλάξει ώστε να τις συμπεριλάβει.

Ορίζεται το συνολικό διάνυσμα παραμέτρων σχήματος προσώπου

$$\tilde{\mathbf{p}} = [\mathbf{t}_{1:4}, \mathbf{p}_{1:n}]^T \quad (7.30)$$

μήκους $4 + n$. Η σύνθεση της συνάρτησης warp $\mathbf{W}(\mathbf{x}; \mathbf{p})$ και του μετασχηματισμού ομοιότητας $\mathbf{S}_t(\mathbf{x})$ συμβολίζεται με $\mathbf{W}(\mathbf{x}; \tilde{\mathbf{p}})$ και παίρνει τη μορφή

$$\mathbf{W}(\mathbf{x}; \tilde{\mathbf{p}}) \equiv \mathbf{S}_t \circ \mathbf{W}(\mathbf{x}; \mathbf{p}) = \mathbf{S}_t(\mathbf{W}(\mathbf{x}; \mathbf{p}), \mathbf{t}) \quad (7.31)$$

Τα βήματα της διαδικασίας αυτής χρησιμοποιώντας τον ορισμό της συνάρτησης warp της εξίσωσης (7.14), το μετασχηματισμό ομοιότητας της εξίσωσης (7.29) και με βάση τον αλγόριθμο 6 φαίνονται στον αλγόριθμο 9.

Αλγόριθμος 9 Σύνθεση συνάρτησης warp και μετασχηματισμού ομοιότητας

- 1: Εφαρμογή συνάρτησης warp στο \mathbf{s}_0 : $\mathbf{W}(\mathbf{s}_0; \mathbf{p}) = \mathbf{s}_0 + \sum_{i=1}^n p_i \mathbf{s}_i$
 - 2: Εφαρμογή μετασχηματισμού ομοιότητας στο \mathbf{s}_0 : $\mathbf{S}_t(\mathbf{s}_0; \mathbf{t}) = \mathbf{s}_0 + \sum_{i=1}^4 t_i \mathbf{s}_i^*$
 - 3: Εφαρμογή Delaunay τριγωνισμού στο μέσο σχήμα \mathbf{s}_0
 - 4: **for** όλα τα τρίγωνα του πριγωνισμού του \mathbf{s}_0 **do**
 - 5: Έστω $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ οι κορυφές του τριγώνου του \mathbf{s}_0 και $\mathbf{x}'_1, \mathbf{x}'_2, \mathbf{x}'_3$ οι κορυφές του τριγώνου του $\mathbf{S}_t(\mathbf{s}_0; \mathbf{t})$
 - 6: Υπολογισμός και αποθήκευση των παραμέτρων $\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6$ από τη σχέση (7.15)
 - 7: **end for**
 - 8: **for** όλα τα landmark σημεία $i = 1, \dots, L$ **do**
 - 9: Έστω $\mathbf{x} = [x, y]^T$ οι συντεταγμένες του i -οστού landmark σημείου του $\mathbf{W}(\mathbf{s}_0; \mathbf{p})$
 - 10: $\mathbf{W}(\mathbf{x}; \tilde{\mathbf{p}}) = \mathbf{S}_t \circ \mathbf{W}(\mathbf{x}; \mathbf{p}) = \begin{pmatrix} \alpha_1 + \alpha_2 x + \alpha_3 y \\ \alpha_4 + \alpha_5 x + \alpha_6 y \end{pmatrix}$
 - 11: **end for**
-

Η αντιστροφή της συνάρτησης $\mathbf{W}(\mathbf{x}; \tilde{\mathbf{p}})$ που χρειάζεται στον αντίστροφο συνθετικό αλγόριθμο (inverse-compositional) της διαδικασίας ταιριάσματος του μοντέλου έχει τη μορφή

$$\mathbf{W}(\mathbf{x}; \tilde{\mathbf{p}})^{-1} = \mathbf{S}_t \circ \mathbf{W}(\mathbf{x}; -\mathbf{p}, -\mathbf{t}) = \mathbf{S}_t(\mathbf{W}(\mathbf{x}; -\mathbf{p}), -\mathbf{t}) \quad (7.32)$$

7.3.2 Μοντέλο Υφής Προσώπου και Συνθήκες Φωτεινότητας

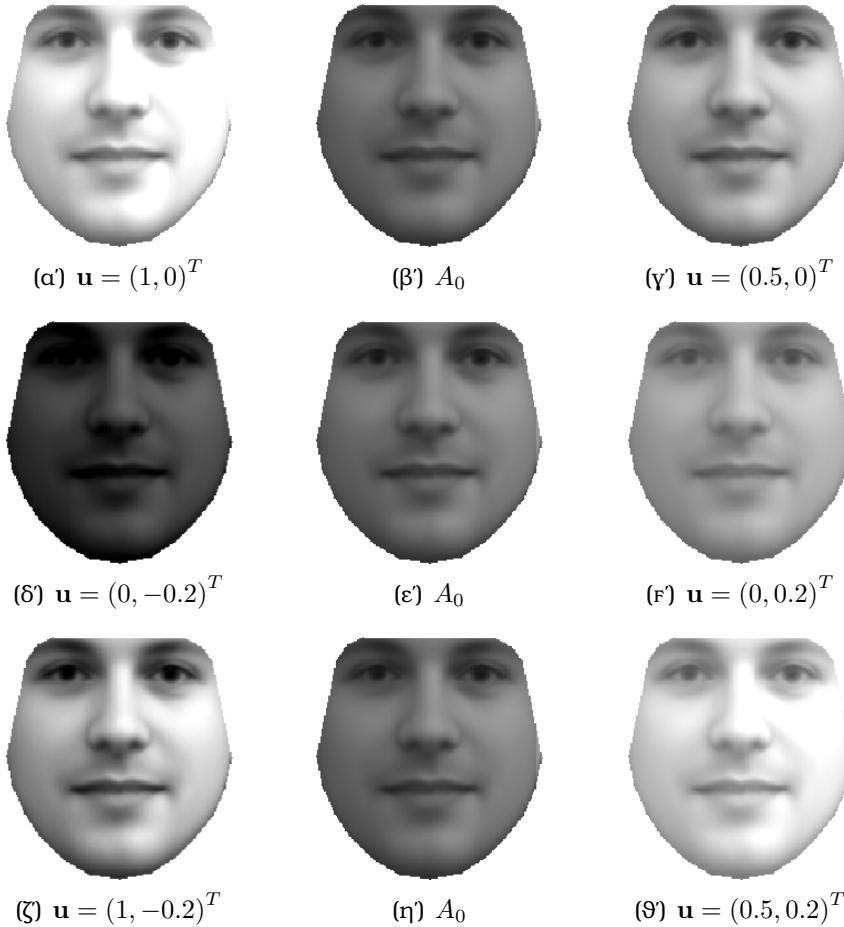
Σε αυτή την παράγραφο αναλύεται ο Ομοπαραλληλικός Μετασχηματισμός Υφής και η επιρροή του στο μοντέλο υφής προσώπου.

7.3.2.1 Ομοπαραλληλικός Μετασχηματισμός Υφής

Ο Ομοπαραλληλικός Μετασχηματισμός Υφής (Global Affine Texture Transform) έχει τη μορφή

$$T_{\mathbf{u}}(I) = (u_1 + 1) I + u_2 \quad (7.33)$$

όπου στο διάνυσμα παραμέτρων $\mathbf{u}_{1:2} = [u_1, u_2]^T$ το u_1 είναι το κέρδος (gain) και το u_2 η αντιστάθμιση (offset). Ο μετασχηματισμός έχει τέτοια μορφή ώστε για $\mathbf{u} = [0, 0]^T$ να μην προκαλεί καμία επίδραση στην εικόνα υφής εισόδου I . Στο Σχήμα 7.18 φαίνονται οι επιδράσεις του μετασχηματισμού υφής στο μέσο διάνυσμα υφής. Συγκεκριμένα, η πρώτη γραμμή δείχνει την επίδραση της παραμέτρου κέρδους u_1 , η δεύτερη γραμμή την επίδραση της αντιστάθμισης u_2 και η τρίτη την ταυτόχρονη επίδραση των δυο πρηγούμενων.



Σχήμα 7.18: Παραδείγματα ομοπαραλληλικού μετασχηματισμού υφής

7.3.2.2 Δειγματοληψία Εικόνας και Ομοπαραλληλικός Μετασχηματισμός Υφής

Ορίζεται το συνολικό διάνυσμα παραμέτρων υφής προσώπου

$$\tilde{\lambda} = [\mathbf{u}_{1:2}, \boldsymbol{\lambda}_{1:m}]^T \quad (7.34)$$

μήκους $2 + m$. Όπως φαίνεται στο επόμενο κεφάλαιο, ο ομοπαραλληλικός μετασχηματισμός υφής εφαρμόζεται στην δειγματοληψία της εικόνας εισόδου στην οποία απαιτείται το ταίριασμα του μοντέλου. Συγκεκριμένα, έχοντας βρει τις απαιτούμενες παραμέτρους σχήματος προσώπου, γίνεται δειγματοληψία των τιμών των pixels που ανήκουν στο σχήμα αυτό και οι τιμές, αφού εφαρμοστεί ο μετασχηματισμός, μεταφέρονται στο μέσο σχήμα προσώπου. Η διαδικασία αυτή με βάση τις σχέσεις (7.17) και (7.33) εκφράζεται από την εξίσωση

$$I'(\mathbf{s}_0) = T_{\mathbf{u}}(I(\mathbf{s}_p)) \Rightarrow I'(\mathbf{s}_0) = T_{\mathbf{u}}(I(\mathbf{W}(\mathbf{s}_0; \mathbf{p}))) \quad (7.35)$$

Βέβαια, ο ομοπαραλληλικός μετασχηματισμός υφής μπορεί να εφαρμοστεί και σε ένα στιγμιότυπο διανύσματος υφής από το αντίστοιχο μοντέλο με τη διαδικασία $T_{\mathbf{u}}(A_{\lambda})$.

7.3.3 Σύνθεση Προσώπου

Η σύνθεση ενός στιγμιότυπου προσώπου γίνεται με την ανεξάρτητη σύνθεση του σχήματος και της υφής του προσώπου ως

$$\mathbf{s} = \mathbf{s}_0 + \sum_{i=1}^n p_i \mathbf{s}_i \quad \text{και} \quad A = A_0 + \sum_{i=1}^m \lambda_i A_i$$

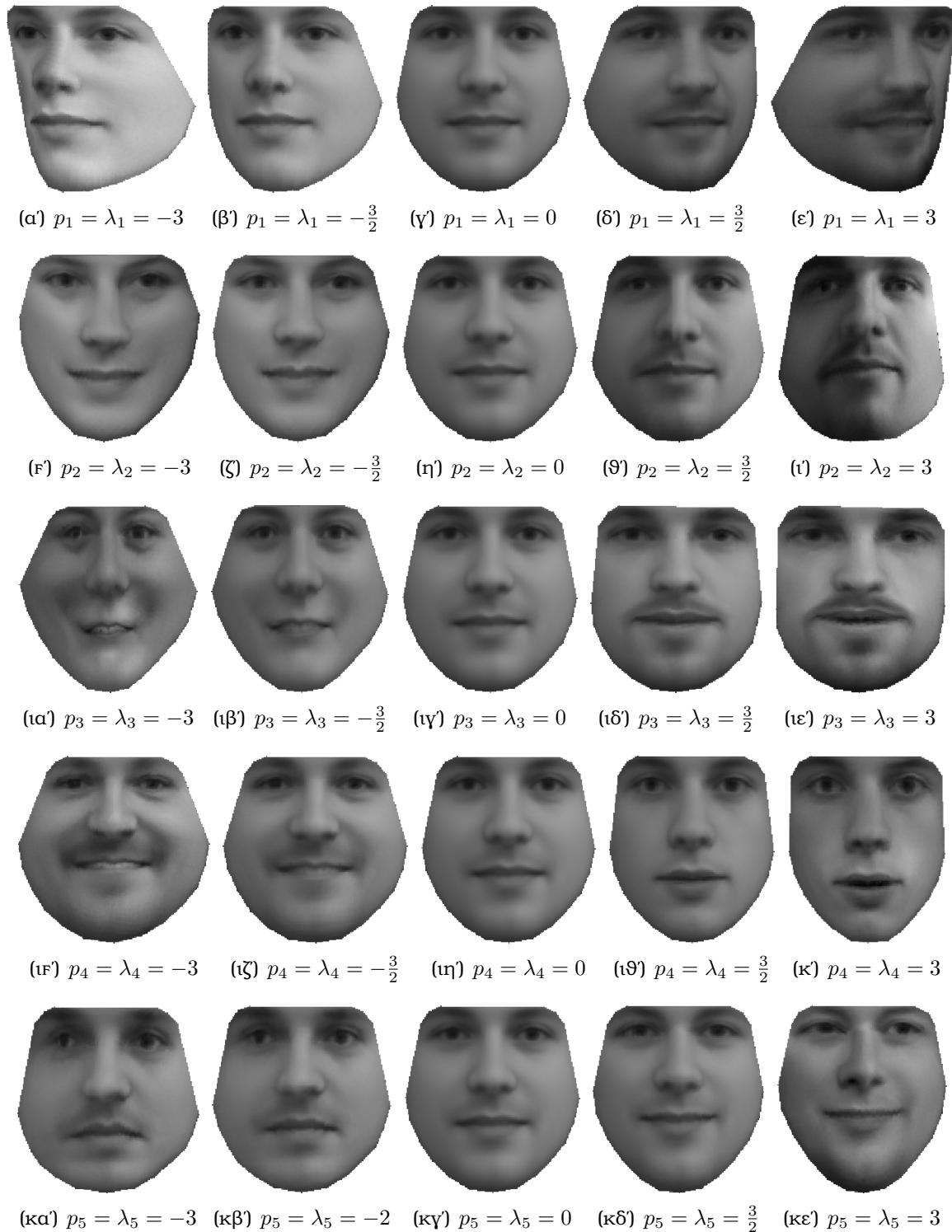
αντίστοιχα. Αρχικά, λοιπόν, οι παράμετροι \mathbf{p} δίνουν το στιγμιότυπο του σχήματος προσώπου. Επομένως, το διάνυσμα παραμέτρων \mathbf{p} ορίζει τη σχέση μεταξύ των σχημάτων \mathbf{s} και \mathbf{s}_0 και κατ' επέκταση τη συνάρτηση piecewise affine warp $\mathbf{W}(\mathbf{x}; \mathbf{p})$ του συνόλου των pixels με συνταταγμένες \mathbf{x} που ανήκουν στο μέσο σχήμα \mathbf{s}_0 . Εισάγωντας στη σχέση αυτή και τις παραμέτρους του μετασχηματισμού ομοιότητας \mathbf{t} για την πόλα του προσώπου, οι συνταταγμένες \mathbf{x}' του συνόλου των pixels που ανήκουν στο σχήμα \mathbf{s} δίνονται από τη σχέση

$$\mathbf{x}' = \mathbf{W}(\mathbf{x}; \tilde{\mathbf{p}})$$

Στη συνέχεια, οι παράμετροι $\boldsymbol{\lambda}$ μας δίνουν την υφή του προσώπου. Χρησιμοποιώντας και τις παραμέτρους \mathbf{u} του ομοπαραλληλικού μετασχηματισμού υφής, το συνθετικό διάνυσμα υφής προσώπου γίνεται $A' = T_{\mathbf{u}}(A)$. Αφού πρώτα μετατρέψουμε το διάνυσμα υφής σε δισδιάστατη εικόνα $I_{\mathbf{s}_0}(\mathbf{x}) = A'$, το μόνο που απομένει είναι η μεταφορά των τιμών των pixels από το μέσο σχήμα προσώπου του A' στο συνθετικό σχήμα με συνταταγμένες σημείων \mathbf{x}' , που γίνεται με τη διαδικασία της δειγματοληψίας

$$I(\mathbf{x}') = I_{\mathbf{s}_0}(\mathbf{x})$$

Παραδείγματα σύνθεσης προσώπου φαίνονται στο Σχήμα 7.19, συνδυάζοντας τα σχήματα 7.6 και 7.15. Οι τιμές των παραμέτρων που αναφέρονται πολλαπλασιάζονται με τη ρίζα των αντίστοιχων ιδιοτιμών.



Σχήμα 7.19: Παραδείγματα σύνθεσης προσώπου

Κεφάλαιο 8

Ταίριασμα Μοντέλου ΑΑΜ

Στο κεφάλαιο αυτό αναλύεται ο αλγόριθμος ταιριάσματος του μοντέλου ΑΑΜ σε νέες εικόνες, ώστε να επιτευχθεί η αναγνώριση του προσώπου. Η ανάλυση βασίζεται στις βελτιώσεις των Παπανδρέου & Μαραγκού [PaMa08]. Ο στόχος είναι η πολύ μικρή χρονική πολυπλοκότητα του αλγορίθμου, ώστε να εκτελείται σε πραγματικό χρόνο (real-time) και η επιτυχία του σε πρόσωπα που παρουσιάζουν μεγάλη διαφορά υφής από τα πρόσωπα της βάσης δεδομένων. Στο τέλος, παρουσιάζονται πλήρη πειραματικά αποτελέσματα σχετικά με την επιτυχία και την χρονική πολυπλοκότητα του αλγορίθμου.

8.1 Ορισμός Προβλήματος και Συμβολισμοί

Πριν δοθεί ο ορισμός του προβλήματος ταιριάσματος του ΑΑΜ, γίνεται μια σύνοψη των μαθηματικών ορισμών και συμβολισμών που χρησιμοποιήθηκαν κατά την εκπαίδευση του μοντέλου στο Κεφάλαιο 7.

Στον Πίνακα 8.1 δίνονται οι επεξηγήσεις όλων των συμβολισμών που αφορούν στο μοντέλο ΑΑΜ. Τα στιγμιότυπα σχήματος και υφής δημιουργούνται ως

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

και

$$A_\lambda = A_0 + \sum_{i=1}^m \lambda_i A_i$$

αντίστοιχα. Ο μετασχηματισμός ομοιότητας ορίζεται ως

$$\mathbf{S}_t(\mathbf{x}) = \begin{pmatrix} 1+t_1 & -t_2 \\ t_2 & 1+t_1 \end{pmatrix} \mathbf{x} + \begin{pmatrix} t_3 \\ t_4 \end{pmatrix} = \mathbf{s}_0 + \sum_{i=1}^4 t_i \mathbf{s}_i^*$$

όπου τα διανύσματα \mathbf{s}_i^* δίνονται από τις σχέσεις (7.28). Ο ομοπαραλληλικός μετασχηματισμός υφής δίνεται από τη σχέση

$$T_u(I) = (u_1 + 1)I + u_2$$

Σχήμα προσώπου	Ιδιοδιανύσματα Πλήθος ιδιοδιανυσμάτων Πλήθος landmark σημείων Παράμετροι σχήματος Στιγμιότυπο σχήματος προσώπου Μετασχηματισμός Ομοιότητας Παράμετροι μετασχηματισμού	s_i n L $p_{n \times 1}$ s_p S_t $t_{4 \times 1}$
Υφή προσώπου	Ιδιοδιανύσματα Πλήθος ιδιοδιανυσμάτων Πλήθος pixels μέσου σχήματος Παράμετροι υφής Στιγμιότυπο υφής προσώπου Ομοπαραλληλικός μετασχηματισμός Παράμετροι μετασχηματισμού	A_i m N $\lambda_{m \times 1}$ A_λ T_u $u_{2 \times 1}$

Πίνακας 8.1: Συμβολισμοί μοντέλου AAM

Τα συνολικά διανύσματα παραμέτρων σχήματος και υφής μαζί με τις παραμέτρους των μετασχηματισμών έχουν τη μορφή

$$\tilde{p} = \begin{bmatrix} t \\ p \end{bmatrix}_{(4+n) \times 1}$$

και

$$\tilde{\lambda} = \begin{bmatrix} u \\ \lambda \end{bmatrix}_{(2+m) \times 1}$$

αντίστοιχα. Προκύπτει λοιπόν το διάνυσμα παραμέτρων

$$q = \begin{bmatrix} \tilde{p} \\ \tilde{\lambda} \end{bmatrix} = \begin{bmatrix} t_{4 \times 1} \\ p_{n \times 1} \\ u_{2 \times 1} \\ \lambda_{m \times 1} \end{bmatrix}_{(n+m+6) \times 1} \quad (8.1)$$

μήκους $(n + m + 6) \times 1$. Τέλος η συνάρτηση

$$W(x; \tilde{p}) = S_t(W(x; p))$$

ορίζει τη εφαρμογή του piecewise affine warp στο σχήμα προσώπου ακολουθούμενη από μετασχηματισμό ομοιότητας. Επομένως, χαρτογραφεί κάθε σημείο x του στιγμιοτύπου του AAM στο αντίστοιχο σημείο της εικόνας.

Ορισμός προβλήματος: Να βρεθεί το διάνυσμα παραμέτρων q που ελαχιστοποιεί τη διαφορά μεταξύ της χαρτογραφημένης ομαλοποιημένης εικόνας $T_u(I(W(\tilde{p})))$ και της υφής του μοντέλου AAM A_λ .

8.2 Ταίριασμα Μοντέλου με Inverse-Compositional Αλγόριθμο

Σε αυτή την παράγραφο παρουσιάζονται συνοπτικά για λόγους πληρότητας οι μεθοδολογίες ταιριάσματος ΑΑΜ που αναπτύχθηκαν κυρίως από τους Baker και Matthews.

8.2.1 Project-Out Inverse Compositional Αλγόριθμος

Ο Project-Out Inverse Compositional (POIC) αλγόριθμος αναπτύχθηκε από τους Matthews και Baker [MaBa04] το 2004. Ουσιαστικά, αποφεύγει την ανανέωση των παραμέτρων υφής $\tilde{\lambda}$ και την αντιστροφή πινάκων σε κάθε βήμα επανάληψης. Συνεπώς είναι εξαιρετικά γρήγορος αλγόριθμος. Ωστόσο, το μεγάλο του μεινέκτημα είναι ότι η απόδοσή του είναι πολύ χαμηλή όταν παρουσιάζεται μεγάλη διαφορά στην υφή του προσώπου. Τα βήματά του φαίνονται στον αλγόριθμο 10.

Αλγόριθμος 10 Αλγόριθμος Project-Out Inverse Compositional

- 1: Υπολογισμός gradient ∇A_0 της μέσης υφής προσώπου $A_0(\mathbf{x})$
 - 2: Υπολογισμός Ιακωβιανών $\frac{\partial \mathbf{W}}{\partial \mathbf{p}}$ και $\frac{\partial \mathbf{S}_t}{\partial t}$ στο $(\mathbf{x}; \mathbf{0})$
 - 3: Υπολογισμός steepest descent εικόνων

$$SD_j(\mathbf{x}) = \nabla A_0 \cdot \frac{\partial \mathbf{S}_t}{\partial t_j} - \sum_{i=1}^m \left[\sum_{\mathbf{x} \in s_0} A_i(\mathbf{x}) \cdot \nabla A_0 \cdot \frac{\partial \mathbf{S}_t}{\partial q_j} \right] A_i(\mathbf{x}), \text{ για } (t_1, t_2, t_3, t_4)$$

$$SD_{j+4}(\mathbf{x}) = \nabla A_0 \cdot \frac{\partial \mathbf{W}}{\partial p_j} - \sum_{i=1}^m \left[\sum_{\mathbf{x} \in s_0} A_i(\mathbf{x}) \cdot \nabla A_0 \cdot \frac{\partial \mathbf{W}}{\partial p_j} \right] A_i(\mathbf{x}), \text{ για } j = 1, \dots, n$$
 - 4: Υπολογισμός Εσσιανού (Hessian) πίνακα με τις SD_j με $j = 1, \dots, n + 4$.
Συγκεκριμένα το (j, k) στοιχείο του $(n + 4) \times (n + 4)$ Εσσιανού πίνακα ισούται με

$$\mathbf{H}_{j,k} = \sum_{\mathbf{x} \in s_0} SD_j(\mathbf{x}) \cdot SD_k(\mathbf{x})$$
 - 5: Αντιστροφή Εσσιανού πίνακα \mathbf{H}^{-1}
 - 6: **repeat**
 - 7: Υπολογισμός $I(\mathbf{W}(\mathbf{x}; \tilde{\mathbf{p}}))$ στην εικόνα εισόδου I
 - 8: Υπολογισμός εικόνας λάθους $E(\mathbf{x}) = I(\mathbf{W}(\mathbf{x}; \tilde{\mathbf{p}})) - A_0(\mathbf{x})$
 - 9: Υπολογισμός εσωτερικού γινομένου SD_i με εικόνα λάθους

$$\sum_{\mathbf{x} \in s_0} SD_i(\mathbf{x}) \cdot E(\mathbf{x}), \text{ για } i = 1, \dots, n + 4$$
 - 10: Υπολογισμός $d\tilde{\mathbf{p}}$ πολλαπλασιάζοντας με τον αντίστροφο Εσσιανό πίνακα

$$d\tilde{\mathbf{p}} = -\mathbf{H}^{-1} \sum_{\mathbf{x} \in s_0} SD_i(\mathbf{x}) \cdot E(\mathbf{x}), \text{ για } i = 1, \dots, n + 4$$
 - 11: Ανανέωση $\mathbf{W}(\mathbf{x}; \tilde{\mathbf{p}}) \leftarrow \mathbf{W}(\mathbf{x}; \tilde{\mathbf{p}}) \circ \mathbf{W}(\mathbf{x}; d\tilde{\mathbf{p}})^{-1}$
 - 12: **until** σύγκλιση
 - 13: Υπολογισμός $\lambda_i = \sum_{\mathbf{x} \in s_0} A_i(\mathbf{x}) [I(\mathbf{W}(\mathbf{x}; \tilde{\mathbf{p}})) - A_0(\mathbf{x})]$
-

8.2.2 Simultaneous Inverse Compositional Αλγόριθμος

Ο Simultaneous Inverse Compositional (SIC) αλγόριθμος αναπτύχθηκε από τους Gross, Matthews και Baker [GMB05] το 2005. Συγκριτικά με τον POIC αλγόριθμο είναι πολύ πιο ακριβής αλλά αρκετά αργός. Τα βήματά του φαίνονται στον αλγόριθμο 11.

Αλγόριθμος 11 Αλγόριθμος Simultaneous Inverse Compositional

- 1: Υπολογισμός των gradient ∇A_0 και ∇A_i για $i = 1, \dots, m$
 - 2: Υπολογισμός Ιακωβιανής $\frac{\partial \mathbf{W}}{\partial \tilde{\mathbf{p}}}$ στο $(\mathbf{x}; \mathbf{0})$
 - 3: **repeat**
 - 4: Υπολογισμός $I(\mathbf{W}(\mathbf{x}; \tilde{\mathbf{p}}))$ στην εικόνα εισόδου I
 - 5: Υπολογισμός εικόνας λάθους $E(\mathbf{x}) = I(\mathbf{W}(\mathbf{x}; \tilde{\mathbf{p}})) - \left[A_0(\mathbf{x}) + \sum_{i=1}^m \lambda_i A_i(\mathbf{x}) \right]$
 - 6: Υπολογισμός steepest descent εικόνων

$$\mathbf{SD}_{sim}(\mathbf{x}) = \left(\nabla A \cdot \frac{\partial \mathbf{W}}{\partial \tilde{p}_1}, \dots, \nabla A \cdot \frac{\partial \mathbf{W}}{\partial \tilde{p}_{n+4}}, A_1(\mathbf{x}), \dots, A_m(\mathbf{x}) \right)$$
 όπου $\nabla A = \nabla A_0 + \sum_{i=1}^m \lambda_i \cdot \nabla A_i$
 - 7: Υπολογισμός Εσσιανού (Hessian) πίνακα με τον \mathbf{SD}_{sim} και αντιστροφή του ως

$$\mathbf{H}_{sim}^{-1} = \sum_{\mathbf{x} \in \mathbf{s}_0} \mathbf{SD}_{sim}^T(\mathbf{x}) \cdot \mathbf{SD}_{sim}(\mathbf{x})$$
 - 8: Υπολογισμός εσωτερικού γινομένου \mathbf{SD}_{sim} με εικόνα λάθους

$$\sum_{\mathbf{x}} \mathbf{SD}_{sim}^T(\mathbf{x}) \cdot E(\mathbf{x})$$
 - 9: Υπολογισμός $d\mathbf{q}$ πολλαπλασιάζοντας με τον αντίστροφο Εσσιανό πίνακα

$$d\mathbf{q} = -\mathbf{H}_{sim}^{-1} \sum_{\mathbf{x}} \mathbf{SD}_{sim}^T(\mathbf{x}) \cdot E(\mathbf{x})$$
 - 10: Ανανέωση $\mathbf{W}(\mathbf{x}; \tilde{\mathbf{p}}) \leftarrow \mathbf{W}(\mathbf{x}; \tilde{\mathbf{p}}) \circ \mathbf{W}(\mathbf{x}; d\tilde{\mathbf{p}})^{-1}$ και $\boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} + d\boldsymbol{\lambda}$
 - 11: **until** σύγκλιση
-

8.2.3 Βήμα Ανανέωσης Inverse Compositional Αλγορίθμου

Στους προηγούμενους αλγορίθμους ταιριάσματος του μοντέλου ΑΑΜ το βήμα ανανέωσης των παραμέτρων σχήματος γίνεται με την αντίστροφη-συνθετική (inverse-compositional) μέθοδο σε αντίθεση με την ανανέωση των παραμέτρων υφής στον SIC που γίνεται με την ευθεία-προσθετική (forwards-additive) μέθοδο.

Ισχύει η εξίσωση

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

Ωστόσο, από τα βήματα του αλγορίθμου φαίνεται ότι έχει ήδη υπολογιστεί (ή μπορεί να υπολογιστεί) η τρέχουσα εκτίμηση του σχήματος προσώπου

$$\mathbf{s}_{0,cur} = \mathbf{W}(\mathbf{x}; \tilde{\mathbf{p}})$$

και η αντίστροφη παραμόρφωση του μέσου σχήματος με παραμέτρους $d\tilde{\mathbf{p}}$

$$\mathbf{s}_{0,inv} = \mathbf{W}^{-1}(\mathbf{x}; d\tilde{\mathbf{p}})$$

Το μόνο που απομένει είναι η σύνθεση των δυο σχημάτων $\mathbf{W}(\mathbf{x}; \tilde{\mathbf{p}}) \circ \mathbf{W}^{-1}(\mathbf{x}; d\tilde{\mathbf{p}})$ με βάση τη μέθοδο της συνάρτησης piecewise affine warp. Συγκεκριμένα, ορίζουμε τις σχέσεις

$$\mathbf{s}_{0,inv} = (x_i^0, y_i^0)^T + (\Delta x_i^0, \Delta y_i^0)^T$$

και

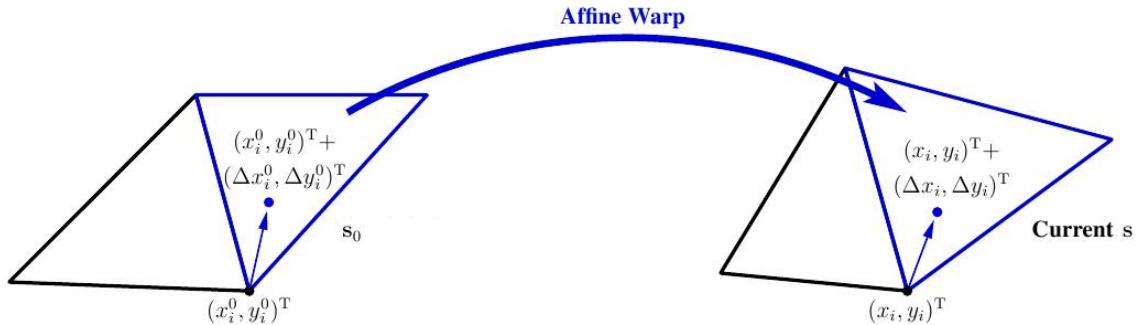
$$\mathbf{s}_{0,cur} = (x_i, y_i)^T$$

Ακόμη ορίζουμε την παραμόρφωση του τρέχοντος σχήματος προσώπου με τις νέες παραμέτρους σχήματος $\tilde{\mathbf{p}}$ ως

$$\mathbf{s}_{0,upd} = (x_i, y_i)^T + (\Delta x_i, \Delta y_i)^T$$

Απαιτείται, λοιπόν, ο υπολογισμός των μεταβολών $(\Delta x_i, \Delta y_i)^T$ στο τρέχων σχήμα $\mathbf{s}_{0,cur}$. Αυτός μπορεί να επιτευχθεί εφαρμόζοντας τη συνάρτηση piecewise affine warp για κάθε τρίγωνο του μέσου σχήματος \mathbf{s}_0 που έχει την i -οστή κορυφή με το σημείο $(x_i^0, y_i^0)^T + (\Delta x_i^0, \Delta y_i^0)^T$ ώστε να βρεθούν πολλές εκτιμήσεις του $(x_i, y_i)^T + (\Delta x_i, \Delta y_i)^T$. Οι εκτιμήσεις αυτές είναι πολλές διότι η i -οστή κορύφη ανήκει σε πολλά τρίγωνα, με αποτέλεσμα η εκτίμηση του τρέχοντος σχήματος να μην είναι μοναδικά ορισμένη. Αυτό το πρόβλημα μπορεί να αντιμετωπιστεί υπολογίζοντας τον μέσο όρο των πολλαπλών εκτιμήσεων $(x_i, y_i)^T + (\Delta x_i, \Delta y_i)^T$, ώστε να βρεθούν οι νέες συντεταγμένες των κορυφών του τρέχοντος σχήματος $\mathbf{s}_{0,cur}$.

Η διαδικασία αυτή απεικονίζεται στο Σχήμα 8.1 και τα βήματά της συνοψίζονται στον αλγόριθμο 12 με βάση τους ορισμούς και τους αλγορίθμους των ενοτήτων 7.2.2.2 και 7.3.1.



Σχήμα 8.1: Βήμα Ανανέωσης Inverse Compositional Αλγορίθμου

Αλγόριθμος 12 Βήμα Ανανέωσης Inverse Compositional Αλγορίθμου

- 1: Παραμόρφωση s_0 με $(S_t \circ W)^{-1}$ δίνει $s_{0,inv} = S_t \circ W(s_0; -d\tilde{p})$
 - 2: Παραμόρφωση s_0 με τις τρέχουσες παραμέτρους \tilde{p} δίνει $s_{0,cur} = S_t \circ W(s_0; \tilde{p})$
 - 3: Εφαρμογή Delaunay τριγωνισμού στο μέσο σχήμα s_0
 - 4: **for** όλα τα τριγώνα του τριγωνισμού του s_0 **do**
 - 5: Έστω x_1, x_2, x_3 οι κορυφές του τριγώνου του s_0 και x'_1, x'_2, x'_3 οι κορυφές του αντίστοιχου τριγώνου του $s_{0,cur}$
 - 6: Υπολογισμός και αποθήκευση των παραμέτρων $\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6$ από τη σχέση (7.15)
 - 7: **end for**
 - 8: **for** όλα τα landmark σημεία $i = 1, \dots, L$ **do**
 - 9: Εύρεση των τριγώνων του s_0 που έχουν το i σημείο ως μια από τις κορυφές τους
 - 10: Αποθήκευση του πλήθους των τριγώνων που βρέθηκαν ως T
 - 11: Υπολογισμός:
$$s_{0,upd}(i, 1) = \frac{\sum_{t=1}^T \alpha_{1,t} + \alpha_{2,t} \cdot s_{0,inv}(i, 1) + \alpha_{3,t} \cdot s_{0,inv}(i, 2)}{T}$$

$$s_{0,upd}(i, 2) = \frac{\sum_{t=1}^T \alpha_{4,t} + \alpha_{5,t} \cdot s_{0,inv}(i, 1) + \alpha_{6,t} \cdot s_{0,inv}(i, 2)}{T}$$
 - 12: **end for**
 - 13: Ανανέωση παραμέτρων μετασχηματισμού ομοιότητας:
$$t_i = [s_{i,orth}^*]^T (s_{0,upd} - s_0), \text{ για } i = 1, \dots, 4$$
 - 14: Ανανέωση παραμέτρων σχήματος:
$$p_i = s_i^T [S_t(s_{0,upd}; t)^{-1} - s_0] = s_i^T [S_t(s_{0,upd}; -t) - s_0], \text{ για } i = 1, \dots, n$$
-

8.3 Προσαρμοστικά Inverse Compositional Μοντέλα ΑΑΜ

8.3.1 Θεωρητική Ανάλυση

Από τους αλγορίθμους του κεφαλαίου 8.2 συμπεραίνεται ότι το πρόβλημα ταιριάσματος του μοντέλου AAM είναι η εύρεση του διανύσματος ανανέωσης παραμέτρων

$$d\mathbf{q} = \begin{bmatrix} d\tilde{\mathbf{p}} \\ d\tilde{\boldsymbol{\lambda}} \end{bmatrix}$$

που ελαχιστοποιεί τη νόρμα της εικόνας λάθους $\|E(\mathbf{q}, d\mathbf{q})\|_2$ ως προς $d\mathbf{q}$, όπου η εικόνα λάθους δίνεται από τον τύπο

$$E(\mathbf{x}; \mathbf{q}, d\mathbf{q}) = T_{\mathbf{u}}(I(\mathbf{W}(\mathbf{x}'\tilde{\mathbf{p}}))) - T_{du}(A_{\boldsymbol{\lambda}+d\boldsymbol{\lambda}}(\mathbf{W}(\mathbf{x}); d\tilde{\mathbf{p}})) \quad (8.2)$$

Εφαρμόζωντας σειρά Taylor πρώτου βαθμού γύρω από μηδενικό διάνυσμα $d\mathbf{q}$ στην εικόνα λάθους και αγνοώντας τους δευτεροβάθμιους όρους προκύπτει

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

Για να αξιοποιηθεί η παραπάνω έκφραση ορίζουμε εκ νέου το διάνυσμα $d\mathbf{q}'$ ως

$$\left. \begin{aligned} d\mathbf{q}' &= \begin{bmatrix} d\tilde{\mathbf{p}} \\ d\tilde{\boldsymbol{\lambda}}' \end{bmatrix} \\ \text{όπου } d\tilde{\boldsymbol{\lambda}}' &= [d\mathbf{u}_{1:2}, d\boldsymbol{\lambda}'_{1:m}]^T \\ \text{και } d\boldsymbol{\lambda}' &= d\boldsymbol{\lambda} - u_1 \boldsymbol{\lambda} \end{aligned} \right\} \Rightarrow d\mathbf{q}' = \begin{bmatrix} dt_{1:4} \\ dp_{1:n} \\ du_{1:2} \\ d\boldsymbol{\lambda}_{1:m} - u_1 \boldsymbol{\lambda}_{1:m} \end{bmatrix} \quad (8.3)$$

Το πρόβλημα ελαχίστων τετραγώνων που πρέπει να λυθεί είναι

$$\min \frac{1}{2\sigma^2} \|E(\mathbf{q}) - H_{\boldsymbol{\lambda}} d\mathbf{q}'\|_2^2 \quad (8.4)$$

όπου

$$E(\mathbf{q}) = T_{\mathbf{u}}(I(\mathbf{W}(\tilde{\mathbf{p}}))) - A_{\boldsymbol{\lambda}} = T_{\mathbf{u}}(I(\mathbf{W}(\tilde{\mathbf{p}}))) - \left(A_0 + \sum_{i=1}^m \lambda_i A_i \right) \quad (8.5)$$

και

$$\left. \begin{aligned} M_{\boldsymbol{\lambda}} &= M_0 + \sum_{i=1}^m \lambda_i M_i && \text{variant} \\ \text{όπου } M_i(\mathbf{x}) &= \left. \frac{\partial A_i(\mathbf{x})}{\partial \mathbf{x}} \frac{\partial \mathbf{W}(\mathbf{x})}{\partial \tilde{\mathbf{p}}} \right|_{\tilde{\mathbf{p}}=0} && \text{fixed} \\ \left. \frac{\partial \mathbf{W}(\mathbf{x})}{\partial \tilde{\mathbf{p}}} \right|_{\tilde{\mathbf{p}}=0} &= \left[\frac{\partial \mathbf{S}_t}{\partial t}, \frac{\partial \mathbf{W}(\mathbf{x})}{\partial p} \right]_{\tilde{\mathbf{p}}=0} && \text{fixed} \\ A(\mathbf{x}) &= [A_0(\mathbf{x}) \ 1 \ A_1(\mathbf{x}) \ \dots \ A_m(\mathbf{x})] && \text{fixed} \end{aligned} \right\} \Rightarrow H_{\boldsymbol{\lambda}} = [M_{\boldsymbol{\lambda}} \ A] \quad (8.6)$$

Οι M_{λ} είναι *steepest descent* εικόνες παραμόρφωσης σχήματος ή *πρότυπα κίνησης* (*motion templates*) όπως αποκαλούνται στο [HaBe98], που αντιστοιχούν στις $4+n$ παραμέτρους σχήματος $d\tilde{\mathbf{p}}$. Έχουν διαστάσεις $N \times (4+n)$ και συντίθενται από γραμμικό συνδυασμό των εικόνων M_i . Κάθε εικόνα $M_i, i = 1, \dots, m$ είναι ένας σταθερός $N \times (n+4)$ πίνακας που απομονώνει τη συμβολή του ιδιοδιανύσματος A_i στα πρότυπα κίνησης. Ο πίνακας M_0 συγκεντρώνει τα μη-προσαρμοσμένα πρότυπα κίνησης που αντιστοιχούν στο διάνυσμα μέσης υφής προσώπου A_0 .

Ακόμα η *Iakabianή* της συνάρτησης *warp* $\frac{\partial \mathbf{W}(\mathbf{x})}{\partial \tilde{\mathbf{p}}} = \left[\frac{\partial \mathbf{S}_t}{\partial t}, \frac{\partial \mathbf{W}(\mathbf{x})}{\partial \mathbf{p}} \right]$ υπολογίζεται για μηδενικό διάνυσμα $\tilde{\mathbf{p}}$ και επομένως, λόγω της μεθόδου *inverse-compositional*, είναι μια σταθερή μήτρα διάστασης $2 \times (n+4)$ ανά pixel, δηλαδή $N \times 2 \times (n+4)$ συνολικά.

Οι εικόνες $A(\mathbf{x}) = [A_0(\mathbf{x}) \ A_1(\mathbf{x}) \ \dots \ A_m(\mathbf{x})]$ είναι *steepest descent* εικόνες διόρθωσης φωτεινότητας και μεταβολής υφής, ανεξάρτητες της εικόνας εισόδου, που αντιστοιχούν στο διάνυσμα $d\lambda'$. Επομένως, είναι σταθεροί πίνακες διάστασης $1 \times (m+2)$ ανά pixel ή $N \times (m+2)$ συνολικά.

Τελικά, η λύση του προβλήματος ελαχίστων τετραγώνων (8.4) είναι

$$d\mathbf{q}' = \sigma^{-2} \Sigma_{\mathbf{q}} H_{\lambda}^T E(\mathbf{q}) \quad (8.7)$$

όπου ο πίνακας συνδιακύμανσης είναι

$$\Sigma_{\mathbf{q}} = \sigma^2 (H_{\lambda}^T H_{\lambda})^{-1} = \sigma^2 \begin{bmatrix} M_{\lambda}^T M_{\lambda} & M_{\lambda}^T A \\ A^T M_{\lambda} & A^T A \end{bmatrix}^{-1} \quad (8.8)$$

Διαχωρίζοντας την ανανέωση των παραμέτρων σχήματος και των παραμέτρων υφής της λύσης (8.7), οι τελικές εκτιμήσεις προκύπτουν

$$\left. \begin{array}{l} A_{temp} = (A^T A)^{-1} A^T \quad [(m+2) \times N \text{ fixed}] \\ P = I - AA_{temp} = I - A (A^T A)^{-1} A^T \quad [N \times N \text{ fixed}] \\ M'_{\lambda} = PM_{\lambda} \quad [N \times (n+4) \text{ variant}] \\ M'_{\lambda} = M'_0 + \sum_{i=1}^m \lambda_i M'_i \quad [N \times (n+4) \text{ variant}] \\ \text{όπου } M'_i = PM_i \end{array} \right\} \Rightarrow$$

$$\left. \begin{array}{l} d\tilde{\mathbf{p}} = \left(M'_{\lambda}^T M'_{\lambda} \right)^{-1} M_{\lambda}^T PE(\mathbf{q}) \\ d\tilde{\lambda}' = (A^T A)^{-1} A^T [E(\mathbf{q}) - M_{\lambda} d\tilde{\mathbf{p}}] \end{array} \right\} \Rightarrow \left. \begin{array}{l} \left[\begin{array}{c} dt \\ d\mathbf{p} \end{array} \right] = \left(M'_{\lambda}^T M'_{\lambda} \right)^{-1} M_{\lambda}^T PE(\mathbf{q}) \\ \left[\begin{array}{c} d\mathbf{u} \\ d\lambda' \end{array} \right] = (A^T A)^{-1} A^T [E(\mathbf{q}) - M_{\lambda} d\tilde{\mathbf{p}}] \end{array} \right\} \Rightarrow$$

$$\left. \begin{array}{c} d\lambda = d\lambda' + u_1 \lambda \\ \xrightarrow{d\lambda = d\lambda' + u_1 \lambda} d\mathbf{q} = \left[\begin{array}{c} d\tilde{\mathbf{p}} \\ d\tilde{\lambda}' \end{array} \right] = \left[\begin{array}{c} dt \\ d\mathbf{p} \\ d\mathbf{u} \\ d\lambda \end{array} \right] \end{array} \right\} \quad (8.9)$$

Στις τελικές εκτιμήσεις (8.9) ο πίνακας P ονομάζεται πίνακας *προβολής* (*projection matrix*), καθώς προβάλλει διανύσματα στον συμπληρωματικό υποχώρο που καλύπτεται από τις στήλες του πίνακα $A = [A_0 \ 1_{N \times 1} \ A_1 \ \dots \ A_m]$. Ο πίνακας αυτός είναι ιδιαίτερα μεγάλος έχοντας διαστάσεις $N \times N$.

Ακόμη, η εικόνα M'_λ είναι η *προβολή της εικόνας κίνησης* (*project-out motion template*) διάστασης $N \times (n+4)$. Ουσιαστικά, είναι η προβολή της εικόνας M_λ με τον πίνακα P .

Η ανανέωση των παραμέτρων του μοντέλου γίνεται με τους ακόλουθους τρόπους:

- m παράμετροι υφής: $\lambda \leftarrow \lambda + d\lambda$
- 2 παράμετροι φωτεινότητας: $T_u \leftarrow T_u \circ T_{du}^{-1}$
- $n+4$ παράμετροι σχήματος: $\mathbf{W}(\mathbf{x}; \tilde{\mathbf{p}}) \leftarrow \mathbf{W}(\mathbf{x}; \tilde{\mathbf{p}}) \circ \mathbf{W}^{-1}(\mathbf{x}; d\tilde{\mathbf{p}})$
 $\tilde{\mathbf{p}} \leftarrow \tilde{\mathbf{p}} + J_{\tilde{\mathbf{p}}} d\tilde{\mathbf{p}}$

Συνεπώς, η ανανέωση των παραμέτρων υφής γίνεται με τον ευθή-προσθετικό τρόπο (forwards-additive) και η ανανέωση των παραμέτρων φωτεινότητας με τον αντίστροφο-συνθετικό (inverse-compositional). Όσον αφορά στις $n+4$ παραμέτρους σχήματος η ανανέωση μπορεί να γίνει είτε με τον αντίστροφο-συνθετικό τρόπο της ενότητας 8.2.3, είτε με τον ευθή-προσθετικό τρόπο που αναλύεται παρακάτω στην ενότητα 8.3.4.3.

8.3.1.1 Τάξη Αλγορίθμου

Από την ανάλυση της προηγούμενης παραγράφου γίνεται ξεκάθαρο ότι χρησιμοποιώντας τις προβολές των προτύπων κίνησης M'_λ , ο αλγόριθμος που αναλύθηκε είναι ισοδύναμος με τον ακριβή αλλά αργό SIC της παραγράφου 8.2.2. Αντίστοιχα, χρησιμοποιώντας τα μη-προσαρμοσμένα πρότυπα κίνησης M_0 που αντιστοιχούν στο διάνυσμα μέσης υφής A_0 , δηλαδή θέτοντας $M'_\lambda = M'_0$, τότε ο αλγόριθμος ισοδυναμεί με την γρήγορη αλλά μη ακριβή μέθοδο POIC της παραγράφου 8.2.1.

Αν όμως διατηρηθούν τα r από τα m συστατικά των πλήρως προσαρμοσένων προτύπων κίνησης M_λ και M'_λ , τότε είναι δυνατή η *μερική προσαρμογή* των προτύπων κίνησης. Θεωρούμε τις προσεγγίσεις

$$M_\lambda \approx M_0 + \sum_{i=1}^r \lambda_i M_i \quad (8.10)$$

και

$$M'_\lambda \approx M'_0 + \sum_{i=1}^r \lambda_i M'_i \quad (8.11)$$

όπου η παράμετρος r ονομάζεται *τάξη* του αλγορίθμου

$$0 \leq r \leq m \quad (8.12)$$

Με τη χρήση της παραμέτρου r ο χρήστης έχει τη δυνατότητα να ρυθμίσει την ταχύτητα του αλγορίθμου έναντι της ακρίβειας των αποτελεσμάτων του. Συγκεκριμένα, όσο πιο

μικρή είναι η τιμή της r τόσο πιο γρήγορος γίνεται ο αλγόριθμος, ενώ όσο αυξάνεται η τιμή της r τόσο πιο ακριβές γίνεται το ταίριασμα του μοντέλου. Για τις ακραίες τιμές ισχύει

$$\begin{aligned} \text{για } r = m &\rightarrow \text{SIC} \\ \text{για } r = 0 &\rightarrow \text{POIC} \end{aligned}$$

Η μεγαλύτερη διαφορά που προκαλεί η παράμετρος r είναι κατά τον υπολογισμό της παράστασης $d\tilde{\lambda}' = (A^T A)^{-1} A^T [E(\mathbf{q}) - M_{\lambda} d\tilde{\mathbf{p}}]$. Απαιτείται ο υπολογισμός μόνο των πρώτων $2 + r$ στοιχείων του διανύσματος $d\tilde{\lambda}'$, με αποτέλεσμα η εικόνα λάθους $E(\mathbf{q})$ να παίρνει τη μορφή

$$E(\mathbf{q}) = T_{\mathbf{u}}(I(\mathbf{W}(\tilde{\mathbf{p}}))) - A_{\lambda} = T_{\mathbf{u}}(I(\mathbf{W}(\tilde{\mathbf{p}}))) - \left(A_0 + \sum_{i=1}^r \lambda_i A_i \right) \quad (8.13)$$

8.3.1.2 Πολυπλοκότητα Αλγορίθμου

Ένα από τα βήματα με το μεγαλύτερο χρονικό κόστος του αλγορίθμου είναι ο υπολογισμός του $M'_{\lambda}^T M'_{\lambda}$. Αν γίνει με ευθύ τρόπο, τότε η πολυπλοκότητα είναι $\mathcal{O}(Nn^2)$. Ωστόσο, αν στα βήματα εκτός του επαναληπτικού βρόγχου υπολογιστούν οι πίνακες συσχέτισης

$$\begin{aligned} R_{ij} = R_{ji} &= M'_i{}^T M'_j + M'_j{}^T M'_i, \quad i \neq j \\ R_{ii} &= M'_i{}^T M'_i, \quad i, j = 0, \dots, m \end{aligned} \quad (8.14)$$

όπου κάθε R_{ij} είναι ένας $(n+4) \times (n+4)$ πίνακας, τότε ο υπολογισμός παίρνει τη μορφή

$$M'_{\lambda}^T M'_{\lambda} \approx R_{00} + \sum_{i=1}^r \lambda_i \left(R_{i0} + \sum_{j=1}^i \lambda_j R_{ij} \right) \quad (8.15)$$

και η πολυπλοκότητα γίνεται $\mathcal{O}(r^2 n^2)$ που συνήθως είναι συμφέρουσα (αν $N > r^2$). Το κόστος αντιστροφής του πίνακα αυτού, δηλαδή $(M'_{\lambda}^T M'_{\lambda})^{-1}$, είναι $\mathcal{O}(n^3)$.

Οι προσεγγίσεις των προτύπων κίνησης (motion templates) μπορούν να υπολογιστούν από τις εξισώσεις (8.10) και (8.11) με κόστος $\mathcal{O}(rN)$. Ωστόσο, είναι πιο συμφέρουσα η χρήση της εξίσωσης

$$M_{\lambda}(\mathbf{x}) = \frac{\partial A_{\lambda}(\mathbf{x})}{\partial \mathbf{x}} \frac{\partial \mathbf{W}(\mathbf{x})}{\partial \tilde{\mathbf{p}}} \quad (8.16)$$

όπου

$$\frac{\partial A_{\lambda}}{\partial \mathbf{x}} = \frac{\partial A_0}{\partial \mathbf{x}} + \sum_{i=1}^r \lambda_i \frac{\partial A_i}{\partial \mathbf{x}} \quad (8.17)$$

Ακόμη το κόστος υπολογισμού της εικόνας λάθους της εξίσωσης (8.13) είναι $\mathcal{O}(rN)$.

Συνεπώς, η συνολική πολυπλοκότητα του αλγορίθμου τάξης r είναι

$$\mathcal{O}(r^2 n^2 + (n+r)N + n^3)$$

ενώ για $r = 0$ η πολυπλοκότητα γίνεται

$$\mathcal{O}(n^3 + nN)$$

8.3.2 Ψευδοκώδικας Αλγορίθμου

Στον αλγόριθμο 13 συνοψίζονται τα βήματα του προσαρμοστικού αλγορίθμου ταιριάσματος ΑΑΜ.

Αλγόριθμος 13 Προσαρμοστικός Αλγόριθμος Ταιριάσματος ΑΑΜ

Input: Επιλογή τάξης r ώστε $0 \leq r \leq m$

1: Υπολογισμός gradients ∇A_0 και ∇A_i για $i = 1, \dots, r$

2: Υπολογισμός Iανωδιανής συνάρτησης *warp* και μετασχηματισμού ομοιότητας

$$\frac{\partial \mathbf{W}(\mathbf{x}; \tilde{\mathbf{p}})}{\partial \tilde{\mathbf{p}}} \Big|_{\tilde{\mathbf{p}}=0} = \left[\frac{\partial \mathbf{S}_t(\mathbf{x})}{\partial t}, \frac{\partial \mathbf{W}(\mathbf{x}; \mathbf{p})}{\partial \mathbf{p}} \right]_{\tilde{\mathbf{p}}=0}$$

3: Υπολογισμός προτύπων κίνησης $M_i(\mathbf{x}) = \frac{\partial A_i(\mathbf{x})}{\partial \mathbf{x}} \frac{\partial \mathbf{W}(\mathbf{x}; \tilde{\mathbf{p}})}{\partial \tilde{\mathbf{p}}} \Big|_{\tilde{\mathbf{p}}=0}$ για $i = 0, \dots, r$

4: Σχηματισμός $A = [A_0 \ \mathbf{1}_{N \times 1} \ A_1 \ \dots \ A_r]$ και υπολογισμός $A_{temp} = (A^T A)^{-1} A^T$

5: Υπολογισμός πίνακα προβολής $P = \mathbf{I}_{N \times N} - AA_{temp} = \mathbf{I}_{N \times N} - A(A^T A)^{-1} A^T$

6: Υπολογισμός προβολών προτύπων κίνησης $M'_i = PM_i$ για $i = 0, \dots, r$

7: **repeat**

8: Εφαρμογή συνάρτησης *warp* στην εικόνα εισόδου I : $I(\mathbf{W}(\mathbf{x}; \tilde{\mathbf{p}}))$

9: Εφαρμογή ομοπαραλληλικού μετασχηματισμού υφής: $T_u(I(\mathbf{W}(\mathbf{x}; \tilde{\mathbf{p}})))$

10: Υπολογισμός εικόνας λάθους

$$E(\mathbf{q}) = T_u(I(\mathbf{W}(\mathbf{x}; \tilde{\mathbf{p}}))) - A_{\lambda} = T_u(I(\mathbf{W}(\mathbf{x}; \tilde{\mathbf{p}}))) - \left(A_0 + \sum_{i=1}^r \lambda_i A_i \right)$$

11: Υπολογισμός στιγμιοτύπου προβολής προτύπου κίνησης $M'_{\lambda} = M'_0 + \sum_{i=1}^r \lambda_i M'_i$,

τον ανάστροφό του M'_{λ}^T και το γινόμενο $M'_{\lambda}^T M'_{\lambda}$

12: Υπολογισμός στιγμιοτύπου προτύπου κίνησης $M_{\lambda} = M_0 + \sum_{i=1}^r \lambda_i M_i$ και τον ανάστροφό του M_{λ}^T

13: Υπολογισμός διαφοράς παραμέτρων σχήματος $d\tilde{\mathbf{p}} = (M'_{\lambda}^T M'_{\lambda})^{-1} M'_{\lambda}^T E(\mathbf{q})$ και σχηματισμός των $dt = d\tilde{\mathbf{p}}_{1:4}$ και $d\mathbf{p} = d\tilde{\mathbf{p}}_{5:n}$

14: Υπολογισμός διαφοράς παραμέτρων υφής $d\tilde{\lambda}' = A_{temp} [E(\mathbf{q}) - M_{\lambda} d\tilde{\mathbf{p}}]$ και σχηματισμός των $du = d\tilde{\lambda}'_{1:2}$ και $d\lambda = d\tilde{\lambda}'_{3:(r+2)} + u_1 \lambda$

15: Ανανέωση των παραμέτρων υφής ως $\lambda \leftarrow \lambda + d\lambda$ και $T_u \leftarrow T_u \circ T_{du}^{-1}$

16: Ανανέωση των παραμέτρων σχήματος ως

$$\mathbf{W}(\mathbf{x}; \tilde{\mathbf{p}}) \leftarrow \mathbf{W}(\mathbf{x}; \tilde{\mathbf{p}}) \circ \mathbf{W}^{-1}(\mathbf{x}; d\tilde{\mathbf{p}}) \text{ ή } \tilde{\mathbf{p}} \leftarrow \tilde{\mathbf{p}} + J_{\tilde{\mathbf{p}}} d\tilde{\mathbf{p}}$$

17: **until** σύγκλιση

Τα βήματα 1-6 του αλγορίθμου αποτελούν τα βήματα εκτός βρόχου επανάληψης, δηλαδή είναι υπολογισμοί μεταβλητών που η τιμές τους παραμένουν σταθερές. Αντίθετα τα βήματα 8-16 είναι τα επαναληπτικά βήματα, δηλαδή αυτά που επαναλαμβάνονται μέχρι να υπάρξει ικανοποιητική σύγκλιση.

Στον πίνακα 8.2 συνοψίζονται τα μεγέθη των σημαντικών πινάκων που χρησιμοποιούνται στον αλγόριθμο.

Πίνακας	Μέγεθος ανά pixel	Συνολικό μέγεθος
$E(\mathbf{q})$	1	$N \times 1$
∇A_i	1×2	$N \times 2$
$\frac{\partial \mathbf{s}_t}{\partial t}$	2×4	$N \times 2 \times 4$
$\frac{\partial \mathbf{W}(\mathbf{p})}{\partial \mathbf{p}}$	$2 \times n$	$N \times 2 \times n$
$\frac{\partial \mathbf{W}(\tilde{\mathbf{p}})}{\partial \tilde{\mathbf{p}}}$	$2 \times (n + 4)$	$N \times 2 \times (n + 4)$
$M_i, M'_i, M_{\lambda}, M'_{\lambda}$	$1 \times (n + 4)$	$N \times (n + 4)$
$M'_{\lambda}^T M'_{\lambda}$	—	$(n + 4) \times (n + 4)$
P	—	$N \times N$
$A = [A_0 \ \mathbf{1} \ A_1 \ \cdots \ A_r]$	$1 \times (r + 2)$	$N \times (r + 2)$
A_{temp}	—	$(r + 2) \times N$

Πίνακας 8.2: Μεγέθη πινάκων προσαρμοστικού αλγορίθμου ταιριάσματος ΑΑΜ

8.3.3 Ανάλυση Βημάτων Εκτός Επαναληπτικού Βρόχου

Στην παράγραφο αυτή αναλύονται οι υπολογιστικές λεπτομέρειες των βημάτων 1-6 του αλγόριθμου 13. Τα βήματα αυτά ανήκουν εκτός του επαναληπτικού βρόχου και έχουν σταθερές τιμές.

8.3.3.1 Gradient Ιδιοδιανυσμάτων Υφής

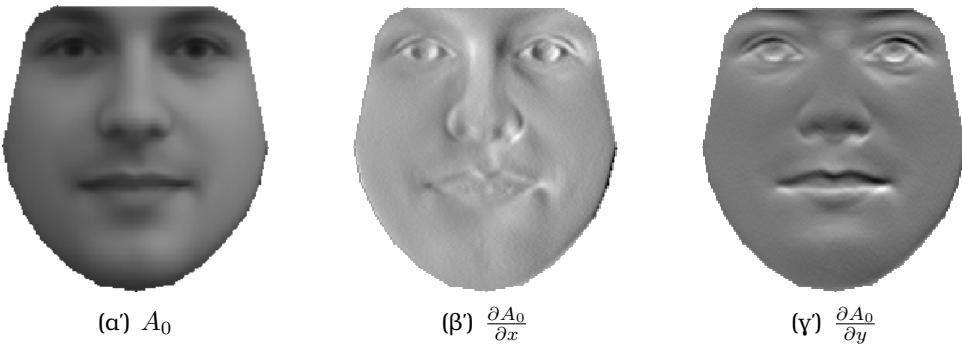
Ο υπολογισμός των gradients των ιδιοδιανυσμάτων του μοντέλου υφής, που γίνεται στο βήμα 2, επιστρέφει δυο εικόνες: την μερική παράγωγο του ιδιοδιανύσματος ως προς τη συντεταγμένη x και τη μερική παράγωγο του ιδιοδιανύσματος ως προς τη συντεταγμένη y , δηλαδή

$$\nabla A_i = \frac{\partial A_i}{\partial \mathbf{x}} = \left[\frac{\partial A_i}{\partial x}, \frac{\partial A_i}{\partial y} \right], \forall i = 0, \dots, r \quad (8.18)$$

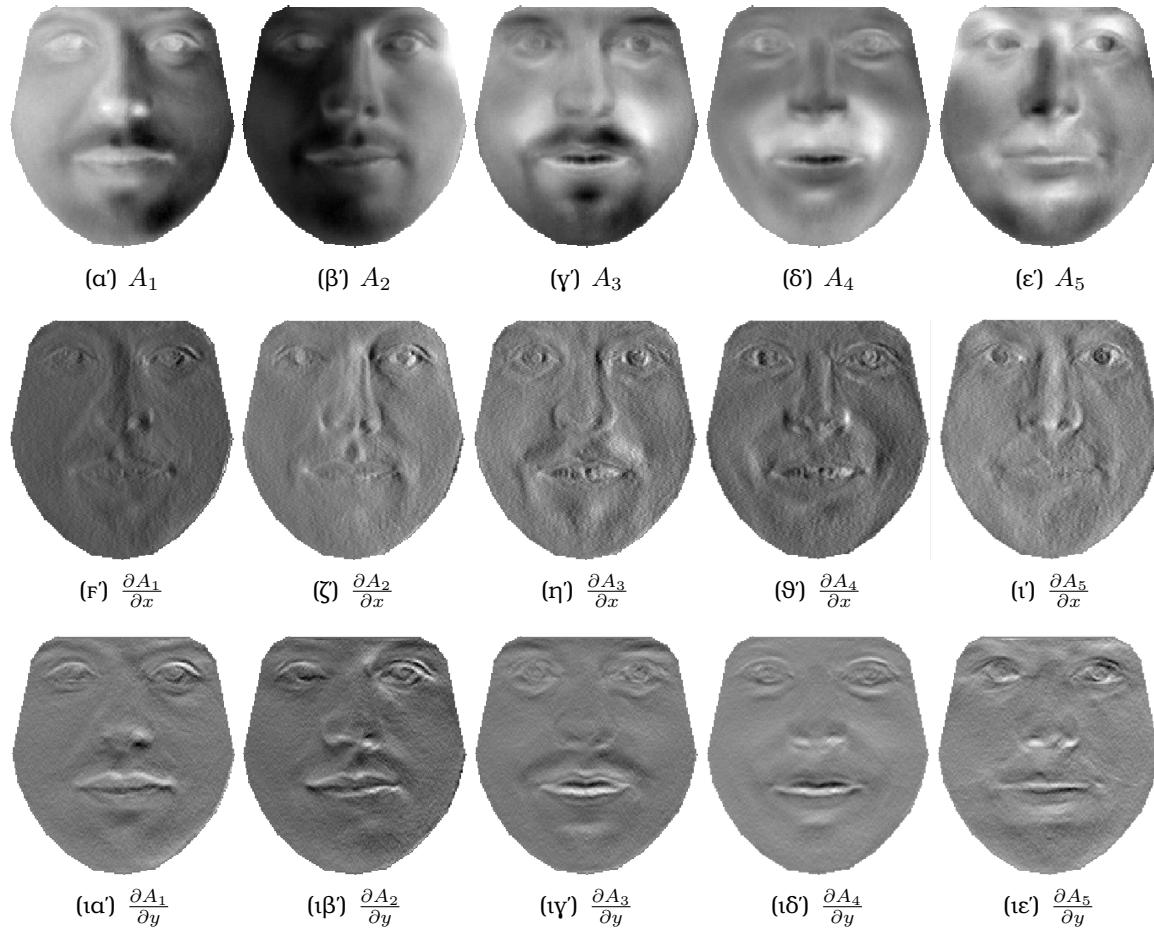
Οι διαστάσεις του gradient του κάθε ιδιοδιανύσματος είναι $N \times 2$, δηλαδή δυο διανύσματα εικόνων μήκους N που αντιστοιχούν σε κάθε συντεταγμένη.

Σημειώνεται ότι λόγω του ορισμού της μερικής παραγώγου, κατά τον υπολογισμό των εικόνων gradient οι τιμές των ακραίων pixels ορίζονται να είναι μηδενικές, ώστε να μην διαφέρουν από την μέση τιμή των υπόλοιπων pixels.

Στο Σχήμα 8.2 φαίνονται οι gradient εικόνες της μέσης υφής προσώπου, ενώ στο Σχήμα 8.3 φαίνονται οι gradient εικόνες των πέντε πιο σημαντικών ιδιοδιανυσμάτων υφής.



Σχήμα 8.2: Εικόνες gradient μέσης υφής προσώπου



Σχήμα 8.3: Εικόνες gradient ιδιοδιανυσμάτων υφής προσώπου

8.3.3.2 Ιακωβιανή Συνάρτησης Warp και Μετασχηματισμού Ομοιότητας

Στο βήμα 2 του αλγορίθμου 13 γίνεται ο υπολογισμός της Ιακωβιανής της συνάρτησης warp και του μετασχηματισμού ομοιότητας, η οποία αποτελείται από δυο επιμέρους Ιακωβιανές

$$\frac{\partial \mathbf{W}(\mathbf{x}; \tilde{\mathbf{p}})}{\partial \tilde{\mathbf{p}}} \Big|_{\tilde{\mathbf{p}}=0} = \left[\frac{\partial \mathbf{S}_t}{\partial t} \Big|_{t=0} \quad \frac{\partial \mathbf{W}(\mathbf{x}; \mathbf{p})}{\partial \mathbf{p}} \Big|_{\mathbf{p}=0} \right] \quad (8.19)$$

Όπως θα αποδειχθεί και παρακάτω η συνολική Ιακωβιανή έχει μέγεθος $2 \times (n + 4)$ ανά pixel ή αλλιώς $N \times 2 \times (n + 4)$ συνολικά.

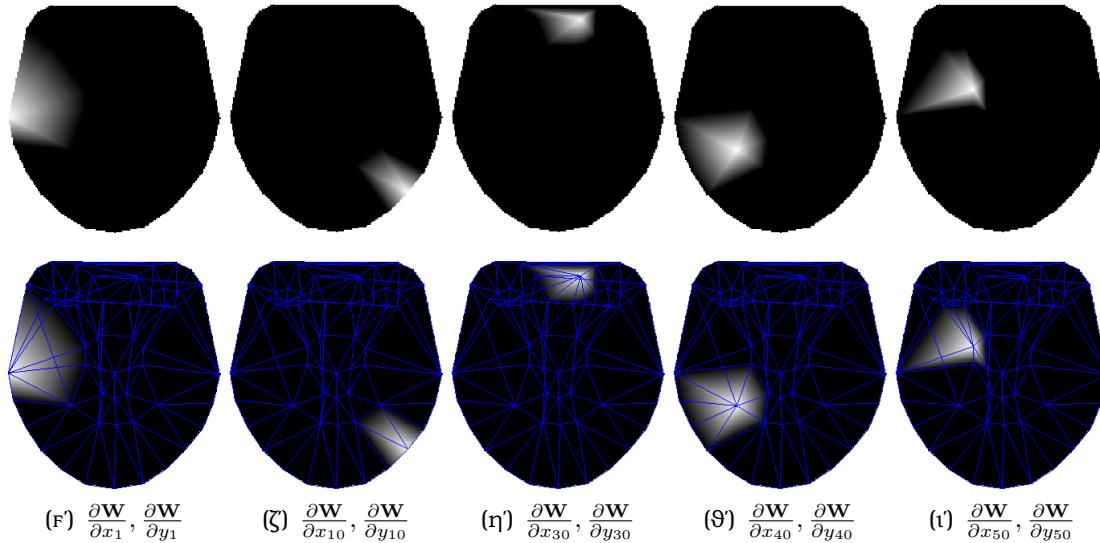
Ιακωβιανή Συνάρτησης Warp Ο υπολογισμός της Ιακωβιανής της συνάρτησης warp με εφαρμογή του κανόνα της αλυσίδας παίρνει τη μορφή

$$\frac{\partial \mathbf{W}}{\partial \mathbf{p}} \Big|_{\mathbf{p}=0} = \sum_{i=1}^L \left[\frac{\partial \mathbf{W}}{\partial x_i} \frac{\partial x_i}{\partial \mathbf{p}} + \frac{\partial \mathbf{W}}{\partial y_i} \frac{\partial y_i}{\partial \mathbf{p}} \right]_{\mathbf{p}=0} \quad (8.20)$$

όπου οι *Iακωβιανές Σχήματος*, δηλαδή οι Ιακωβιανές ως προς τις κορυφές του σχήματος \mathbf{s} , από την εξίσωση (7.13), είναι

$$\frac{\partial \mathbf{W}}{\partial x_i} = \begin{bmatrix} \alpha \\ 0 \end{bmatrix} = \begin{bmatrix} 1 - \beta - \gamma \\ 0 \end{bmatrix}, \quad \frac{\partial \mathbf{W}}{\partial y_i} = \begin{bmatrix} 0 \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ 1 - \beta - \gamma \end{bmatrix} \quad (8.21)$$

όπου οι βαρυκεντρικές συντεταγμένες α, β, γ ορίζονται από τη σχέση (7.12). Προφανώς το μέγεθος των Ιακωβιανών Σχήματος είναι 2×1 ανά pixel του μέσου σχήματος. Στο Σχήμα 8.4 φαίνονται οι Ιακωβιανές Σχήματος για τα landmark σημεία 1, 10, 30, 40 και 50 με βάση την αρίθμηση του Σχήματος 7.2 και o Delaunay τριγωνισμός.



Σχήμα 8.4: Ιακωβιανές Σχήματος landmark σημείων 1, 10, 30, 40 και 50

Οι δεύτερες Ιακωβιανές της εξίσωσης (8.20) ορίζονται ως

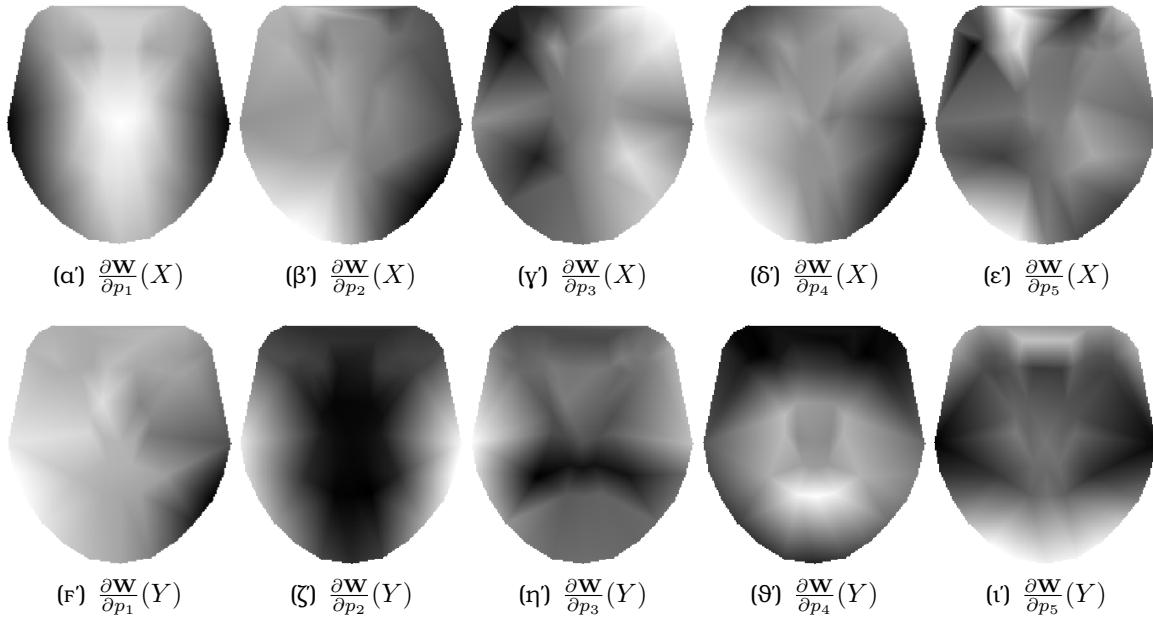
$$\frac{\partial x_i}{\partial \mathbf{p}} = (s_1^{x_i}, s_2^{x_i}, \dots, s_n^{x_i}), \frac{\partial y_i}{\partial \mathbf{p}} = (s_1^{y_i}, s_2^{y_i}, \dots, s_n^{y_i}) \quad (8.22)$$

όπου $s_j^{x_i}$ το στοιχείο του ιδιοδιανύσματος \mathbf{s}_j που αντιστοιχεί στο landmark σημείο x_i (το i -οςτό στοιχείο). Προφανώς το μέγεθος αυτών των Ιακωβιανών είναι $1 \times n$ ανά pixel του μέσου σχήματος προσώπου. Από τις εξισώσεις (8.20), (8.21) και (8.22) προκύπτει

$$\left. \frac{\partial \mathbf{W}}{\partial \mathbf{p}} \right|_{\mathbf{p}=0} = \sum_{i=1}^L \alpha \begin{bmatrix} s_1^{x_i}, s_2^{x_i}, \dots, s_n^{x_i} \\ s_1^{y_i}, s_2^{y_i}, \dots, s_n^{y_i} \end{bmatrix} \quad (8.23)$$

Συνεπώς, το μέγεθος της Ιακωβιανής της συνάρτησης warp της εξίσωσης (8.20) είναι $2 \times n$ ανά pixel ή $N \times 2 \times n$ συνολικά.

Στο Σχήμα 8.5 φαίνονται οι Ιακωβιανές της συνάρτησης warp για τα πέντε πιο σημαντικά ιδιοδιανύσματα του μοντέλου σχήματος \mathbf{s}_i , $i = 1, \dots, 5$. Η πάνω γραμμή δείχνει την X εικόνα της Ιακωβιανής και η κάτω γραμμή την Y .



Σχήμα 8.5: Ιακωβιανές συνάρτησης warp για τα πέντε πιο σημαντικά ιδιοδιανύσματα σχήματος

Ιακωβιανή Μετασχηματισμού Ομοιότητας Ο υπολογισμός της Ιακωβιανής του Μετασχηματισμού Ομοιότητας γίνεται από τη σχέση

$$\left. \frac{\partial \mathbf{S}_t}{\partial t} \right|_{t=0} = \left[\frac{\partial \mathbf{S}_t}{\partial t_1} \quad \frac{\partial \mathbf{S}_t}{\partial t_2} \quad \frac{\partial \mathbf{S}_t}{\partial t_3} \quad \frac{\partial \mathbf{S}_t}{\partial t_4} \right]_{t=0} \quad (8.24)$$

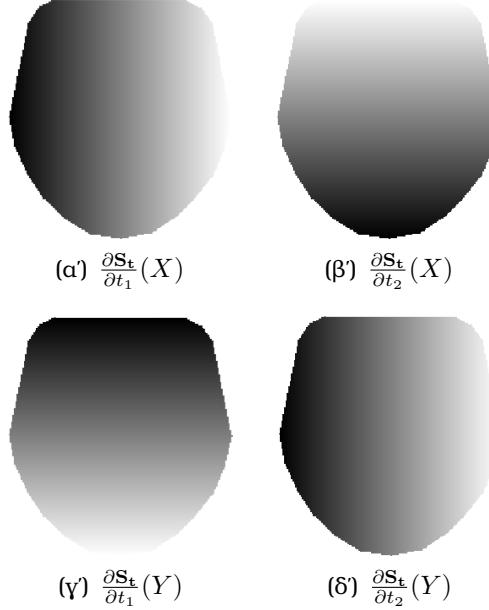
όπου σύμφωνα με την εξίσωση (7.27) προκύπτει

$$\frac{\partial \mathbf{S}_t}{\partial t_1} = \begin{bmatrix} x \\ y \end{bmatrix}, \quad \frac{\partial \mathbf{S}_t}{\partial t_2} = \begin{bmatrix} -y \\ x \end{bmatrix}, \quad \frac{\partial \mathbf{S}_t}{\partial t_3} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \frac{\partial \mathbf{S}_t}{\partial t_4} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Συνεπώς ισχύει

$$\left. \frac{\partial \mathbf{S}_t}{\partial \mathbf{t}} \right|_{\mathbf{t}=0} = \begin{bmatrix} x & -y & 1 & 0 \\ y & x & 0 & 1 \end{bmatrix} \quad (8.25)$$

Προφανώς το μέγεθος της Ιακωβιανής είναι 2×4 για κάθε pixel του μέσου σχήματος προσώπου ή $N \times 2 \times 4$ συνολικά. Στο Σχήμα 8.6 φαίνονται οι Ιακωβιανές των δυο πρώτων παραμέτρων του μετασχηματισμού ομοιότητας.



Σχήμα 8.6: Ιακωβιανές μετασχηματισμού ομοιότητας για τις δυο πρώτες παραμέτρους

Το ίδιο αποτέλεσμα θα είχε επιτευχθεί ακολουθώντας την υπολογιστική μέθοδο της Ιακωβιανής της συνάρτησης warp. Συγκεκριμένα, από τον κανόνα τις αλυσίδας ισχύει

$$\left. \frac{\partial \mathbf{S}_t}{\partial \mathbf{t}} \right|_{\mathbf{t}=0} = \sum_{i=1}^L \left[\frac{\partial \mathbf{S}_t}{\partial x_i} \frac{\partial x_i}{\partial \mathbf{t}} + \frac{\partial \mathbf{S}_t}{\partial y_i} \frac{\partial y_i}{\partial \mathbf{t}} \right]_{\mathbf{t}=0} \quad (8.26)$$

όπως στην εξίσωση (8.20). Επομένως, από τον ορισμό του μετασχηματισμού ομοιότητας της σχέσης (7.29) προκύπτει

$$\begin{aligned} \frac{\partial \mathbf{S}_t}{\partial x_i} &= \begin{bmatrix} \alpha \\ 0 \end{bmatrix}, \quad \frac{\partial \mathbf{S}_t}{\partial y_i} = \begin{bmatrix} 0 \\ \alpha \end{bmatrix} \\ \frac{\partial x_i}{\partial \mathbf{t}} &= [s_1^{*x_i}, s_2^{*x_i}, \dots, s_n^{*x_i}], \quad \frac{\partial y_i}{\partial \mathbf{t}} = [s_1^{*y_i}, s_2^{*y_i}, \dots, s_n^{*y_i}] \end{aligned}$$

8.3.3.3 Πίνακας Προβολής και Πρότυπα Κίνησης

Ο πίνακας προβολής του βήματος 5 του αλγορίθμου 13

$$P = \mathbf{I}_{N \times N} - AA_{temp} = \mathbf{I}_{N \times N} - A(A^T A)^{-1} A^T \quad (8.27)$$

έχει μέγεθος $N \times N$. Αυτός είναι ένας πίνακας πολύ μεγάλων διαστάσεων αν συνυπολιστεί το γεγονός ότι κάθε αριθμός απαιτεί 32 bits μνήμης. Επομένως, η αποθήκευση ολόκληρου του πίνακα προβολής P απαιτεί μνήμη RAM χωρητικότητας M

$$M = N \cdot N \cdot 4\text{bytes} = 4N^2\text{bytes} \stackrel{N=22477}{\Rightarrow} M \simeq 2GB$$

Η λύση της αποθήκευσης του πίνακα στον σκληρό δίσκο του υπολογιστή έχει σαν αποτέλεσμα την πολύ αργή φόρτωση των δεδομένων, το οποίο είναι ανεπιθύμητο.

Για να αντιμετωπιστεί αυτό το πρόβλημα, αντί να αποθηκευτεί ο πίνακας προβολής P , αποθηκεύονται στα βήματα εκτός του επαναληπτικού βρόχου οι πίνακες A και A_{temp} που ορίζονται ως

$$A = [A_0 \ 1_{N \times 1} \ A_1 \ \dots \ A_r] \text{ και } A_{temp} = (A^T A)^{-1} A^T$$

Οι πίνακες αυτοί έχουν μεγέθη $N \times (r + 2)$ ανδ $(r + 2) \times N$ αντίστοιχα με αποτέλεσμα να δεσμέουν περίπου 6MB μνήμης RAM συνολικά. Συνεπώς, ο υπολογισμός των προβολών των προτύπων κίνησης που βασίζεται στη σχέση

$$M'_i = P M_i, \text{ για } i = 0, \dots, r$$

γίνεται με τη διαδικασία που φαίνεται στον αλγόριθμο 14. Ο αλγόριθμος υπολογίζει τις γραμμές του πίνακα προβολής P και των προβολών των προτύπων κίνησης M'_i σε ομάδες.

Αλγόριθμος 14 Υπολογισμός προβολών προτύπων κίνησης χωρίς την αποθήκευση του πίνακα προβολής

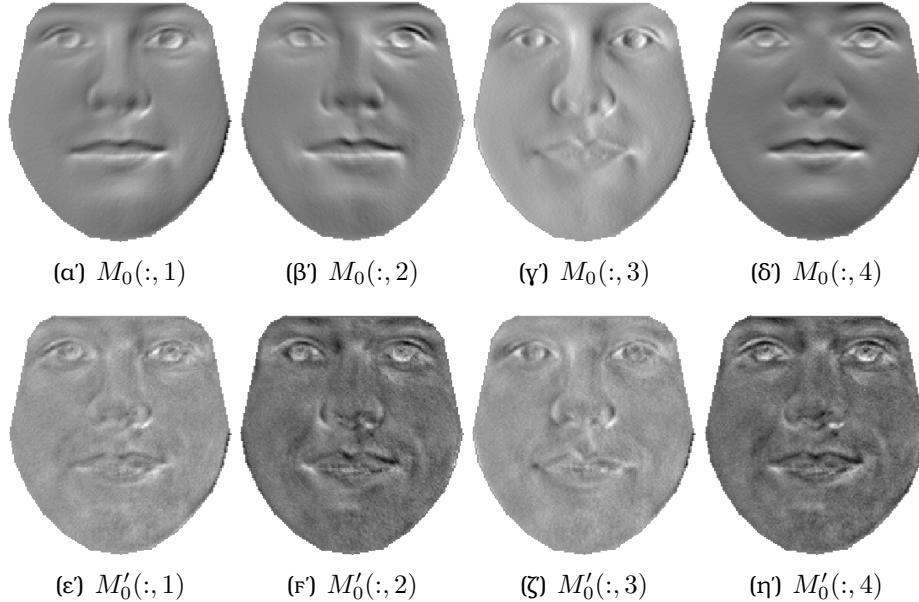
Input: Επιλογή μεταβλητής *rowGroupStep*

- 1: **for** *row* = *rowGroupStep* ... N με βήμα *rowGroupStep* **do**
 - 2: *from* = *row* - *rowGroupStep* + 1
 - 3: *to* = *row*
 - 4: $P_{temp} = A[from \dots to, 1 \dots (r + 2)] \cdot A_{temp}$
 - 5: $I = zeros(to - from + 1, N)$ πίνακας με μηδενικά μεγέθους $[(to - from + 1) \times N]$
 - 6: Τοποθέτηση μοναδιαίου πίνακα $\mathbf{I}_{(to - from + 1) \times (to - from + 1)}$ στην κατάλληλη θέση του πίνακα I : $I(1 \dots (to - from + 1), from \dots to) = eye(to - from + 1)$
 - 7: Υπολογισμός των τιμών του P για τις επιλεγμένες γραμμές: $P_{row} = I - P_{temp}$
 - 8: **for** *i* = 0 ... r **do**
 - 9: $M'_i(from \dots to, :) = P_{row} M_i$
 - 10: **end for**
 - 11: **end for**
-

Η τιμή της παραμέτρου $rowGroupStep$ επιλέγεται ανάλογα με τη χωρητικότητα της μνήμης RAM. Μετά από αρκετές δοκιμές, μια κατάλληλη αλλά ενδεικτική τιμή της παραμέτρου είναι $rowGroupStep = 600$.

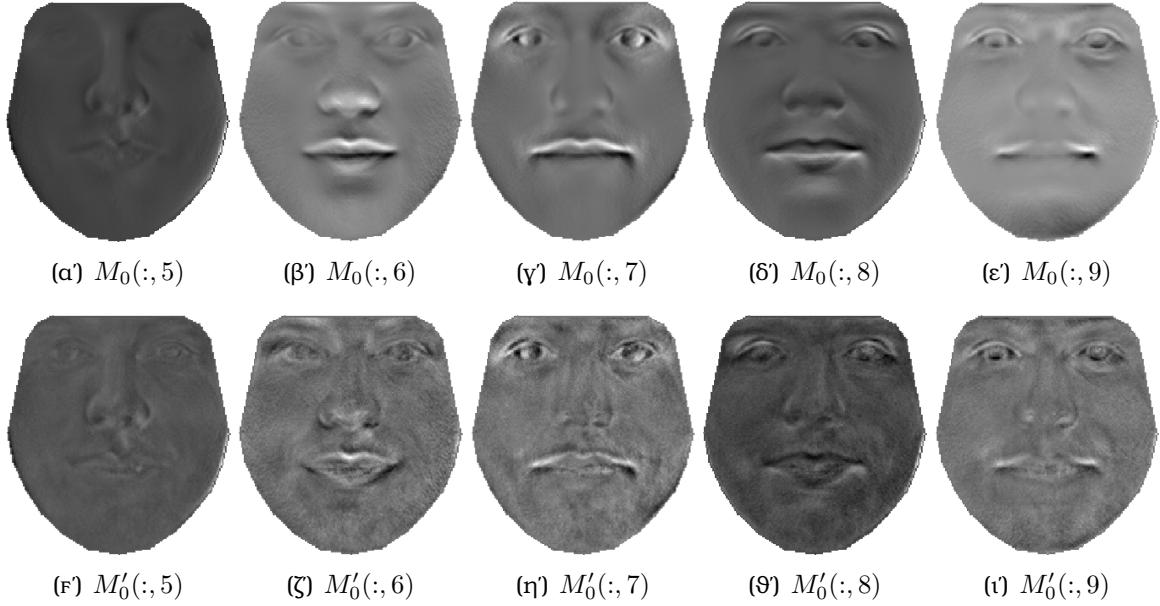
Τα πρότυπα κίνησης και οι προβολές τους είναι πίνακες διαστάσεων $N \times (n + 4)$. Ουσιαστικά λοιπόν αποτελούνται από $(n + 4)$ διανυσματικές εικόνες των N pixels, καθεμιά από τις οποίες αντιπροσωπεύει την αντίστοιχη παράμετρο του μοντέλου σχήματος και του μετασχηματισμού ομοιότητας. Επίσης, κάθε πρότυπο κίνησης M_i , για $i = 1, \dots, r$ απομονώνει την συνεισφορά του A_i ιδιοδιανύσματος υφής στα πρότυπα κίνησης, ενώ το M_0 συγκεντρώνει τα μη προσαρμοσμένα πρότυπα κίνησης που αντιστοιχούν στο μέσο διάνυσμα υφής A_0 του μοντέλου. Από την άλλη, οι προβολές των προτύπων κίνησης M'_i , για $i = 1, \dots, r$ προκύπτουν πολλάπλασιάζοντας τα M_i με τον πίνακα προβολής P , ο οποίος προβάλλει διανύσματα στον συμπληρωματικό υποχώρο που καλύπτεται από τις στήλες του πίνακα $A = [A_0 \ 1_{N \times 1} \ A_1 \ \dots \ A_m]$.

Στο Σχήμα 8.7 φαίνονται οι διανυσματικές εικόνες του προτύπου κίνησης M_0 και της προβολής του M'_0 για τις παραμέτρους του μετασχηματισμού ομοιότητας $t_{1:4}$.



Σχήμα 8.7: Πρότυπα κίνησης και οι προβολές τους για τις παραμέτρους του μετασχηματισμού ομοιότητας

Αντίστοιχα στο Σχήμα 8.8 φαίνονται οι διανυσματικές εικόνες του προτύπου κίνησης M_0 και της προβολής του M'_0 για τις πέντε πρώτες παραμέτρους σχήματος $p_{1:5}$ που αντιστοιχούν στα πέντε πιο σημαντικά ιδιοδιανύσματα του μοντέλου σχήματος.



Σχήμα 8.8: Πρότυπα κίνησης και οι προβολές τους για τις παραμέτρους σχήματος

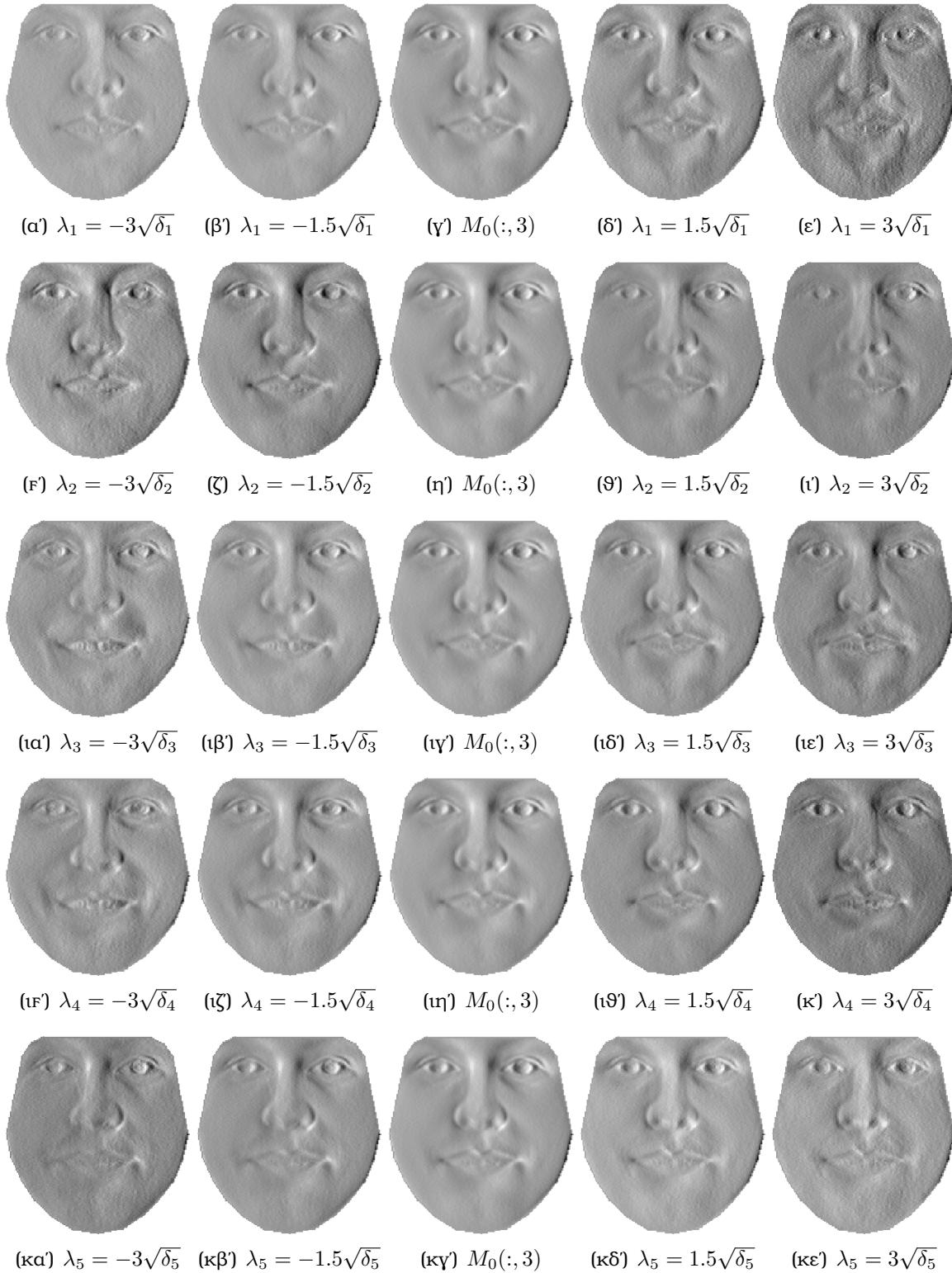
Στα βήματα 11 και 12 του αλγορίθμου 13 απαιτείται η δημιουργία του στιγμιοτύπου του προτύπου κίνησης και της προβολής του. Η σύνθεση του στιγμιοτύπου γίνεται προσθέτωντας στο M_0 ένα γραμμικό συνδυασμό των υπόλοιπων προτύπων $M_i, i = 1, \dots, r$ με βάρη τις αντίστοιχες παραμέτρους του μοντέλου υφής λ_i . Ισχύουν δηλαδή οι σχέσεις

$$M_{\lambda} = M_0 + \sum_{i=1}^r \lambda_i M_i$$

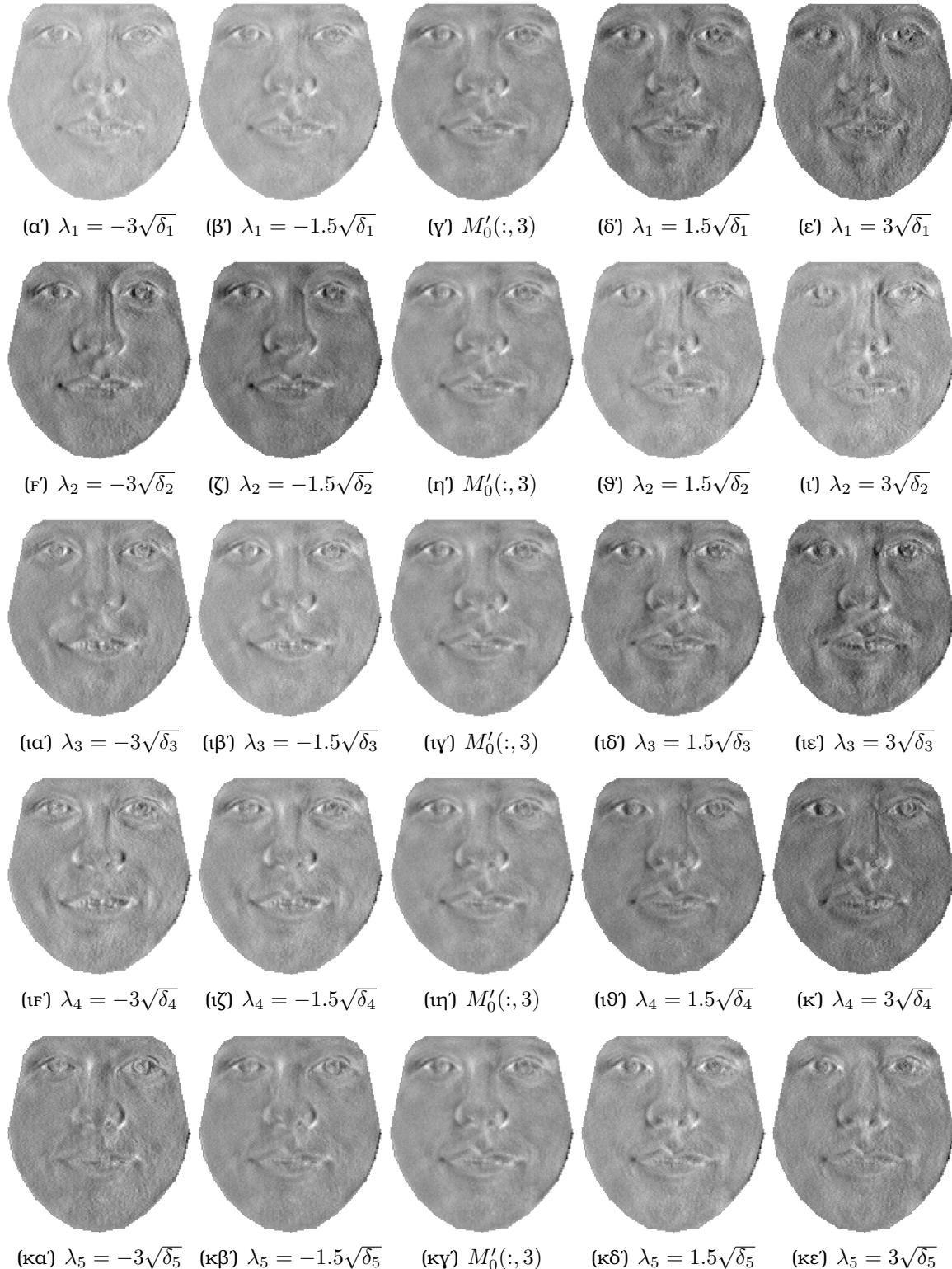
και

$$M'_{\lambda} = M'_0 + \sum_{i=1}^r \lambda_i M'_i$$

Στα Σχήματα 8.9 και 8.10 φαίνεται το στιγμιότυπο των προτύπων κίνησης και των προβολών τους αντίστοιχα μεταβάλλοντας τις πέντε πρώτες παραμέτρους υφής λ_i , που αντιστοιχούν στα πέντε πιο σημαντικά ιδιοδιανύσματα υφής, στο διάστημα $\pm \sqrt{\delta_i}$, όπου δ_i οι αντίστοιχες ιδιοτιμές. Απεικονίζεται η τρίτη στήλη των στιγμιοτύπων που αντιστοιχεί στην παράμετρο του μετασχηματισμού ομοιότητας t_3 , δηλαδή τα διανύσματα $M_{\lambda}(:, 3)$ και $M'_{\lambda}(:, 3)$.



Σχήμα 8.9: Απεικόνιση στιγμιοτύπου προτύπου κίνησης $M_\lambda(:, 3)$ μεταβάλλοντας τις πέντε πρώτες παραμέτρους υφής



Σχήμα 8.10: Απεικόνιση στιγμιοτύπου προσολής προτύπου κίνησης $M'_\lambda(:, 3)$ μεταβάλλοντας τις πέντε πρώτες παραμέτρους υφής

8.3.4 Ανάλυση Βημάτων Επαναληπτικού Βρόχου

Στην παράγραφο αυτή αναλύονται οι υπολογιστικές λεπτομέρειες των βημάτων 8-14 του αλγόριθμου 13. Τα βήματα αυτά ανήκουν στον επαναληπτικό βρόχο.

8.3.4.1 Υπολογισμός Μεταβολών Παραμέτρων Σχήματος και Υφής Προσώπου

Στα βήματα 13 και 14 του αλγορίθμου 13 γίνεται ο υπολογισμός των μεταβολών των διανύσματων παραμέτρων σχήματος και υφής προσώπου. Ο σκοπός είναι να χρησιμοποιηθούν αυτές οι μεταβολές $d\tilde{\mathbf{p}}$ και $d\tilde{\lambda}$ στα αμέσως επόμενα βήματα της ανανέωσης.

Η μεταβολή του διανύσματος των παραμέτρων σχήματος και του μετασχηματισμού ομοιότητας γίνεται με τη σχέση

$$d\tilde{\mathbf{p}} = \left(M'_{\lambda}^T M'_{\lambda} \right)^{-1} M'_{\lambda}^T P E(\mathbf{q}) \quad (8.28)$$

στην οποία γίνεται χρήση του πίνακα προβολής P . Ωστόσο για τους λόγους που αναφέρθηκαν στην παράγραφο 8.3.3.3, πρέπει ο υπολογισμός αυτός να γίνει χωρίς τον πίνακα αυτό. Γνωρίζοντας ότι ισχύει $M'_{\lambda} = PM_{\lambda}$ και $P^T = P$, προκύπτει

$$M_{\lambda}^T P = (P^T M_{\lambda})^T = (PM_{\lambda})^T = M'_{\lambda}^T$$

Επομένως, ο υπολογισμός του διανύσματος μεταβολής $d\tilde{\mathbf{p}}$ παίρνει τη μορφή

$$d\tilde{\mathbf{p}} = \left(M'_{\lambda}^T M'_{\lambda} \right)^{-1} M'_{\lambda}^T E(\mathbf{q}) \quad (8.29)$$

Τα 4 πρώτα στοιχεία του διανύσματος που προκύπτει μας δίνουν το διάνυσμα μεταβολής των παραμέτρων του μετασχηματισμού ομοιότητας και τα υπόλοιπα n στοιχεία διαμορφώνουν το διάνυσμα μεταβολής των παραμέτρων σχήματος του μοντέλου προσώπου. Επομένως, ισχύει

$$dt_{1:4} = d\tilde{\mathbf{p}}_{1:4} \text{ και } d\mathbf{p}_{1:n} = d\tilde{\mathbf{p}}_{5:(n+4)}$$

Η μεταβολή του διανύσματος των παραμέτρων υφής και του μετασχηματισμού φωτεινότητας γίνεται με τη σχέση

$$d\tilde{\lambda}' = A_{temp} [E(\mathbf{q}) - M_{\lambda} d\tilde{\mathbf{p}}] \quad (8.30)$$

δηλαδή απαιτείται η χρήση του διανύσματος της εξίσωσης (8.29). Τα 2 πρώτα στοιχεία του διανύσματος που προκύπτει μας δίνουν το διάνυσμα μεταβολής των παραμέτρων του ομοπαραλληλικού μετασχηματισμού υφής και τα υπόλοιπα r στοιχεία διαμορφώνουν το διάνυσμα μεταβολής των παραμέτρων υφής του μοντέλου προσώπου. Επομένως, ισχύει

$$d\mathbf{u}_{1:2} = d\tilde{\lambda}'_{1:2} \text{ και } d\lambda_{1:r} = d\tilde{\lambda}'_{3:(r+2)} + u_1 \lambda_{1:r}$$

8.3.4.2 Ανανέωση Παραμέτρων Υφής Προσώπου

Στο βήμα 15 του αλγορίθμου 13 γίνεται η ανανέωση των παραμέτρων υφής προσώπου ώστε να φτιαχτεί το νέο στιγμιότυπο. Συγκεκριμένα, η ανανέωση των r παραμέτρων του μοντέλου υφής προσώπου γίνεται με τον ευθύ-προσθετικό τρόπο (forwards-additive), δηλαδή

$$\lambda \leftarrow \lambda + d\lambda \quad (8.31)$$

Η ανανέωση των παραμέτρων του ομοπαραλληλικού μετασχηματισμού υφής T_u γίνεται με την ανάστροφη-συνθετική μέθοδο (inverse-compositional), δηλαδή

$$T_u \leftarrow T_u \circ {T_{du}}^{-1} \quad (8.32)$$

Για εικόνα εισόδου I η σχέση αυτή παίρνει τη μορφή

$$\left. \begin{array}{l} T_u = (u_1 + 1)I + u_2 \\ {T_{du}}^{-1} = \frac{I - du_2}{du_1 + 1} \end{array} \right\} \Rightarrow T_u \circ {T_{du}}^{-1} = T_u ({T_{du}}^{-1}) = \frac{u_1 + 1}{du_1 + 1} (I - du_2) + u_2 \quad (8.33)$$

Τονίζεται ότι ο μετασχηματισμός T_u εφαρμόζεται κατά τη δειγματοληψία της εικόνας εισόδου και όχι στο στιγμιότυπο του μοντέλου υφής προσώπου, κάτι που φαίνεται και από τη σχέση υπολογισμού της εικόνας λάθους του βήματος 10 του αλγορίθμου 13.

8.3.4.3 Ανανέωση Παραμέτρων Σχήματος Προσώπου με Priors

Όπως έχει αναφερθεί προηγουμένως, η διαδικασία ανανέωσης των παραμέτρων σχήματος του βήματος 16 του αλγορίθμου 13, μπορεί να γίνει με την ανάστροφη-συνθετική μέθοδο (inverse-compositional) της παραγράφου 8.2.3. Ωστόσο στην παράγραφο αυτή παρουσιάζεται η τεχνική της χρησιμοποίησης priors σε ανάστροφους-συνθετικούς αλγορίθμους, όπως αναλύεται από τους Παπανδρέου & Μαραγκό στο [PaMa08]. Το αποτέλεσμα της μεθόδου θα είναι η μετατροπή της inverse-compositional μεταβολής $d\tilde{p}$ στην πρώτου βαθμού ισοδύναμη προσθετική της $J_{\tilde{p}}d\tilde{p}$.

Θεωρητική Ανάλυση Χρησιμοποιώντας prior περιορισμούς η συνάρτηση λάθους που πρέπει να ελαχιστοποιηθεί έχει την εξίσωση

$$f(\mathbf{q}) = \frac{1}{2\sigma^2} \|E(\mathbf{q})\|_2^2 + Q(\mathbf{q}) \quad (8.34)$$

όπου $E(\mathbf{q})$ είναι η εικόνα λάθους και

$$Q(\mathbf{q}) = \frac{1}{2} (\mathbf{q} - \mathbf{q}_0)^T \Sigma_{\mathbf{q},0}^{-1} (\mathbf{q} - \mathbf{q}_0) \quad (8.35)$$

είναι τετραγωνική συνάρτηση που αντιστοιχή σε Gaussian prior με μέση τιμή \mathbf{q}_0 και πίνακα συνδιακύμανσης $\Sigma_{\mathbf{q},0}$. Η επαναληπτική βελτίωση της $f(\mathbf{q})$ για το ταίριασμα του μοντέλου μπορεί να γίνει με δυο τρόπους. Είτε με τον ευθή τρόπο όπου ελαχιστοποιείται η $f(\mathbf{q} + d\mathbf{q})$ με το forwards-additive διάνυσμα παραμέτρων $d\mathbf{q}$ είτε με τον

ανάστροφο τρόπο όπου ελαχιστοποιείται η $f(\mathbf{q} + J_{\mathbf{q}}d\mathbf{q})$ με το inverse-compositional διάνυσμα παραμέτρων $d\mathbf{q}$. Οι εξισώσεις υπολογισμού ανανέωσης των παραμέτρων με μέγιστο a posteriori είναι

$$\mathbf{q}_{MAP} \leftarrow \Sigma_{\mathbf{q},MAP} \left[\Sigma_{\mathbf{q},0}^{-1} \mathbf{q}_0 + (J_{\mathbf{q}} \Sigma_{\mathbf{q}} J_{\mathbf{q}}^T)^{-1} (\mathbf{q}_{MAP} + J_{\mathbf{q}} d\mathbf{q}) \right] \quad (8.36a')$$

$$\Sigma_{\mathbf{q},MAP}^{-1} = \Sigma_{\mathbf{q},0}^{-1} + (J_{\mathbf{q}} \Sigma_{\mathbf{q}} J_{\mathbf{q}}^T)^{-1} \quad (8.36b')$$

Για τον υπολογισμό της μήτρας $J_{\tilde{\mathbf{p}}}$ αρχικά γίνεται η προσέγγιση

$$\mathbf{W}(\mathbf{x}; \tilde{\mathbf{p}} + J_{\tilde{\mathbf{p}}} d\tilde{\mathbf{p}}) \approx \mathbf{W}(\mathbf{W}(\mathbf{x}; -d\tilde{\mathbf{p}}); \tilde{\mathbf{p}}) \quad (8.37)$$

Παραγωγίζοντας ως προς $d\tilde{\mathbf{p}}$ και λύνοντας με τη μέθοδο των ελαχίστων τετραγώνων προκύπτει η εκτίμηση της Ιακωβιανής

$$\begin{aligned} & \underbrace{\frac{\partial \mathbf{W}}{\partial \tilde{\mathbf{p}}}\Big|_{(\mathbf{x}; \tilde{\mathbf{p}})}}_{2 \times (n+4)} \underbrace{J_{\tilde{\mathbf{p}}}}_{(n+4) \times (n+4)} \approx - \underbrace{\frac{\partial \mathbf{W}}{\partial \mathbf{x}}\Big|_{(\mathbf{x}; \tilde{\mathbf{p}})}}_{2 \times 2} \underbrace{\frac{\partial \mathbf{W}}{\partial \tilde{\mathbf{p}}}\Big|_{(\mathbf{x}; \tilde{\mathbf{p}}=0)}}_{2 \times (n+4)} \Rightarrow \\ & \Rightarrow \underbrace{\begin{bmatrix} \frac{\partial \mathbf{W}}{\partial \tilde{\mathbf{p}}}\Big|_{(\mathbf{x}_1; \tilde{\mathbf{p}})} \\ \vdots \\ \frac{\partial \mathbf{W}}{\partial \tilde{\mathbf{p}}}\Big|_{(\mathbf{x}_L; \tilde{\mathbf{p}})} \end{bmatrix}}_{2L \times (n+4)} \underbrace{J_{\tilde{\mathbf{p}}}}_{(n+4) \times (n+4)} \approx - \underbrace{\begin{bmatrix} \frac{\partial \mathbf{W}}{\partial \mathbf{x}}\Big|_{(\mathbf{x}_1; \tilde{\mathbf{p}})} & \frac{\partial \mathbf{W}}{\partial \tilde{\mathbf{p}}}\Big|_{(\mathbf{x}_1; \tilde{\mathbf{p}}=0)} \\ \vdots & \vdots \\ \frac{\partial \mathbf{W}}{\partial \mathbf{x}}\Big|_{(\mathbf{x}_L; \tilde{\mathbf{p}})} & \frac{\partial \mathbf{W}}{\partial \tilde{\mathbf{p}}}\Big|_{(\mathbf{x}_L; \tilde{\mathbf{p}}=0)} \end{bmatrix}}_{2L \times (n+4)} \Rightarrow \\ & \Rightarrow \underbrace{\frac{\partial \mathbf{W}}{\partial \tilde{\mathbf{p}}}\Big|_{(\mathbf{x}_{1:L}; \tilde{\mathbf{p}})}}_{2L \times (n+4)} \underbrace{J_{\tilde{\mathbf{p}}}}_{(n+4) \times (n+4)} \approx - \underbrace{\frac{\partial \mathbf{W}}{\partial \mathbf{x}}\Big|_{(\mathbf{x}_{1:L}; \tilde{\mathbf{p}})}}_{(n+4) \times 2L} \odot \underbrace{\frac{\partial \mathbf{W}}{\partial \tilde{\mathbf{p}}}\Big|_{(\mathbf{x}_{1:L}; \tilde{\mathbf{p}}=0)}}_{2L \times (n+4)} \Rightarrow \\ & J_{\tilde{\mathbf{p}}} = - \underbrace{\left(\frac{\partial \mathbf{W}}{\partial \tilde{\mathbf{p}}}\Big|_{(\mathbf{x}_{1:L}; \tilde{\mathbf{p}})}^T \frac{\partial \mathbf{W}}{\partial \tilde{\mathbf{p}}}\Big|_{(\mathbf{x}_{1:L}; \tilde{\mathbf{p}})} \right)}_{(n+4) \times (n+4)}^{-1} \underbrace{\frac{\partial \mathbf{W}}{\partial \tilde{\mathbf{p}}}\Big|_{(\mathbf{x}_{1:L}; \tilde{\mathbf{p}})}^T}_{(n+4) \times 2L} \underbrace{\left(\frac{\partial \mathbf{W}}{\partial \mathbf{x}}\Big|_{(\mathbf{x}_{1:L}; \tilde{\mathbf{p}})} \odot \frac{\partial \mathbf{W}}{\partial \tilde{\mathbf{p}}}\Big|_{(\mathbf{x}_{1:L}; \tilde{\mathbf{p}}=0)} \right)}_{2L \times (n+4)} \quad (8.38) \end{aligned}$$

Η εκτίμηση ελαχίστων τετραγώνων της Ιακωβιανής έχει διαστάσεις $(n+4) \times (n+4)$. Η Ιακωβιανή $\frac{\partial \mathbf{W}}{\partial \tilde{\mathbf{p}}}\Big|_{(\mathbf{x}_{1:L}; \tilde{\mathbf{p}})}$ είναι ένας $2L \times (n+4)$ πίνακας που περιέχει τις μερικές παραγώγους υπολογισμένες στα L landmark οημεία του μέσου σχήματος προσώπου για τις παραμέτρους του διανύσματος $\tilde{\mathbf{p}}$. Το σύμβολο \odot δηλώνει έναν κατάλληλο τιμηματικό πολλαπλασιασμό πινάκων.

Υπολογιστική Ανάλυση Παρακάτω αναλύεται ο υπολογισμός των μερών της εκτίμησης ελαχίστων τετραγώνων της Ιακωβιανής $J_{\tilde{\mathbf{p}}}$, που υπολογίστηκε στην εξίσωση (8.38).

Ιακωθιανή συνάρτησης warp W ως προς $\tilde{\mathbf{p}}$ Ο υπολογισμός αυτού του όρου για ένα σημείο \mathbf{x} , εφαρμόζοντας τον κανόνα της αλυσίδας στην εξίσωση (7.31) και λαμβάνοντας υπόψην ανεξάρτητα τα διανύσματα \mathbf{t} και \mathbf{p} , έχει τη μορφή

$$\frac{\partial \mathbf{W}}{\partial \tilde{\mathbf{p}}} \Big|_{(\mathbf{x}; \tilde{\mathbf{p}})} = \left[\frac{\partial \mathbf{S}_t}{\partial \mathbf{t}} \Big|_{(\mathbf{W}(\mathbf{x}, \mathbf{p}); \mathbf{t})} \quad \left(\frac{\partial \mathbf{S}_t}{\partial \mathbf{x}} \Big|_{(\mathbf{W}(\mathbf{x}, \mathbf{p}); \mathbf{t})} \frac{\partial \mathbf{W}}{\partial \mathbf{p}} \Big|_{(\mathbf{x}; \mathbf{p})} \right) \right] \quad (8.39)$$

Από την εξίσωση (7.27) και δεδομένου ότι $\mathbf{W}(\mathbf{x}; \mathbf{p}) = (x', y')^T = \mathbf{x}_0 + \sum_{i=1}^n \mathbf{x}_i$ προκύπτει

$$\frac{\partial \mathbf{S}_t}{\partial \mathbf{t}} \Big|_{(\mathbf{W}(\mathbf{x}, \mathbf{p}); \mathbf{t})} = \begin{bmatrix} \frac{\partial \mathbf{S}_t}{\partial t_1} & \frac{\partial \mathbf{S}_t}{\partial t_2} & \frac{\partial \mathbf{S}_t}{\partial t_3} & \frac{\partial \mathbf{S}_t}{\partial t_4} \end{bmatrix}_{(\mathbf{W}(\mathbf{x}, \mathbf{p}); \mathbf{t})} = \begin{bmatrix} x' & -y' & 1 & 0 \\ y' & x' & 0 & 1 \end{bmatrix} \quad (8.40)$$

Ακόμη

$$\frac{\partial \mathbf{S}_t}{\partial \mathbf{x}} \Big|_{(\mathbf{W}(\mathbf{x}, \mathbf{p}); \mathbf{t})} = \begin{bmatrix} \frac{\partial \mathbf{S}_t}{\partial x} & \frac{\partial \mathbf{S}_t}{\partial y} \end{bmatrix}_{(\mathbf{W}(\mathbf{x}, \mathbf{p}); \mathbf{t})} = \begin{bmatrix} 1+t_1 & -t_2 \\ t_2 & 1+t_1 \end{bmatrix} \quad (8.41)$$

Συνεπώς, χρησιμοποιώντας την εξίσωση (8.23) ισχύει

$$\frac{\partial \mathbf{S}_t}{\partial \mathbf{x}} \Big|_{(\mathbf{W}(\mathbf{x}, \mathbf{p}); \mathbf{t})} \frac{\partial \mathbf{W}}{\partial \mathbf{p}} \Big|_{(\mathbf{x}; \mathbf{p})} = \alpha(1+t_1) \begin{bmatrix} s_1^{x_i} & \dots & s_n^{x_i} \\ s_1^{y_i} & \dots & s_n^{y_i} \end{bmatrix} + \alpha t_2 \begin{bmatrix} -s_1^{y_i} & \dots & -s_n^{y_i} \\ s_1^{x_i} & \dots & s_n^{x_i} \end{bmatrix} \quad (8.42)$$

Τελικά από τις εξισώσεις (8.40) και (8.42) η Ιακωθιανή της εξίσωσης (8.39) ορισμένη μόνο για τα L landmark σημεία του σχήματος (που σημαίνει $\alpha = 1$) παίρνει τη μορφή

$$\frac{\partial \mathbf{W}}{\partial \tilde{\mathbf{p}}} \Big|_{(\mathbf{x}_{1:L}; \tilde{\mathbf{p}})} = [[\mathbf{s}_p \quad \mathbf{s}_p^+ \quad \mathbf{1}_x \quad \mathbf{1}_x^+] \quad (1+t_1) [\mathbf{s}_1 \dots \mathbf{s}_n] + t_2 [\mathbf{s}_1^+ \dots \mathbf{s}_n^+]] \quad (8.43)$$

όπου $\mathbf{s}_p = \mathbf{s}_0 + \sum_{i=1}^n p_i \mathbf{s}_i$ είναι το στιγμιότυπο του μοντέλου σχήματος, \mathbf{s}^+ συμβολίζει το σχήμα \mathbf{s} με περιστροφή 90° σύμφωνα με τη φορά των δεικτών του ρολογιού και $\mathbf{1}_x = [1 \ 0 \ \dots \ 1 \ 0]^T$ είναι το σχήμα με άσους στη συντεταγμένη x και μηδενικά στη συντεταγμένη y . Το μέγεθος αυτής της Ιακωθιανής είναι $2L \times (n+4)$. Η ανάστροφη της Ιακωθιανής $\frac{\partial \mathbf{W}}{\partial \tilde{\mathbf{p}}} \Big|_{(\mathbf{x}_{1:L}; \tilde{\mathbf{p}})}$ έχει μέγεθος $(n+4) \times 2L$. Επομένως το αντεστραμμένο γινόμενο των δυο $\left(\frac{\partial \mathbf{W}}{\partial \tilde{\mathbf{p}}} \Big|_{(\mathbf{x}_{1:L}; \tilde{\mathbf{p}})} \frac{\partial \mathbf{W}}{\partial \tilde{\mathbf{p}}} \Big|_{(\mathbf{x}_{1:L}; \tilde{\mathbf{p}})} \right)^{-1}$ έχει μέγεθος $(n+4) \times (n+4)$.

Ιακωθιανή συνάρτησης warp W ως προς \mathbf{x} Αυτός ο όρος της εξίσωσης (8.38) με εφαρμογή του κανόνα της αλυσίδας στο $\mathbf{W}(\mathbf{x}, \tilde{\mathbf{p}}) = \mathbf{S}_t(\mathbf{W}(\mathbf{x}, \mathbf{p}))$ δίνει

$$\frac{\partial \mathbf{W}}{\partial \mathbf{x}} \Big|_{(\mathbf{x}; \tilde{\mathbf{p}})} = \frac{\partial \mathbf{S}_t}{\partial \mathbf{x}} \Big|_{(\mathbf{W}(\mathbf{x}, \mathbf{t}); \mathbf{t})} \frac{\partial \mathbf{W}}{\partial \mathbf{x}} \Big|_{(\mathbf{x}; \mathbf{p})} = \begin{bmatrix} 1+t_1 & -t_2 \\ t_2 & 1+t_1 \end{bmatrix} \frac{\partial \mathbf{W}}{\partial \mathbf{x}} \Big|_{(\mathbf{x}; \mathbf{p})}$$

Από τον ορισμό της συνάρτησης piecewise affine warp της παραγράφου 7.2.2.2 με παραμέτρους $\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6$ ως

$$\mathbf{W}(\mathbf{x}, \mathbf{p}) = \begin{bmatrix} \alpha_1 + \alpha_2 x + \alpha_3 y \\ \alpha_4 + \alpha_5 x + \alpha_6 y \end{bmatrix}$$

αποδεικνύεται ότι η Ιακωβιανή της συνάρτησης warp ως προς \mathbf{x} για τα landmark σημεία $\mathbf{x} = \mathbf{x}_l$, $l = 1, \dots, L$ είναι

$$\left. \frac{\partial \mathbf{W}}{\partial \mathbf{x}} \right|_{(\mathbf{x}_l; \mathbf{p})} = \begin{bmatrix} \frac{\partial \mathbf{W}}{\partial x} & \frac{\partial \mathbf{W}}{\partial y} \end{bmatrix}_{(\mathbf{x}_l; \mathbf{p})} = \begin{bmatrix} \alpha_2 & \alpha_3 \\ \alpha_5 & \alpha_6 \end{bmatrix}$$

Συνεπώς, ο συνολικός υπολογισμός αυτού του όρου της εξίσωσης έχει τη μορφή

$$\left. \frac{\partial \mathbf{W}}{\partial \mathbf{x}} \right|_{(\mathbf{x}_l; \tilde{\mathbf{p}})} = \begin{bmatrix} 1 + t_1 & -t_2 \\ t_2 & 1 + t_1 \end{bmatrix} \begin{bmatrix} \alpha_2 & \alpha_3 \\ \alpha_5 & \alpha_6 \end{bmatrix}, \quad l = 1, \dots, L \quad (8.44)$$

Τονίζεται ότι οι παράμετροι $\alpha_2, \alpha_3, \alpha_5, \alpha_6$ εξαρτώνται από το επιλεγμένο landmark σημείο. Προφανώς το μέγεθος αυτού του όρου είναι 2×2 για κάθε σημείο landmark.

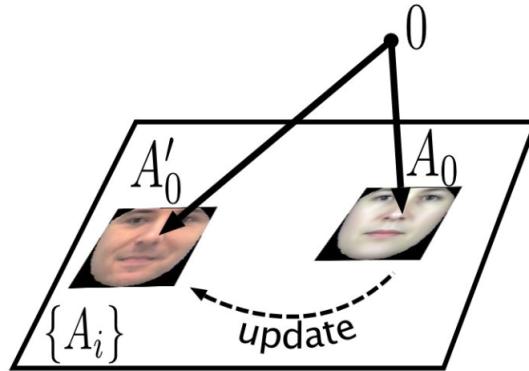
Ιακωβιανή συνάρτησης warp \mathbf{W} ως προς $\tilde{\mathbf{p}}$ στο $\tilde{\mathbf{p}} = 0$ Ο υπολογισμός αυτής της Ιακωβιανής με βάση την εξίσωση (8.43) για ένα σημείο landmark έχει τη μορφή

$$\left. \frac{\partial \mathbf{W}}{\partial \tilde{\mathbf{p}}} \right|_{(\mathbf{x}_l; \tilde{\mathbf{p}}=0)} = \left[\begin{bmatrix} x_l & -y_l & 1 & 0 \\ y_l & x_l & 0 & 1 \end{bmatrix} \left((1 + t_1) \begin{bmatrix} s_1^{x_l} & s_2^{x_l} & \dots & s_n^{x_l} \\ s_1^{y_l} & s_2^{y_l} & \dots & s_n^{y_l} \end{bmatrix} \right) \right] \quad (8.45)$$

Ο υπολογισμός αυτός πρέπει να γίνει για όλα τα landmark σημεία \mathbf{x}_l , $l = 1, \dots, L$.

8.4 Ανανέωση Διανύσματος Μέσης Υφής Προσώπου

Με τη μέθοδο των προσαρμοστικών inverse compositional AAM μοντέλων που αναλύθηκαν, το ταίριασμα του μοντέλου γίνεται με τον αλγόριθμο βαθμού r , ο οποίος επιτρέπει τη χρησιμοποίηση r ιδιοδιανυσμάτων του μοντέλου υφής για τη μείωση της διαφοράς μεταξύ της εικόνας εισόδου και της υφής του μοντέλου. Ωστόσο μια ακόμα διαδικασία που βοηθάει στο ταχύτερο ταίριασμα είναι να προσαρμοστεί η μέση υφή προσώπου A_0 του μοντέλου ώστε να είναι μικρότερη η διαφορά της με την εικόνα εισόδου. Αυτό φαίνεται και στο Σχήμα 8.11.



Σχήμα 8.11: Ανανέωση μέσης υφής προσώπου από το διάνυσμα A_0 στο A'_0

Ο σκοπός της διαδικασίας αυτής είναι να γίνουν οι μη χρησιμοποιημένες παράμετροι λ_i για $i > r$ αρκετά μικρές ώστε το λάθος του αλγορίθμου βαθμού r να γίνει αμελητέο. Μάλιστα, στην ιδανική περίπτωση που οι μη χρησιμοποιημένες παράμετροι γίνουν μηδενικές, τότε ο αλγόριθμος είναι ακριβής. Η ανανέωση του διανύσματος μέσης υφής προσώπου A_0 γίνεται με την προσαρμοσμένη μορφή του

$$A'_0 = A_0 + \sum_{i=r+1}^m \lambda_i A_i \quad (8.46)$$

Το διάνυσμα A'_0 έχει μικρότερη διαφορά από την εικόνα εισόδου ενώ ταυτόχρονα βρίσκεται μέσα στον ομοπαραλληλικό χώρο (affine space) $A_0 + \text{span}\{A_i\}$. Συνεπώς, το εύρος αναπαράστασης του μοντέλου υφής παραμένει ίδιο και μειώνεται η διαφορά του προτύπου κίνησης του αλγορίθμου βαθμού r .

Στην περίπτωση που χρησιμοποιούνται οι πίνακες συσχέτισης R_{ij} των εξισώσεων (8.14), απαιτείται η ανανέωση αυτών που εξαρτώνται από το διάνυσμα A'_0 και χρησιμοποιούνται από την προσεγγιστική σχέση (8.15). Η ανανέωση γίνεται με τις σχέσεις

$$\begin{aligned} R'_{i0} &= R_{i0} + \sum_{j=r+1}^m \lambda_j R_{ij}, \quad \text{για } i = 1, \dots, r \\ R'_{00} &= R_{00} + \sum_{i=r+1}^m \lambda_i \left(R_{i0} + \sum_{j=r+1}^i \lambda_j R_{ij} \right) \end{aligned} \quad (8.47)$$

Η πολυπλοκότητα της διαδικασίας αυτής είναι

$$\mathcal{O}(m(m-r)n^2)$$

Επομένως, η πολυπλοκότητα είναι αντιστρόφως ανάλογη του βαθμού r και όταν $r = m$ (SIC αλγόριθμος) το κόστος μηδενίζεται.

8.5 Πειραματικά Αποτελέσματα

Στην παράγραφο αυτή παρουσιάζονται τα πειραματικά αποτελέσματα του προσαρμοστικού αλγορίθμου ταιριάσματος ΑΑΜ.

8.5.1 Παράμετροι Επιλογής Χρήστη

Στα προηγούμενα κεφάλαια έγινε εκτενής ανάλυση της μεθόδου ανίχνευσης προσώπου, της διαδικασίας εκπαίδευσης του μοντέλου ΑΑΜ και του προσαρμοστικού αλγορίθμου ταιριάσματος του ΑΑΜ. Το πιο σημαντικό στοιχείο της υλοποίησης του συστήματος είναι ότι έχει δοθεί ελευθερία αποφάσεων στο χρήστη της εφαρμογής. Συγκεκριμένα, υπάρχουν έξι διαφορετικές παράμετροι, που καθεμιά επηρεάζει διαφορετικό στάδιο της συνολικής διαδικασίας. Οι τιμές που δίνει ο χρήστης σε αυτές τις παραμέτρους καθορίζουν κατά κύριο λόγο την ταχύτητα εκτέλεσης της μεθόδου και την ακρίβεια των αποτελεσμάτων της. Παρακάτω παρουσιάζονται οι παράμετροι επιλογής προς τον χρήστη.

Τάξη r του αλγορίθμου ταιριάσματος: Αυτή είναι η πιο σημαντική παράμετρος. Καθορίζει τον αριθμό των ιδιοδιανυσμάτων του μοντέλου υφής που χρησιμοποιούνται στη διαδικασία του ταιριάσματος. Όσο μεγαλύτερη είναι η τιμή του r , τόσο ο αλγόριθμος γίνεται πιο αργός αλλά πιο ακριβής. Αντίθετα, όσο μικρότερη είναι η τιμή του r , τόσο ταχύτερος γίνεται ο αλγόριθμος αλλά λιγότερο ακριβής στα αποτελέσματά του. Εκτενής παρουσιάση της παραμέτρου αυτής γίνεται στην παράγραφο 8.3.1.1. Οι τιμές της τάξης r κυμαίνονται στο διάστημα $[0, m]$.

Καθολική Κλίμακα μοντέλου AAM s_{global} : Αυτή η παράμετρος είναι πολύ σημαντική για την ταχύτητα εκτέλεσης του αλγορίθμου ταιριάσματος. Η τιμή της καθολικής κλίμακας επηρεάζει τον αριθμό των pixels N στο εσωτερικό του μέσου σχήματος προσώπου, δηλαδή την ανάλυσή του. Το μέσο σχήμα προσώπου είναι το πρότυπο στο οποίο γίνεται ο υπολογισμός της εικόνας λάθους, δηλαδή είναι το σχήμα αναφοράς. Έτσι, η ανάλυσή του παίζει σημαντικό ρόλο στη διαδικασία της δειγματοληψίας της εικόνας εισόδου και καθορίζει τη μνήμη RAM που απαιτεί ο αλγόριθμος ταιριάσματος. Ακόμη, είναι προφανές ότι μειώνονται οι διαστάσεις των διανυσμάτων και των πινάκων που αφορούν στην υφή του προσώπου, όπως τα ιδιοδιανύσματα, οι Ιακωβιανές και τα πρότυπα κίνησης.

Η παράμετρος s_{global} παίρνει τιμές στο διάστημα $[0.1, 1]$, διότι ο σκοπός είναι η μείωση της ανάλυσης ώστε να έχουμε λιγότερα pixels N . Η αρχική ανάλυση από την εκπαίδευση με τη βάση δεδομένων IMM είναι $N = 22477$ pixels εσωτερικά του μέσου σχήματος προσώπου. Έτσι με τη μείωση της καθολικής κλίμακας, ο αριθμός των pixels προσεγγίζεται με τη σχέση

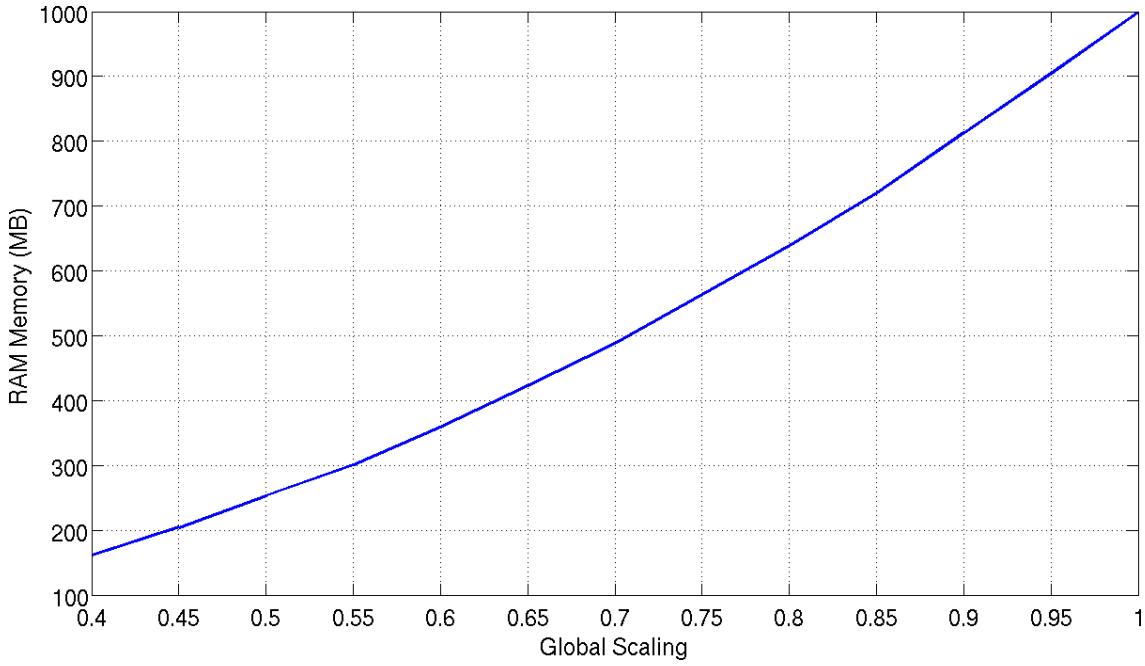
$$N \approx 22477 s_{global}^2$$

Ο πίνακας 8.3 δείχνει τη σχέση αυτή μεταξύ της καθολικής κλίμακας και του αριθμού των pixels. Στο Σχήμα 8.12 φαίνεται η καμπύλης της μνήμης RAM που απαιτούν οι μεταβλητές εκτός επαναληπτικού βρόχου και το μοντέλο AAM ως προς την τιμή της καθολικής κλίμακας s_{global} .

s_{global}	N pixels	s_{global}	N pixels
1.00	22477	0.65	9532
0.95	20356	0.60	8099
0.90	18281	0.55	6776
0.85	16215	0.50	5709
0.80	14380	0.45	4607
0.75	12699	0.40	3644
0.70	10996	0.35	2753

Πίνακας 8.3: Σχέση καθολικής κλίμακας και ανάλυσης του μέσου σχήματος προσώπου

Το μεγάλο πλεονέκτημα της παραμέτρου αυτής είναι η μεγάλη βελτίωση της ταχύτητας του αλγορίθμου ταιριάσματος με πολύ μικρή μείωση της ακρίβειας. Ωστόσο



Σχήμα 8.12: Καμπύλη απαιτούμενης μνήμης RAM ως προς την τιμή της καθολικής κλίμακας

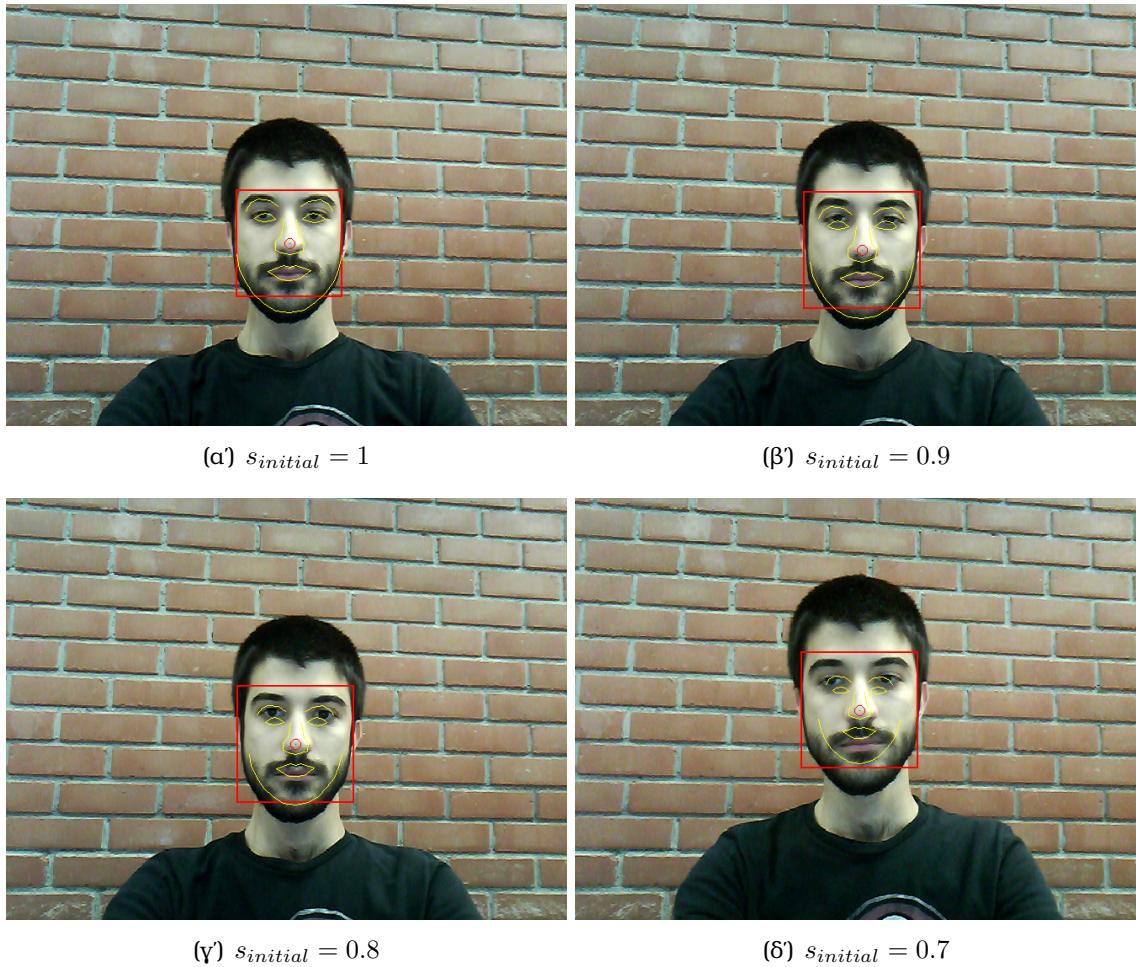
το μειονέκτημα είναι ότι απαιτεί την εκ νέου εκπαίδευση του μοντέλου AAM. Για λόγους ευκολίας του χρήστη, η εκπαίδευση έχει εκτελεστεί και αποθηκευτεί για τις τιμές του s_{global} που φαίνονται στον πίνακα 8.3, με αποτέλεσμα να απαιτείται μόνο η εκ νέου φόρτωση των αρχείων δεδομένων από το σκληρό δίσκο.

Κατώφλι Ανίχνευσης Προσώπου θ_{V-J} και Κλίμακα Αρχικοποίησης $s_{initial}$: Αυτές οι δύο παράμετροι αντιστοιχούν στο *πρόβλημα αρχικοποίησης* του αλγορίθμου ταιριάσματος. Όπως αναλύθηκε και στο κεφάλαιο 4, η παράμετρος θ_{V-J} καθορίζει την ευαισθησία του αλγορίθμου ανίχνευσης προσώπου Viola-Jones και παίρνει τιμές στο διάστημα $[-7, 3]$. Η τιμή του πρέπει να επιλεχθεί κατάλληλα ώστε να είναι σίγουρη η ανίχνευση όλων των προσώπων στην εικόνα εισόδου. Σε περιπτώσεις εικόνων με πολύπλοκο παρασκήνιο χρειάζεται και η χρησιμοποίηση του ανιχνευτή δέρματος, όπως εξηγήθηκε στο κεφάλαιο 5.

Η παράμετρος της κλίμακας αρχικοποίησης $s_{initial}$ καθορίζει τις αρχικές τιμές των παραμέτρων του μετασχηματισμού ομοιότητας t σύμφωνα με το αποτέλεσμα της ανίχνευσης προσώπου. Συγκεκριμένα, δίνεται κατάλληλη κλίμακα και μετατόπιση στο μέσο σχήμα προσώπου ώστε να ταιριάζει μέσα στο τετράγωνο που επιστρέφει ο ανιχνευτής προσώπου. Ωστόσο, κατόπιν πειραματικών δοκιμών έχει παρατηρηθεί ότι σε κάποιες εικόνες εισόδου αν μικρύνει η αρχική κλίμακα του μέσου σχήματος προσώπου τότε ο αλγόριθμος ταιριάσματος είναι πιο αποτελεσματικός. Έτσι η παράμετρος $s_{initial}$ αλλάζει την κλίμακα του s_0 ώστε να μικρύνει παίρνοντας τιμές στο διάστημα $[0.7, 1]$.

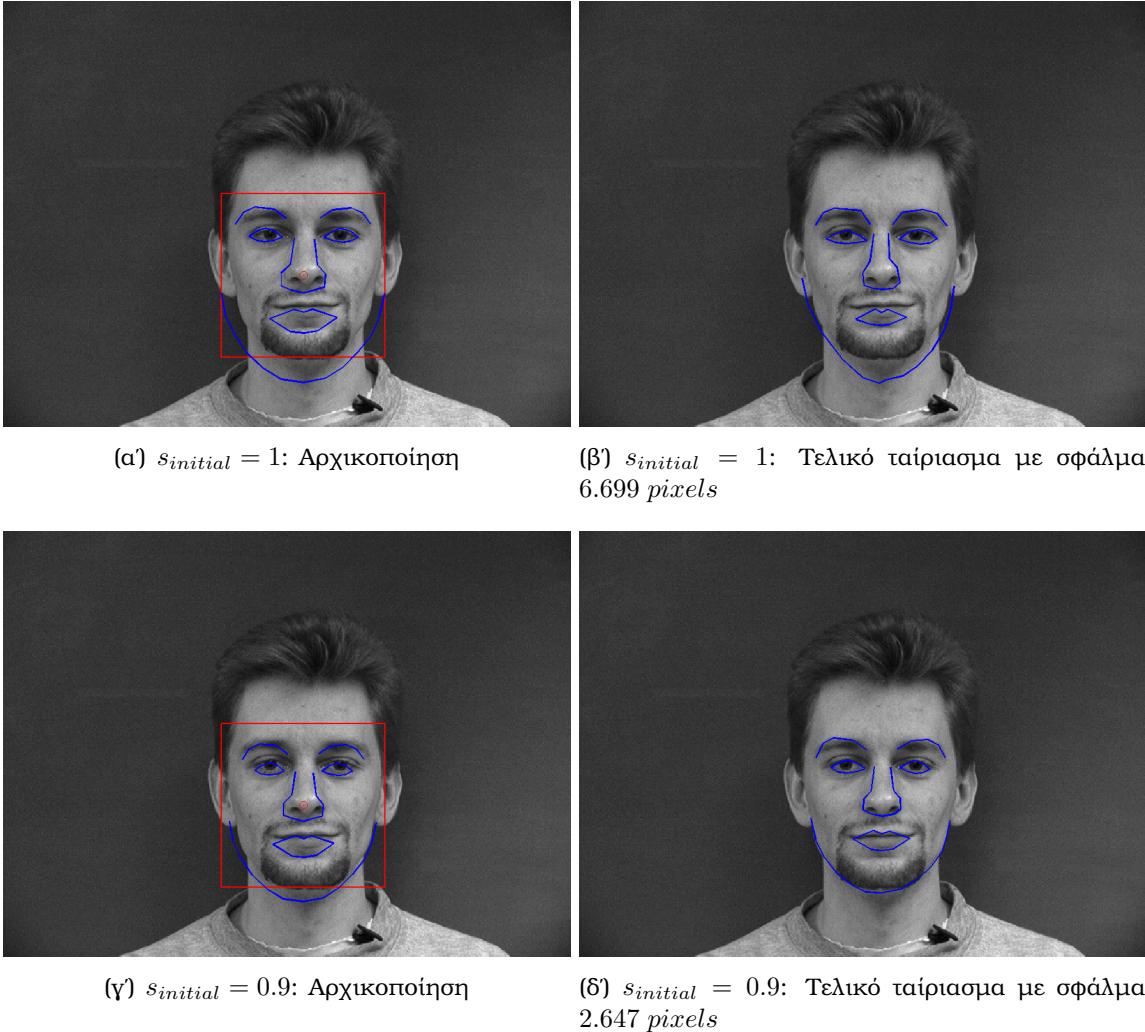
Στο Σχήμα 8.13 φαίνεται ένα παράδειγμα αρχικοποίησης του μέσου σχήματος

προσώπου για τιμές της $s_{initial}$ παραμέτρου 0.7, 0.8, 0.9 και 1. Στο Σχήμα 8.14 φαίνεται ένα παραδειγμα του αλγορίθμου ταιριάσματος που αποτυγχάνει για $s_{initial} = 1$ και συγκλίνει για $s_{initial} = 0.9$.



Σχήμα 8.13: Παραδείγματα αρχικοποίησης του αλγορίθμου ταιριάσματος για $s_{initial} = 1, 0.9, 0.8, 0.7$

Μέγιστος Αριθμός Επαναλήψεων M_{iter} και Ελάχιστη Μετακίνηση Σημείων Σχήματος M_{displ} : Οι δύο αυτές παράμετροι καθορίζουν τον αριθμό των επαναλήψεων που εκτελεί ο αλγόριθμος ταιριάσματος για μια εικόνα εισόδου, δηλαδή αποτελούν τη συνδήκη τερματισμού. Πιο συγκεκριμένα, ο αλγόριθμος σταματά όταν οι επαναλήψεις έχουν φτάσει την τιμή M_{iter} ή η μέγιστη μετακίνηση των σημείων του σχήματος που κάνει το ταίριασμα έχει φτάσει την τιμή M_{displ} . Η παράμετρος του Μέγιστου Αριθμού Επαναλήψεων παίρνει τιμές στο διάστημα $M_{iter} \in [8, 12]$ και η παράμετρος της Ελάχιστης Μετακίνησης Σημείων Σχήματος στο διάστημα $M_{displ} \in [1, 4]$.



Σχήμα 8.14: Παράδειγμα εκτέλεσης αλγορίθμου ταιριάσματος με τιμές αρχικοποίησης $s_{initial} = 1$ και 0

8.5.2 Ποσοστό Επιτυχίας

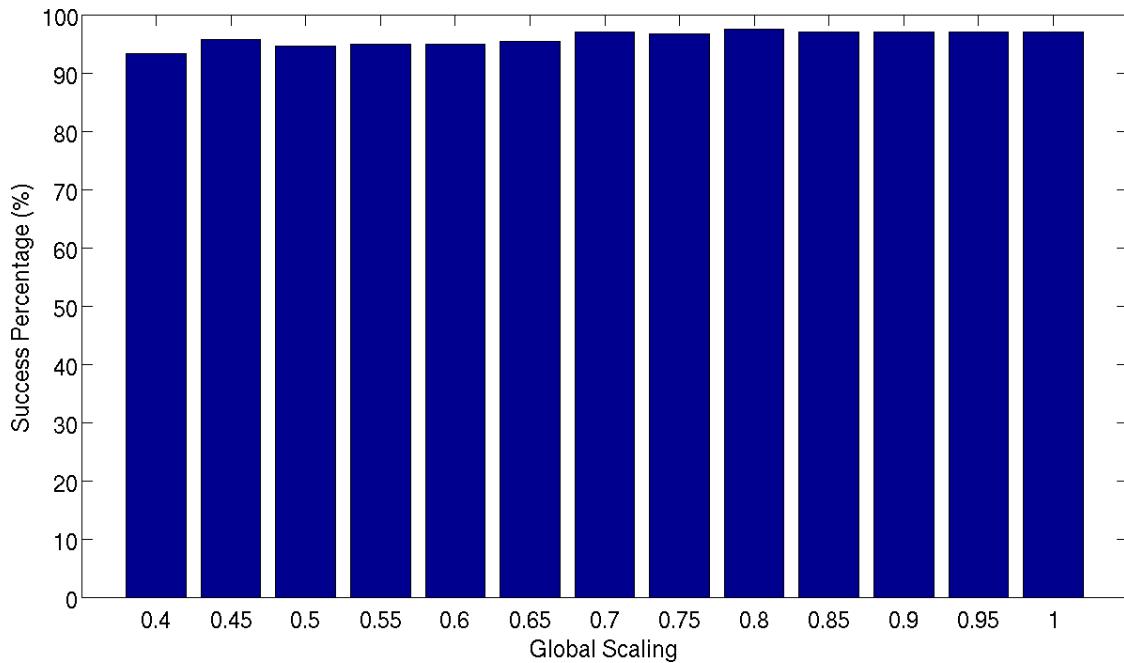
Όπως έχουμε προαναφέρει, η εκπαίδευση του μοντέλου AAM έγινε στη βάση δεδομένων IMM [NLSS04]. Η βάση δεδομένων αποτελείται από 6 εικόνες από 40 άτομα, δηλαδή περιέχει ένα σύνολο 240 εικόνων. Οι 6 εικόνες ανά άτομο εξηγούνται στον Πίνακα 8.4.

Για να βρεθεί το ποσοστό επιτυχίας του αλγορίθμου το μοντέλο AAM εκπαιδεύτηκε για τις εικόνες 30 ατόμων από τη βάση δεδομένων (σύνολο 180 εικόνων) και το ταίριασμα έγινε για τις εικόνες των υπόλοιπων 10 ατόμων (σύνολο 40 εικόνων). Οι άξονες μεταβλητήτας του μοντέλου σχήματος που προέκυψαν είναι $n = 21$ και οι αντίστοιχοι του μοντέλου υφής προσώπου παίρνουν τιμές στο διάστημα $m = \{91, 92, 93\}$ ανάλογα με την τιμή της καθολικής κλίμακας s_{global} . Οι τιμές της καθολικής κλίμακας

Εικόνα	Πόζα	Έκφραση	Φωτισμός
1	Μετωπική	ουδέτερη	διάχυτος
2	Μετωπική	χαρούμενη	διάχυτος
3	Περιστροφή 30° προς τα δεξιά	ουδέτερη	διάχυτος
4	Περιστροφή 30° προς τα αριστερά	ουδέτερη	διάχυτος
5	Μετωπική	ουδέτερη	από τα αριστερά
6	Μετωπική	αυθαίρετη	διάχυτος

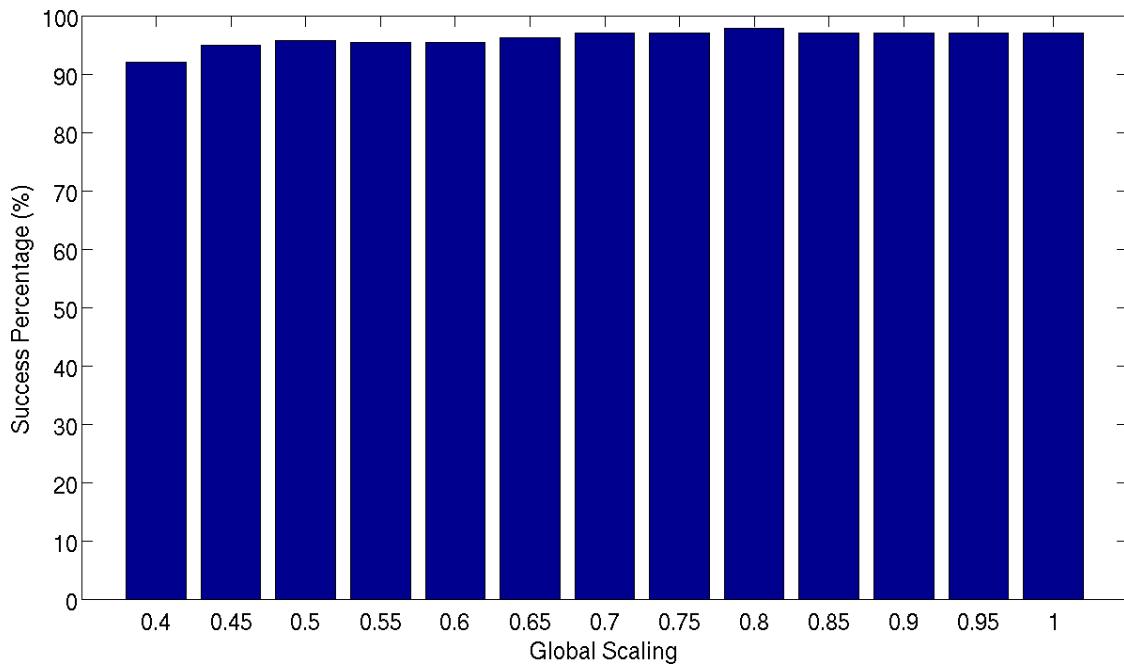
Πίνακας 8.4: Επεξήγηση εικόνων της βάσης δεδομένων IMM

που χρησιμοποιήθηκαν είναι $s_{global} = 0.4, \dots, 1$ με βήμα 0.05. Η τιμή της παραμέτρου κατωφλίου του ανιχνευτή προσώπου που χρησιμοποιείται είναι $\theta_{V-J} = -1$. Τέλος ο αλγόριθμος ταιριάσματος εκτελείται για τιμές της τάξης του $r = \{m, \frac{m}{2}, 0\}$. Τονίζεται ότι τα ποσοστά επιτυχίας που παρουσιάζονται παρακάτω για $r = 0$ δεν χρησιμοποιούν τη βελτιστοποίηση της ανανέωσης του διανύσματος μέσης υφής προσώπου που αναλύθηκε στην παράγραφο 8.4.

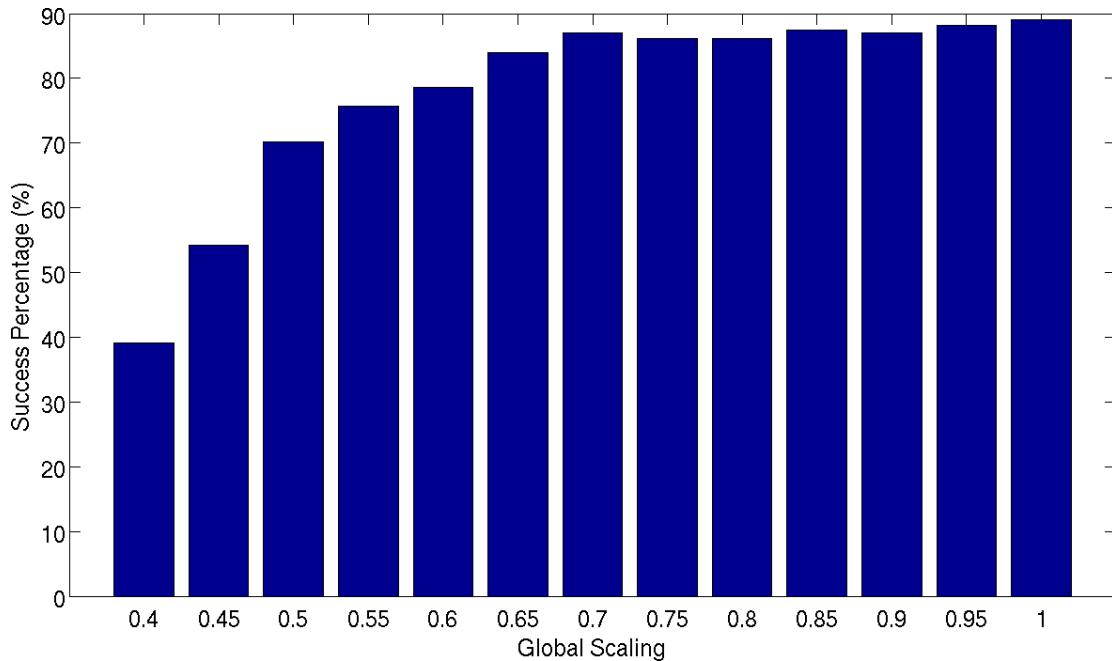


Σχήμα 8.15: Ποσοστό επιτυχίας ως προς την καθολική κλίμακα s_{global} για $r = m$

Για κάθε εικόνα εισόδου, ο αλγόριθμος εκτελείται για τιμές της παραμέτρου της κλίμακας αρχικοποίησης $s_{initial} = \{0.8, 0.9, 1\}$. Θεωρείται ότι ο αλγόριθμος είναι επιτυχημένος για μια εικόνα εισόδου αν έχει συγκλίνει σε τουλάχιστον μια από τις τρεις αυτές τιμές της κλίμακας αρχικοποίησης $s_{initial}$. Για κάθε εικόνα εισόδου έχουμε τις αληθινές συντεταγμένες των $L = 58$ landmark σημείων του σχήματος. Έτσι το λάθος σχήματος υπολογίζεται ως η μέση τιμή όλων των διαφορών των αληθινών συντεταγμένων των landmark σημείων από τις συντεταγμένες που προέκυψαν από τον



Σχήμα 8.16: Ποσοστό επιτυχίας ως προς την καθολική κλίμακα s_{global} για $r = \frac{m}{2}$



Σχήμα 8.17: Ποσοστό επιτυχίας ως προς την καθολική κλίμακα s_{global} για $r = 0$

αλγόριθμο. Θεωρείται ότι το ταίρασμα έγινε με επιτυχία αν το λάθος σχήματος έχει τιμή $ShapeError \leq 10$ pixels. Αντίστοιχα, το λάθος υφής προσώπου υπολογίζεται ως η νόρμα της εικόνας λάθους $\|E(\mathbf{q})\|$.

Ο Πίνακας 8.5 δείχνει τα ποσοστά επιτυχίας για της μορφές πειραμάτων που εξη-

γήθηκαν. Στα Σχήματα 8.15, 8.16 και 8.17 φαίνονται τα διαγράμματα του ποσοστού επιτυχίας του αλγορίθμου ως προς την καθολική κλίμακα s_{global} για τις τιμές τάξης $r = \{m, \frac{m}{2}, 0\}$.

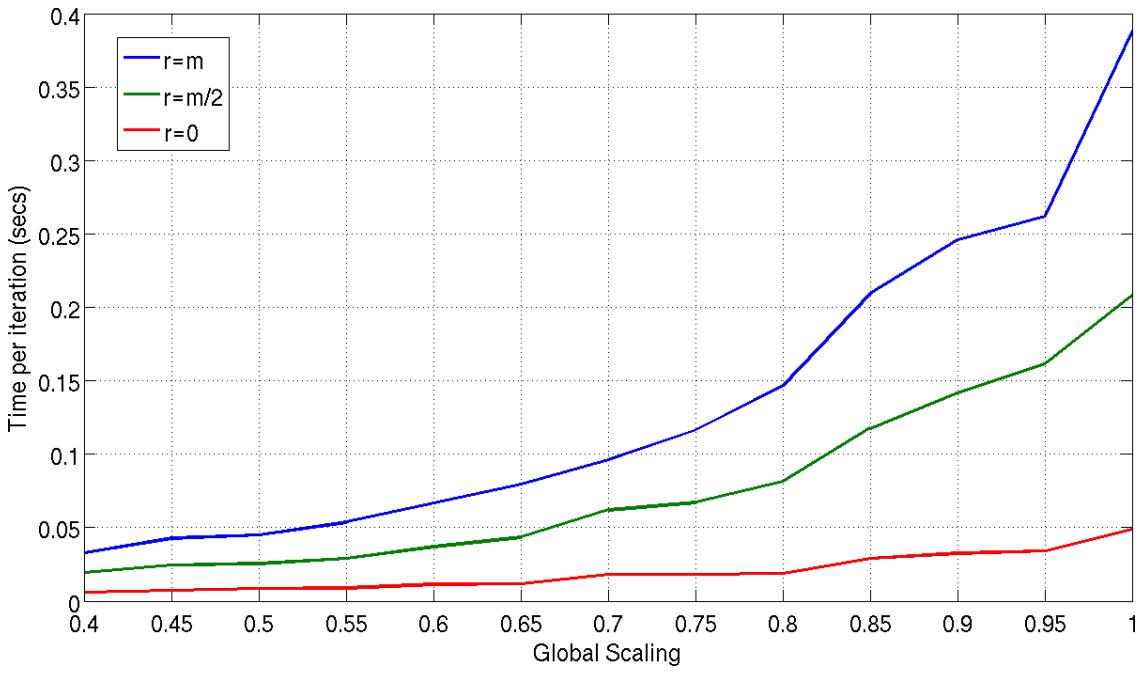
s_{global}	N	r	Ποσοστό Επιτυχίας (%)	s_{global}	N	r	Ποσοστό Επιτυχίας (%)
1	22477	93	97.06	0.65	9532	92	95.38
		47	97.06			46	96.22
		0	89.08			0	84.03
0.95	20356	92	97.06	0.6	8099	92	94.96
		46	97.06			46	95.38
		0	88.24			0	78.57
0.9	18281	93	97.06	0.55	6776	92	94.96
		47	97.06			46	95.38
		0	86.97			0	75.63
0.85	16215	93	97.06	0.5	5709	92	94.54
		47	97.06			46	95.80
		0	87.40			0	70.17
0.8	14380	93	97.48	0.45	4607	92	95.80
		47	97.90			46	94.96
		0	86.13			0	54.20
0.75	12699	92	96.64	0.4	3644	91	93.28
		46	97.06			46	92.02
		0	86.13			0	39.08
0.7	10996	92	97.06				
		46	97.06				
		0	86.97				

Πίνακας 8.5: Ποσοστό επιτυχίας αλγορίθμου ταιριάσματος με μέγιστη τιμή λάθους σχήματος 10 pixels

Από τα αποτελέσματα παρατηρείται ότι η μείωση της τιμής της παραμέτρου καθολικής κλίμακας s_{global} δεν επιφέρει σημαντική μείωση στην απόδοση του αλγορίθμου, παρά μόνο στην περίπτωση του $r = 0$, όπου για $s_{global} < 0.7$ το ποσοστό επιτυχίας πέφτει σχεδόν εκθετικά.

8.5.3 Χρονικό Κόστος

Το Σχήμα 8.18 δείχνει το διάγραμμα το χρονικού κόστους ανά επανάληψη του αλγορίθμου ταιριάσματος για τιμές της καθολικής κλίμακας $s_{global} = 0.4, \dots, 1$ με βήμα 0.05. Το διάγραμμα έχει τρεις καμπύλες καθεμιά από τις οποίες αντιστοιχεί στις τιμές της τάξης του αλγορίθμου $r = \{m, \frac{m}{2}, 0\}$.

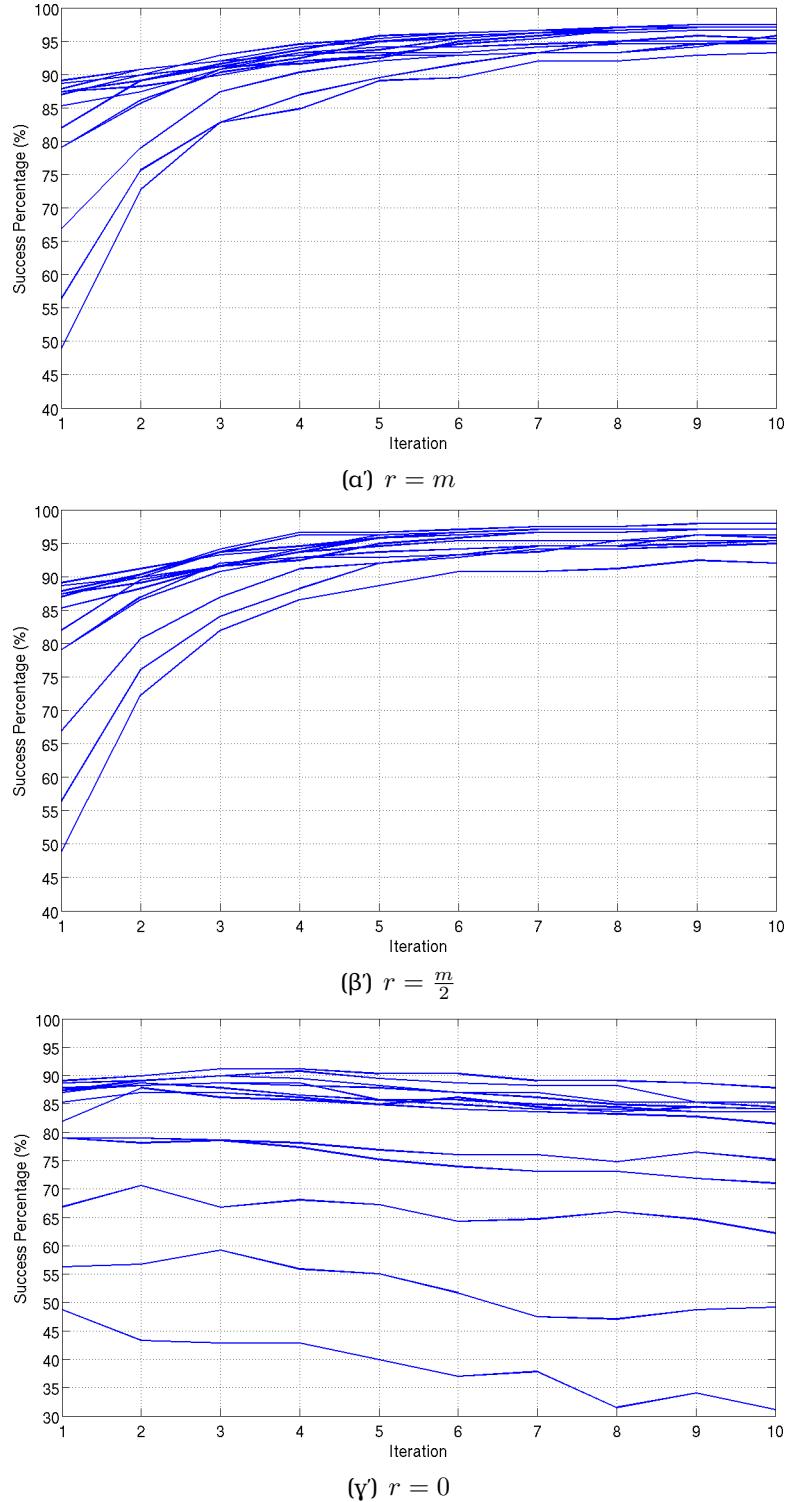


Σχήμα 8.18: Χρονικό κόστος ανά επανάληψη για $r = m$, $r = \frac{m}{2}$ και $r = 0$

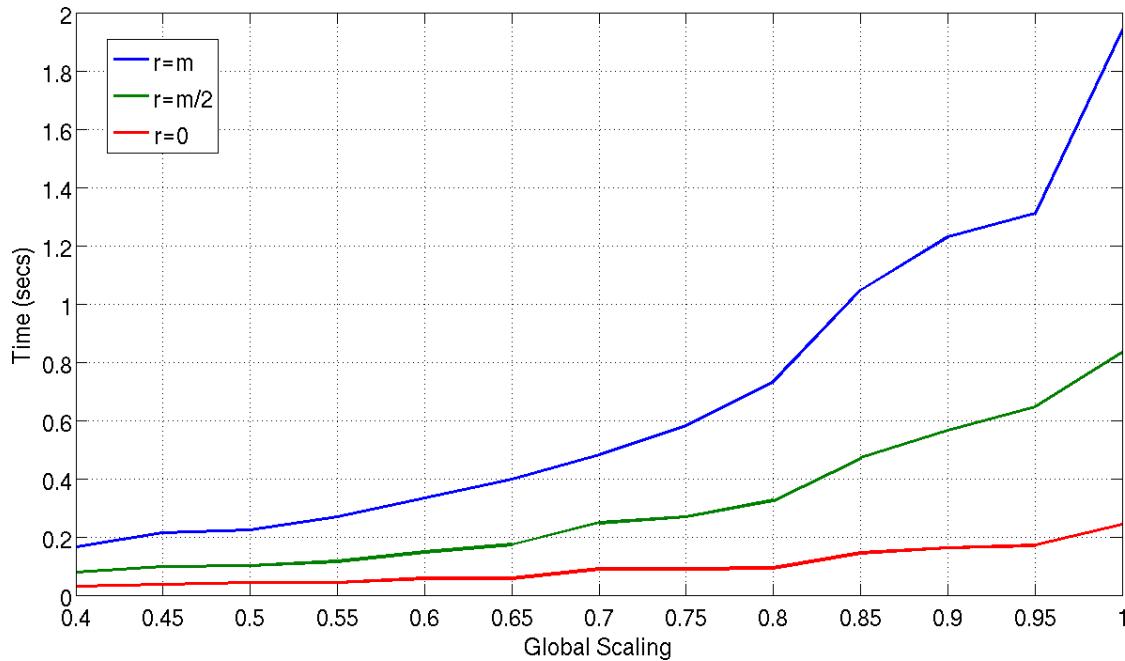
Με το διάγραμμα αυτό γίνεται ξεκάθαρη η χρησιμότητα των παραμέτρων r και s_{global} καθώς και η ευελιξία επιλογών που δίνουν στο χρήστη του τελικού συστήματος. Ωστόσο, για να υπολογιστεί το χρονικό κόστος του αλγορίθμου ανά εικόνα εισόδου και κατ' επέκταση η τιμή $frames/sec$, θα πρέπει να υπολογιστούν οι τιμές των παραμέτρων της συνθήκης τερματισμού, M_{iter} και M_{displ} , ώστε να γνωρίζουμε πόσες επαναλήψεις χρειάζεται να εκτελεστούν για να γίνει σύγκλιση.

Στο Σχήμα 8.19 φαίνονται όλες οι καμπύλες του ποσοστού επιτυχίας για όλες τις τιμές της s_{global} ως προς τον αριθμό επαναλήψεων του αλγορίθμου. Τα διαγράμματα αντιστοιχούν σε τιμές $r = m$, $r = \frac{m}{2}$ και $r = 0$. Παρατηρείται ότι στα διαγράμματα των τιμών $r = m$ και $r = \frac{m}{2}$, για τις περισσότερες τιμές της παραμέτρου s_{global} , το ποσοστό επιτυχίας του αλγορίθμου ξεπερνάει το 90% από την τρίτη επανάληψη και προσεγγίζει την τελική του τιμή από την τέταρτη ή πέμπτη επανάληψη. Αντίθετα, στο διάγραμμα για $r = 0$ φαίνεται ότι η απόδοση του αλγορίθμου παραμένει ίδια όσο και να αυξάνεται ο αριθμός των επαναλήψεων. Η απόδοση του αλγορίθμου μηδενικής τάξης μπορεί να βελτιωθεί με τη μέθοδο της ανανέωσης του διανύσματος μέσης υφής προσώπου. Συνεπώς, πέντε συνολικά επαναλήψεις είναι ένας ασφαλής αριθμός ώστε να υπάρχει σύγκλιση του αλγορίθμου.

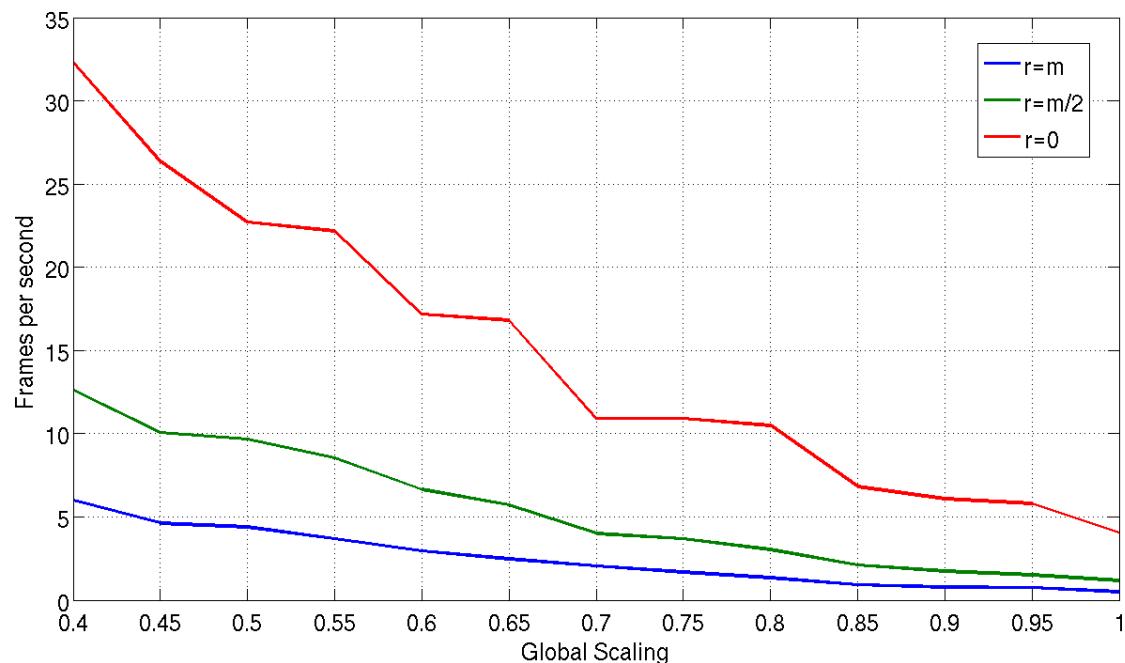
Στο Σχήμα 8.20 φαίνεται το διάγραμμα του συνολικού χρονικού κόστους του αλγορίθμου ως προς την καθολική κλίμακα, ενώ στο Σχήμα 8.21 φαίνεται η αντίστοιχη απόδοση σε frames ανά δευτερόλεπτο.



Σχήμα 8.19: Παραδείγματα αρχικοποίησης του αλγορίθμου ταιριάσματος για $s_{initial} = 1, 0.9, 0.8, 0.7$



Σχήμα 8.20: Συνολικό χρονικό κόστος αλγορίθμου ταιριάσματος για $r = m$, $r = \frac{m}{2}$ και $r = 0$



Σχήμα 8.21: Απόδοση frames/second αλγορίθμου ταιριάσματος για $r = m$, $r = \frac{m}{2}$ και $r = 0$

Το μόνο που απομένει είναι ο προσδιορισμός της τιμής της παραμέτρου ελάχιστης μετατόπισης σημείων σχήματος M_{displ} για τις πέντε επαναλήψεις του αλγορίθμου που υπολογίστηκε ότι είναι αρκετές. Ο Πίνακας 8.6 δείχνει τη μέση τιμή της M_{displ} για όλες τις τιμές $s_{global} = 1, \dots, 0.4$ με βήμα 0.05 και για τις τιμές τάξης $r = \{m, \frac{m}{2}, 0\}$.

s_{global}	Τάξη αλγορίθμου		
	$r = m$	$r = \frac{m}{2}$	$r = 0$
1	1.8416	2.2712	7.9511
0.95	2.0431	2.5059	8.8748
0.9	2.2798	2.7756	10.337
0.85	2.5447	2.9881	12.133
0.8	2.8521	3.4375	15.043
0.75	3.3691	3.9092	17.211
0.7	3.7652	4.3758	21.589
0.65	4.4829	5.4159	24.151
0.6	5.1646	6.3290	28.967
0.55	6.3679	7.9595	33.969
0.5	7.1854	8.8680	39.063
0.45	9.3659	11.625	44.184
0.4	11.389	13.893	42.911

Πίνακας 8.6: Τιμές παραμέτρου M_{displ} ως προς την καθολική κλίμακα και την τάξη του αλγορίθμου

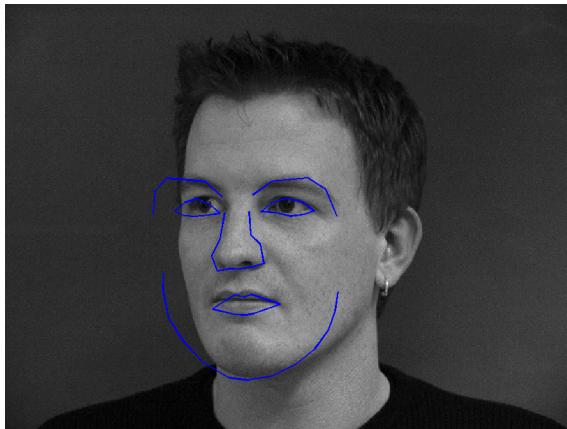
Τέλος στον Πίνακα 8.7 φαίνονται οι τελικές προτεινόμενες τιμές των έξι παραμέτρων με ελευθερία επιλογής στον χρήστη.

Περιγραφή	Σύμβολο	Τιμή
Καθολική κλίμακα	s_{global}	0.7
Τάξη αλγορίθμου	r	$m/2$
Κατώφλι ανιχνευτή προσώπου	θ_{V-J}	-1
Κλίμακα αρχικοποίησης	$s_{initial}$	0.9 ή 1
Μέγιστος αριθμός επαναλήψεων	M_{iter}	10
Ελάχιστη μετατόπιση σημείων σχήματος	M_{displ}	4.5

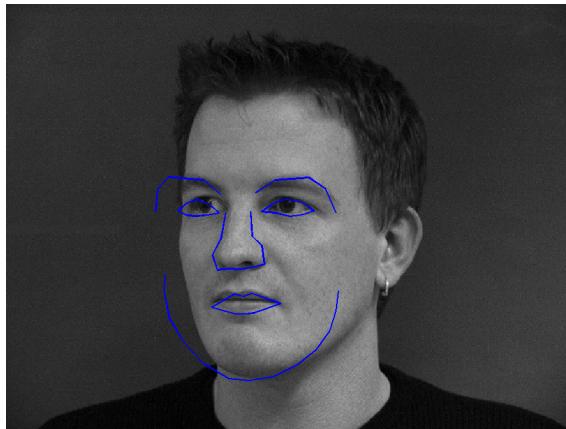
Πίνακας 8.7: Προτεινόμενες τιμές παραμέτρων συστήματος αναγνώρισης προσώπου

8.5.4 Παράδειγμα Εκτέλεσης Αλγορίθμου Ταιριάσματος ΑΑΜ

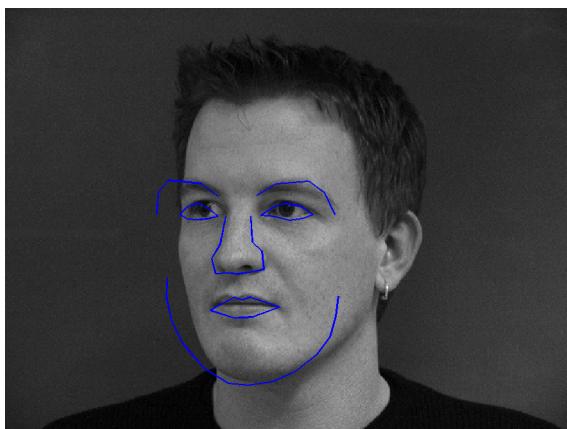
Στο Σχήμα 8.22 φαίνονται τα διαδοχικά βήματα της διαδικασίας ταιριάσματος σε μια εικόνα της βάσης δεδομένων IMM. Απεικονίζονται 10 επαναλήψεις με $r = m$ και $s_{global} = 1$.



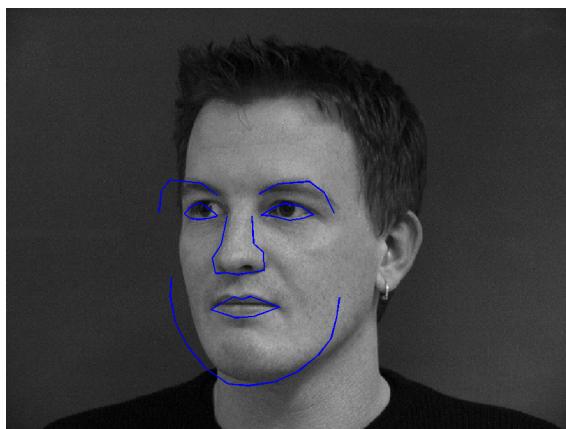
(α') Επανάληψη 1/10, Σφάλμα σχήματος 6.88



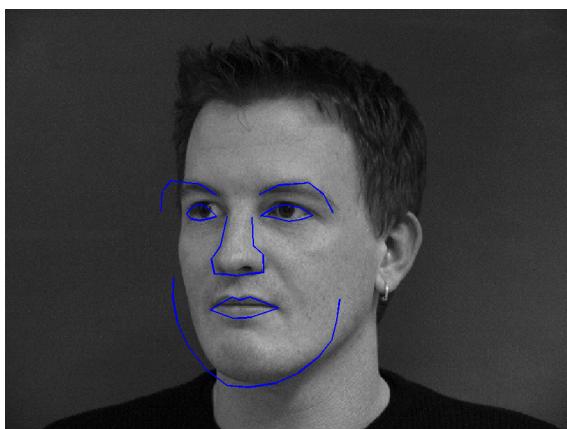
(β') Επανάληψη 2/10, Σφάλμα σχήματος 6.37



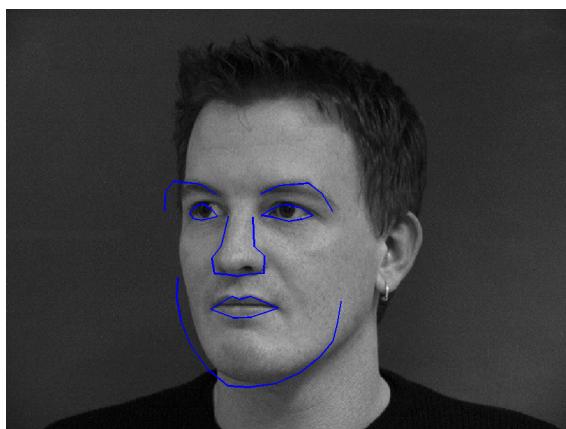
(γ') Επανάληψη 3/10, Σφάλμα σχήματος 5.83



(δ') Επανάληψη 4/10, Σφάλμα σχήματος 5.32



(ε') Επανάληψη 5/10, Σφάλμα σχήματος 4.83



(φ') Επανάληψη 6/10, Σφάλμα σχήματος 4.39



(ζ) Επανάληψη 7/10, Σφάλμα σχήματος 3.91

(η) Επανάληψη 8/10, Σφάλμα σχήματος 3.50



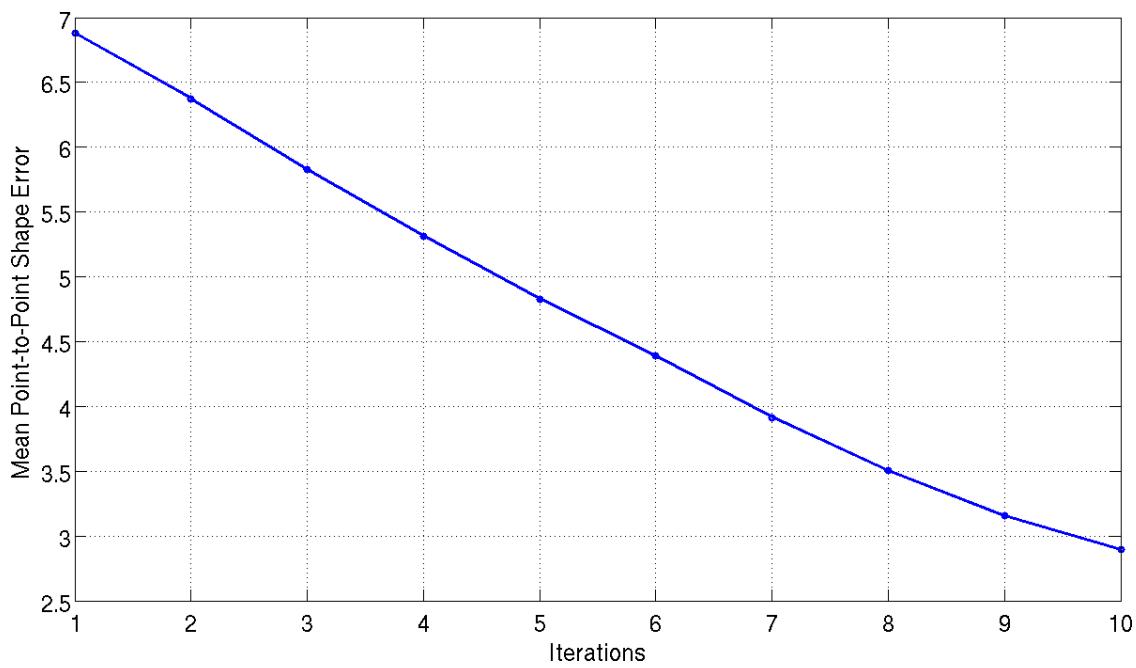
(θ) Επανάληψη 9/10, Σφάλμα σχήματος 3.16

(ι) Επανάληψη 10/10, Σφάλμα σχήματος 2.89

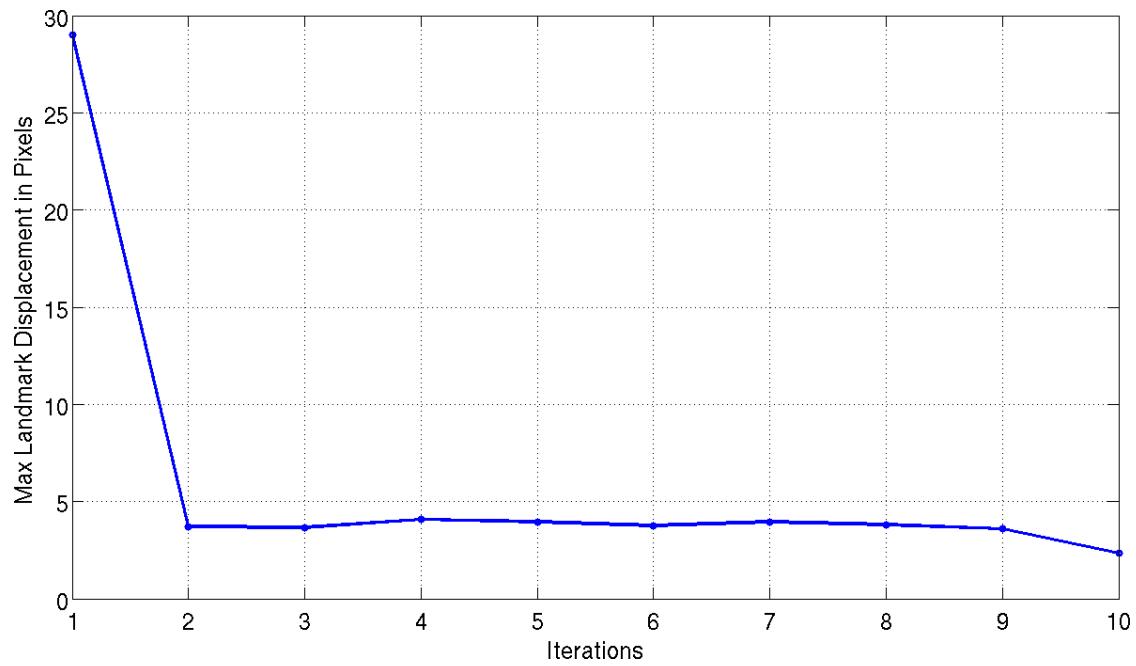
Σχήμα 8.22: Παράδειγμα εκτέλεσης αλγορίθμου ταιριάσματος AAM

Το Σχήμα 8.23 δείχνει την καμπύλη του σφάλματος σχήματος και το Σχήμα 8.24 δείχνει την καμπύλη της μέγιστης μετατόπισης των landmark σημείων του σχήματος σε κάθε επαναληπτικό βήμα.

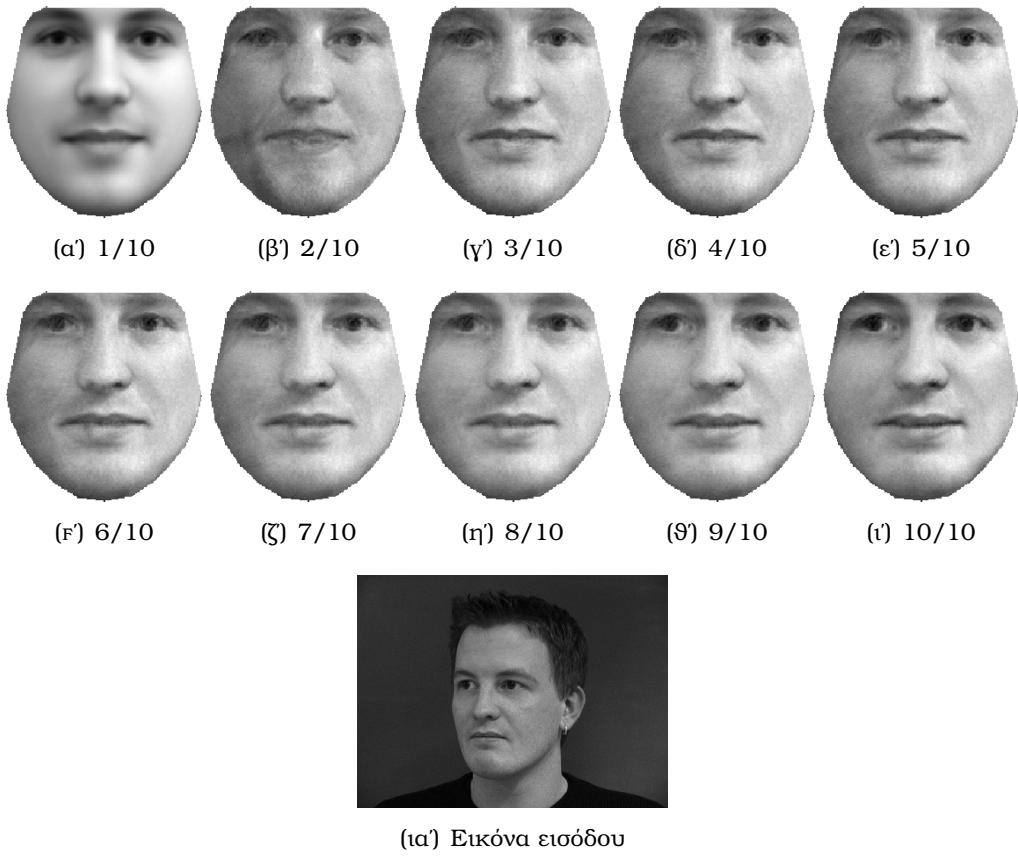
Οι διαδοχικές εικόνες του Σχήματος 8.25 απεικονίζουν τη διαδικασία ταιριάσματος του μοντέλου υφής προσώπου καθώς και την αρχική υφή για σύγκριση. Τέλος, το Σχήμα 8.26 δείχνει την καμπύλη του σφάλματος υφής όπως υπολογίζεται με τη νόρμα της εικόνας λάθους $E(\mathbf{q})$.



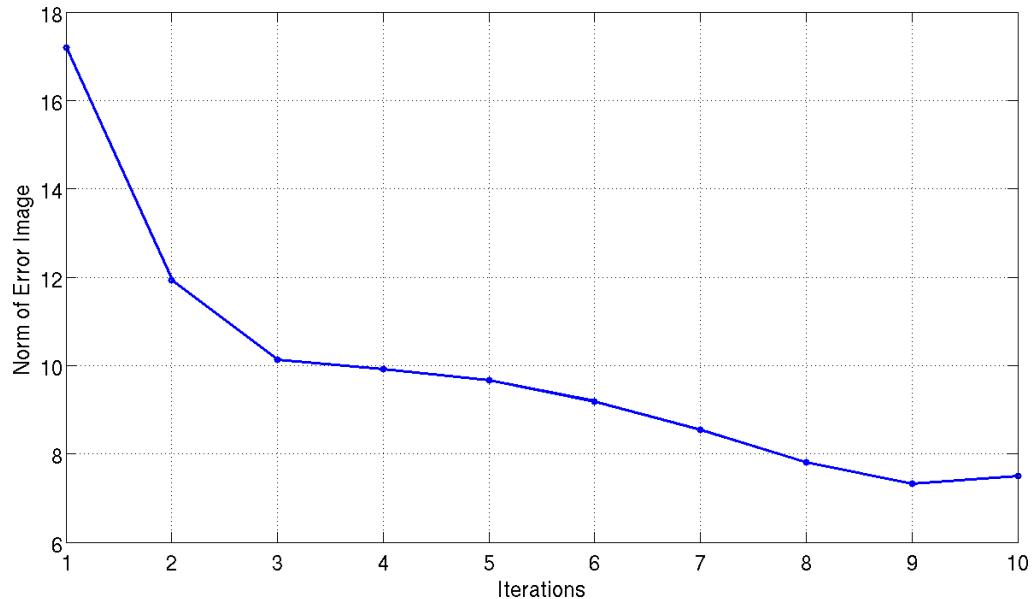
Σχήμα 8.23: Σφάλμα σχήματος προσώπου για το παράδειγμα του σχήματος 8.22



Σχήμα 8.24: Μέγιστη μετατόπιση landmark σημείων για το παράδειγμα του σχήματος 8.22



Σχήμα 8.25: Παράδειγμα ταιριάσματος υφής προσώπου



Σχήμα 8.26: Σφάλμα υφής προσώπου για το παράδειγμα του σχήματος 8.25

Μέρος III

**Εφαρμογή Μοντελοποίησης
Προσώπου σε Αναγνώριση
Εκφράσεων και Συναισθημάτων**

Κεφάλαιο 9

Εισαγωγή

Όπως είδαμε στα προηγούμενα κεφάλαια χρησιμοποιήθηκαν μέθοδοι και θεωρίες από τους τομείς της Όρασης Υπολογιστών και της Αναγνώρισης Προτύπων. Ωστόσο το φάσμα των προβλημάτων στα οποία η Οπτική Μοντελοποίηση του ανθρώπινου προσώπου θεωρείται απαραίτητη εξαπλώνεται σε πολλούς ακόμα τομείς, όπως η Ιατρική, η Ρομποτική και η Αλληλεπίδραση Ανθρώπου-Ρομπότ.

Κάποια πολύ σημαντικά επιστημονικά προβλήματα στον τομέα της αναγνώρισης, που στηρίζονται στην οπτική μοντελοποίηση προσώπου και έχουν επικεντρώσει την προσοχή της επιστημονικής κοινότητας, είναι αυτά της αναγνώρισης προσώπου (με την έννοια της ταυτοποίησης), της αυτόματης αναγνώρισης εκφράσεων προσώπου και της αυτόματης αναγνώρισης της συναισθηματικής κατάστασης ενός ανθρώπου μέσω των εκφράσεων του προσώπου. Τα τελευταία χρόνια μάλιστα έχουν αναπτυχθεί πολλές εμπορικές εφαρμογές και προϊόντα τα οποία απαιτούν την μοντελοποίηση του προσώπου, όπως φωτογραφικές μηχανές με αναγνώριση προσώπου, ρομπότ για οικιακή χρήση, συστήματα ασφαλείας σε αυτοκίνητα αλλά και συστήματα αυτοματοποιημένης ιατρικής διάγνωσης.

9.1 Προεπισκόπηση

Στο μέρος αυτό της διπλωματικής εργασίας αναλύονται δυο πολύ σημαντικές εφαρμογές της οπτικής μοντελοποίησης ανθρώπινου προσώπου: η αυτόματη αναγνώριση εκφράσεων του προσώπου και η αυτόματη αναγνώριση συναισθηματικής κατάστασης ενός ανθρώπου μέσω των εκφράσεων του προσώπου. Η κωδικοποίηση των εκφράσεων γίνεται με το σύστημα FACS (Facial Action Coding System) που αναπτύχθηκε από τους Ekman και Friesen το 1978 [EkFr78]. Το FACS είναι ένα σύστημα ταξινόμησης των εκφράσεων προσώπου και κατηγοριοποίησης των φυσικών εκφράσεων των συναισθημάτων με βάση τις κινήσεις όλων των ορατών μυών του ανθρώπινου προσώπου. Η βάση δεδομένων που χρησιμοποιείται είναι η CK+ (Extended Cohn-Kanade Dataset) [Luc+10] που ειδικεύεται σε αναγνώριση συναισθημάτων. Το πρόβλημα της αυτόματης ταξινόμησης κεφράσεων και συναισθημάτων αντιμετωπίζεται με τη μέθοδο Support Vector Machine.

Κεφάλαιο 10

Βάση δεδομένων CK+

Στο κεφάλαιο αυτό παρουσιάζεται η βάση δεδομένων CK+ (Extended Cohn-Kanade Dataset) και δίνονται παραδείγματα χρήσης της στο σύστημα οπτικής μοντελοποίησης ανθρώπινου προσώπου.

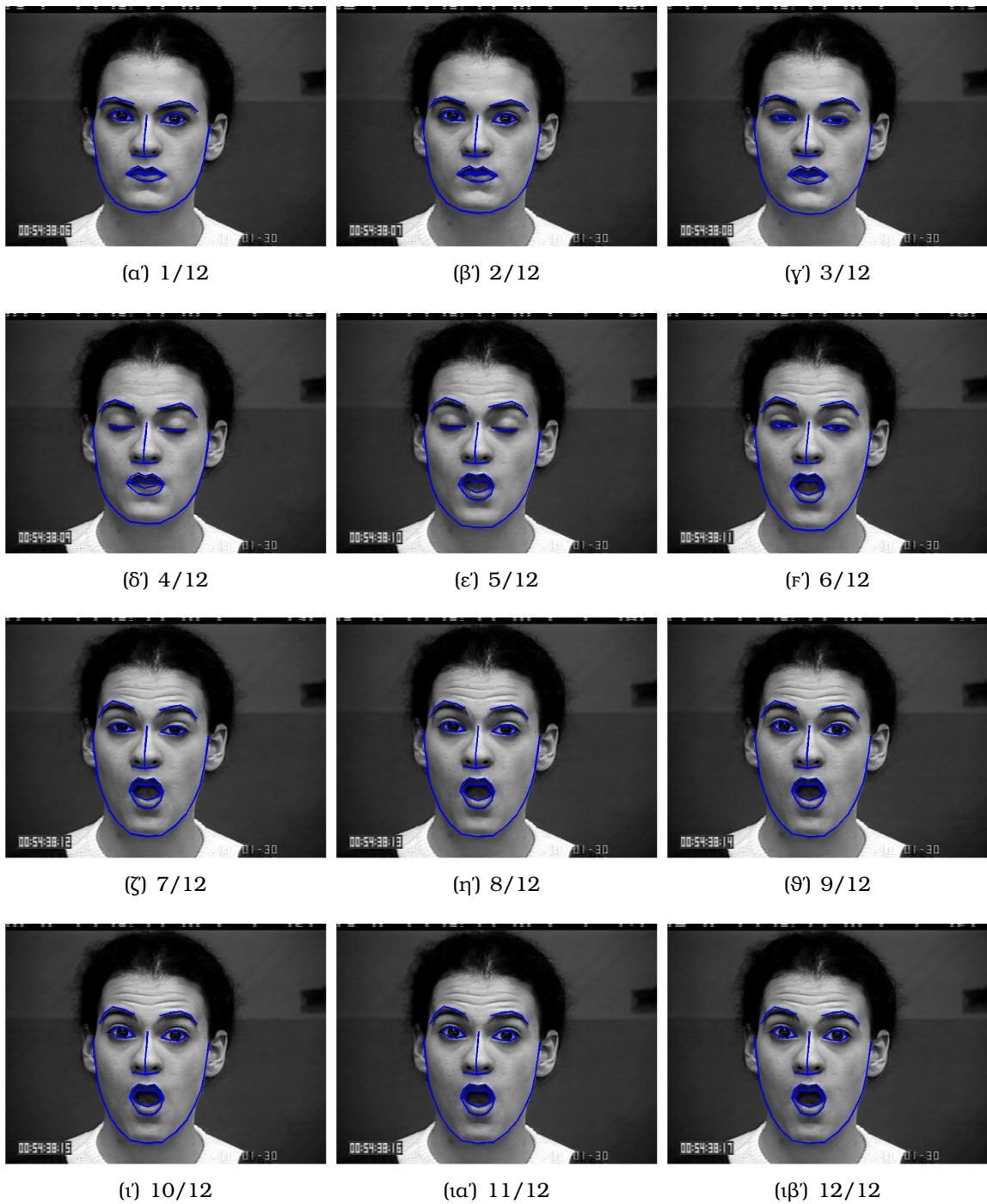
10.1 Περιγραφή Χαρακτηριστικών, Πλεονεκτήματα και Μειονεκτήματα

Η βάση δεδομένων CK+ [Luc+10] είναι ειδικά σχεδιασμένη για αναγνώριση της συναισθηματικής κατάστασης ενός ανθρώπου μέσω αναγνώρισης των εκφράσεων του προσώπου. Αποτελείται από 593 ακολουθίες εικόνων με τη συμμετοχή 123 ατόμων. Οι ηλικίες των ανθρώπων που συμμετέχουν είναι μεταξύ 18 και 50 ετών. Το 69% των συμμετεχόντων είναι γένους θηλυκού και το υπόλοιπο 31% γένους αρσενικού. Ακόμη, το 81% είναι είτε Ευρωπαίοι είτε Αμερικάνοι, το 13% είναι αφρο-Αμερικάνοι και το υπόλοιπο 6% ανήκουν σε άλλη φυλετική κατηγορία.

Οι ακολουθίες των εικόνων έχουν διαφορετική διάρκεια στο διάστημα από 10 έως 60 frames. Υπάρχουν μετωπικές εικόνες προσώπων και εικόνες με πρόσωπα στραμμένα κατά 30 μοίρες. Όλες οι εικόνες είναι ψηφιακές με ανάλυση είτε 640×490 είτε 640×480 και τιμές χρωμάτων γκρι των 8-bit ή RGB των 24-bit.

Επιπροσθέτως, όλες οι ακολουθίες εικόνων ξεκινούν από την ουδέτερη έκφραση (neutral expression) και καταλήγουν στην κορύφωση της έκφρασης ή του συνδυασμού εκφράσεων (peak expression) που έχει ζητηθεί στο άτομο. Ένα τέτοιο χαρακτηριστικό παράδειγμα φαίνεται στο Σχήμα 10.1, όπου δημιουργείται το συναίσθημα της έκπληξης (surprise).

Το μεγάλο πλεονέκτημα της βάσης δεδομένων CK+ είναι ο μεγάλος αριθμός σημείων landmark με τα οποία ορίζεται το σχήμα του προσώπου σε κάθε εικόνα, τα οποία είναι 68. Μάλιστα, δίνεται πολύ μεγάλη έμφαση στις περιοχές του προσώπου που συμβάλλουν περισσότερο στις εκφράσεις. Συνεπώς τα χείλη τονίζονται με εσωτερικά και εξωτερικά σημεία, τα φρύδια αντιμετωπίζονται με ένα περίγραμμα σημείων και δεν θεωρούνται μια γραμμή και η μύτη αντιμετωπίζεται πολύ απλοϊκά δεδομένου ότι δεν μπορεί να κινηθεί σε μεγάλο βαθμό ώστε να συμβάλλει στις εκφράσεις. Ένα ακόμα



Σχήμα 10.1: Παράδειγμα ακολουθίας εικόνων από ουδέτερη έκφραση σε έκπληξη της βάσης δεδομένων CK+

πλεονέκτημα είναι ο μεγάλος αριθμός ανθρώπων και εικόνων που υπάρχουν στη βάση δεδομένων, ο οποίος της δίνει μεγάλη ποικιλία. Μάλιστα όλες οι εικόνες συνοδεύονται από κατάλληλα αρχεία που έχουν ορισμένα τα landmark σημεία, ενώ οι εικόνες με τις

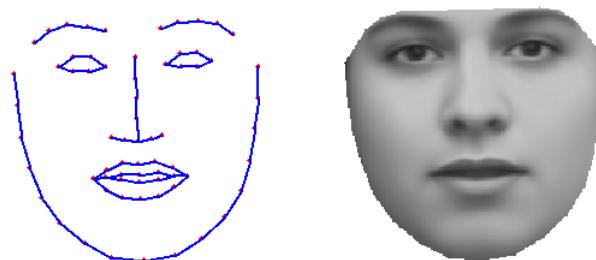
κορυφώσεις των συναισθημάτων/εκφράσεων συνοδεύονται επό επιπλέον αρχεία που αναφέρουν το συναίσθημα αλλά και το/τα action units του συστήματος FACS.

Επίσης, ένα πολύ σημαντικό πλεονέκτημα της βάσης αυτής έναντι άλλων ανταγωνιστικών είναι το γεγονός ότι τα συναισθήματα που έχουν δοθεί ως ετικέτες στις εικόνες δεν είναι αυτά τα οποία ζητήθηκαν από τα άτομα που απεικονίζονται, αλλά έχουν καθοριστεί και επιβεβαιωθεί από κατάλληλους επιστήμονες με βάση τα AU (Action Units) που υπάρχουν στην εικόνα και την κωδικοποίηση FACS. Αυτό έχει ως συνέπεια μόλις 327 από τις 593 ακολουθίες εικόνων να έχουν ετικέτες συναισθημάτων και κατά συνέπεια να συμπεριληφθηνούν στην αναγνώριση συναισθημάτων.

Από την άλλη πλευρά, το μεγάλο και εμφανές μειονέκτημα της βάσης δεδομένων CK+ είναι η άσχημη ποιότητα εικόνων που περιλαμβάνει. Στη συντριπτική πλειονότητα των εικόνων, ο φωτισμός και η απόχρωση του γκρι είναι τόσο χαμηλής ποιότητας που η υφή κρίνεται ακατάλληλη για μοντελοποίηση του δέρματος του προσώπου. Το γεγονός αυτό καθιστά πολύ δύσκολη τη χρήση της βάσης δεδομένων στο μοντέλο AAM για ταυτόχρονη μοντελοποίηση του σχήματος αλλά και της υφής προσώπου.

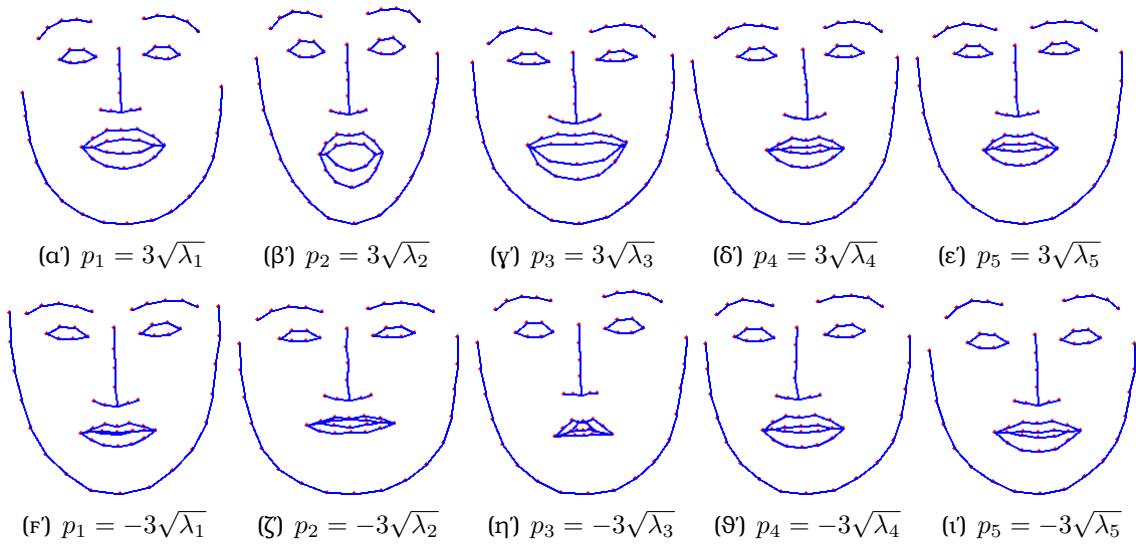
10.2 Οπτική Μοντελοποίηση Προσώπου με τη Βάση Δεδομένων CK+

Φυσικά για την αυτόματη αναγνώριση εκφράσεων προσώπου απαιτείται πρώτα η εκπαίδευση του μοντέλου AAM με τη βάση δεδομένων CK+. Η εκπαίδευση αναμένεται να διαρκεί πολύ περισσότερη ώρα συγκριτικά με τη βάση δεδομένων IMM που χρησιμοποιήθηκε στο προηγούμενο μέρος λόγω του μεγάλου αριθμού εικόνων (10734 έναντι 240). Για να αποφευχθεί αυτή η χρονοβόρα διαδικασία η εκπαίδευση έγινε συμπεριλαμβάνοντας την πρώτη και την τελευταία εικόνα καθεμιάς από τις 593 ακολουθίες, δηλαδή ένα σύνολο 1186 εικόνων. Το μέσο σχήμα και η μέση υφή προσώπου που προέκυψαν φαίνονται στο Σχήμα 10.2.

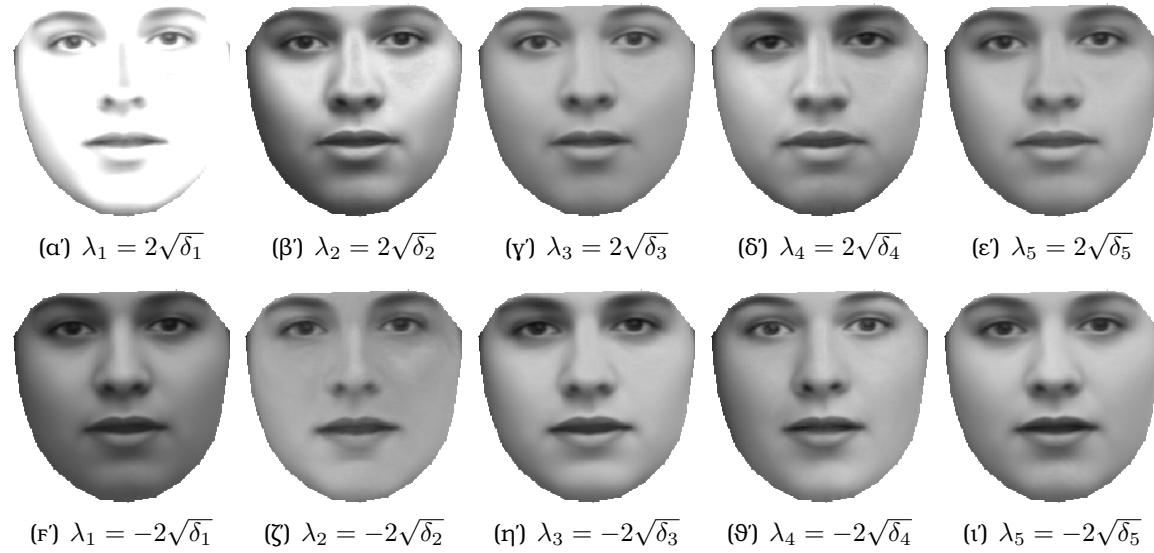


Σχήμα 10.2: Μέσο σχήμα και μέση υφή προσώπου της βάσης δεδομένων CK+

Στα Σχήματα 10.3 και 10.4 φαίνονται στιγμιότυπα του μοντέλου σχήματος και υφής προσώπου αντίστοιχα για τα πέντε πιο σημαντικά ιδιοδιανύσματα. Οι παράμετροι του μοντέλου μεταβάλλονται σε διάστημα που ορίζεται από τις αντίστοιχες ιδιοτιμές. Ο αριθμός των αξόνων μεταβλητότητας του μοντέλου σχήματος προέκυψε 21 και ο αντίστοιχος του μοντέλου υφής προέκυψε 69.



Σχήμα 10.3: Στιγμιότυπα του μοντέλου σχήματος προσώπου CK+ για τα πέντε πιο σημαντικά ιδιοδιανύσματα



Σχήμα 10.4: Στιγμιότυπα του μοντέλου υφής προσώπου CK+ για τα πέντε πιο σημαντικά ιδιοδιανύσματα

Από το Σχήμα 10.4 γίνεται ξεκάθαρο το μειονέκτημα της βάσης δεδομένων CK+. Το μοντέλο υφής προσώπου είναι πολύ ασθενές με ελάχιστη μεταβλητότητα. Χαρακτηριστικό είναι το γεγονός ότι κανένα αρσενικό άτομο της βάσης δεδομένων δεν έχει μουστάκι ή γένια. Επομένως, το μοντέλο υφής κρίνεται ακατάλληλο για το ταίριασμα του AAM πλήρους τάξης.

Κεφάλαιο 11

Σύστημα Κωδικοποίησης FACS

Στο κεφάλαιο αυτό παρουσιάζεται συνοπτικά το σύστημα κωδικοποίησης FACS (Facial Action Coding System). Αναλύεται η χρησιμότητά του, η κατηγοριοποίηση των AU (Action Units) του προσώπου και ο ορισμός των συναισθημάτων με βάση αυτά.

11.1 Χρησιμότητα Συστήματος FACS

Το σύστημα κωδικοποίησης δράσεων του προσώπου FACS (Facial Action Coding System) [EkFr78] δημιουργήθηκε για την εύκολη και κατανοητή διαφοροποίηση όλων των πιθανών οπτικά διακριτών κινήσεων του προσώπου. Όπως χαρακτηριστικά αναφέρουν οι επιστήμονες Ekman και Friesen, η ανάπτυξη του συστήματος FACS στηρίχθηκε στην "ανάγκη για διαχωρισμό του συνπεράσματος από την περιγραφή". Με άλλα λόγια, η αναζήτηση του συναισθήματος που προκάλεσε μια συγκεκριμένη έκφραση προσώπου πρέπει να καθοριστεί από την περιγραφή της ίδιας της έκφρασης προσώπου.

Η δημιουργία του συστήματος κωδικοποίησης FACS βασίστηκε σε μια λεπτομερή μελέτη διάρκειας οκτώ χρόνων πάνω στους μύες που ελέγχουν το πρόσωπο. Ουσιαστικά, μετράει κάθε ορατή κίνηση του ανθρώπινου προσώπου που οφείλεται σε κίνηση αντίστοιχων μυών του προσώπου. Το FACS βασίστηκε μόνον στους ορατούς μύες διότι η μέτρηση μη ορατών μυών είναι αδύνατη χωρίς ηλεκτρομυογράφημα και κατά δεύτερο ο στόχος ήταν η διερεύνηση της κίνησης των μυών του προσώπου σε άτομα που δεν γνωρίζουν ότι υπόκεινται σε αυτόν. Τονίζεται ότι η κωδικοποίηση και ανάλυση των κινήσεων του προσώπου του συστήματος FACS μοιάζει σε πολύ μεγάλο βαθμό με τον τρόπο που ο ανθρώπινος εγκέφαλος αντιλαμβάνεται τις εκφράσεις του προσώπου.

11.2 Δομή του Συστήματος FACS

Το σύστημα κωδικοποίησης FACS σχεδιάστηκε με τέτοιο τρόπο ώστε να καθορίζει τις ελάχιστες μονάδες της συμπεριφοράς (κίνησης) του προσώπου. Αυτές οι μονάδες ονόμαζονται μονάδες δράσης. Ο όρος που θα χρησιμοποιούμε στη συνέχεια είναι ο

αγγλικός όρος που πηγάζει από το όνομα του FACS, δηλαδή *action units* ή σε συντομογραφία *AUs*. Κάποιες *AUs* έχουν ένα-προς-ένα αντιστοίχιση με συγκεκριμένους μύες του προσώπου, ενώ άλλες *AUs* προκαλούνται από ένα συνδυασμό κινήσεων δύο ή περισσότερων μυών, ακόμα και αν οι επιδράσεις των μυών αυτών είναι μη διακριτές στο πρόσωπο. Σε άλλες περιπτώσεις, ορισμένες *AUs* μπορεί να συνδέονται με τον ίδιο μυ, σε περίπτωση που διαφορετικά τμήματα του συγκεκριμένου μυ ενεργοποιούνται ανεξάρτητα.

Στο σύστημα FACS έχει γίνει αρίθμηση κάθε μονάδας *AU* για λόγους ευκολίας. Στην αρχική έκδοση του FACS το 1978 υπήρχαν 44 *AUs*, με τους αριθμούς να είναι από 1 έως 46 γιατί δεν χρησιμοποιήθηκαν το 3 και το 40. Η ανανεωμένη έκδοση του συστήματος το 2002 [EFH02], η οποία περιέχει κινήσεις της κόρης των ματιών και του κεφαλιού, περιλαμβάνει επιπλέον 12 *AUs* με αρίθμηση από το 51 και πάνω. Και στις δυο εκδόσεις οι *AUs* 1 έως 7 αναφέρονται στις κινήσεις του πάνω μέρους του προσώπου, ενώ οι *AUs* 8 έως 46 στις κινήσεις του κάτω μέρους του προσώπου.

Στον Πίνακα 11.1 φαίνονται οι επεξηγήσεις των δράσεων *AUs* που εμφανίζονται στις εικόνες της βάσης δεδομένων CK+. Υπενθυμίζεται ότι οι δράσεις αυτές σημειώνονται μόνο στην εικόνα κορύφωσης της εκάστοτε ακολουθίας. Στην τρίτη στήλη του Πίνακα 11.1 σημειώνεται και οι συχνότητα εμφάνισης *N* της κάθε *AU* στις εικόνες της CK+.

Εκτός από την ύπαρξη μιας ή περισσότερων *AU* σε μια εικόνα προσώπου είναι χρήσιμος και ο προσδιορισμός της έντασης (*intensity*) της κάθε *AU*. Η ένταση μετριέται σε μια κλίμακα από A, που αντιπροσωπεύει την ελάχιστη ένταση, έως E που αντιπροσωπεύει τη μέγιστη ένταση. Τα κριτήρια του προσδιορισμού της έντασης για κάθε *AU* δίνονται στον αντίστοιχο οδηγό του συτήματος FACS.

<i>AU</i>	Περιγραφή	<i>N</i>	<i>AU</i>	Περιγραφή	<i>N</i>
1	Ανύψωση εσωτερικού του φρυδιού	173	18	Συρρίκνωση χειλιών	9
2	Ανύψωση εξωτερικού του φρυδιού	116	20	Τέντωμα χειλιών	77
4	Χαμήλωμα φρυδιού	191	21	Σφίξιμο λαιμού	3
5	Ανύψωση άνω μέρους βλέφαρου	102	23	Σφίξιμο χειλιών	59
6	Ανύψωση μάγουλου	122	24	Πίεση χειλιών	57
7	Σφίξιμο βλέφαρου	119	25	Χείλη χωριστά	287
9	Ζάρωμα μύτης	74	26	Κατέβασμα σαγονιού	48
10	Ανύψωση άνω χειλιού	21	27	Τέντωμα στόματος	81
11	Ρινοχειλικό βαθούλωμα	33	28	Ρούφηγμα χειλιών	1
12	Τράβηγμα γωνίας χειλιών	111	29	Σπρώξιμο σαγονιού	1
13	Τράβηγμα μάγουλου	2	31	Σφίξιμο σαγονιού	3
14	Λακκάκι στο μάγουλο	29	34	Φούσκωμα μάγουλου	1
15	Συμπίεση γωνίας χειλιών	89	38	Διαστολή ρουθουνιών	29
16	Συμπίεση κάτω χειλιού	24	39	Συμπίεση ρουθουνιών	16
17	Ανύψωση πηγουνιού	196	43	Κλειστά μάτια	9

Πίνακας 11.1: Περιγραφή και συχνότητα εμφάνισης των δράσεων *AUs* στη βάση δεδομένων CK+

11.3 Κατηγοριοποίηση Συναισθημάτων

Από τις AUs που αναφέρονται στον πίνακα 11.1 ότι περιλαμβάνονται στη βάση δεδομένων CK+, οι απαραίτητες για την περιγραφή των συναισθημάτων είναι μόνο οι AUs 1, 2, 4, 5, 6, 7, 9, 10, 12, 15, 16, 17, 20, 23, 24, 25 και 26. Πιο συγκεκριμένα οι 17 αυτές δράσεις AUs είναι αρκετές για την περιγραφή των 7 βασικών κατηγοριών συναισθημάτων: Θυμός, Περιφρόνηση, Αηδία, Φόβος, Ευτυχία, Θλίψη και Έκπληξη.

Στον Πίνακα 11.2 φαίνονται τα αναλυτικά κριτήρια αναγνώρισης των συναισθημάτων βάση των AUs, όπως αναφέρονται στο [PaRo00]. Στον πίνακα αυτό το σύμβολο "+" δείχνει το λογικό KAI.

Συναισθημα	Κριτήρια AUs
Θυμός	$4 + 7 + (((23 \text{ ή } 24) \text{ με ή χωρίς } 17) \text{ ή } (16 + (25 \text{ ή } 26)) \text{ ή } (10 + 16 + (25 \text{ ή } 26))) \text{ με ή χωρίς } 2$
Αηδία	$((10 \text{ με ή χωρίς } 17) \text{ ή } (9 \text{ με ή χωρίς } 17)) + (25 \text{ ή } 26)$
Φόβος	$(1 + 4) + (5 + 7) + 20 + (25 \text{ ή } 26)$
Ευτυχία	$6 + 12 + 16 + (25 \text{ ή } 26)$
Θλίψη	$1 + 4 + (6 \text{ ή } 7) + 15 + 17 + (25 \text{ ή } 26)$
Έκπληξη	$(1 + 2) + (5 \text{ χωρίς } 7) + 26$

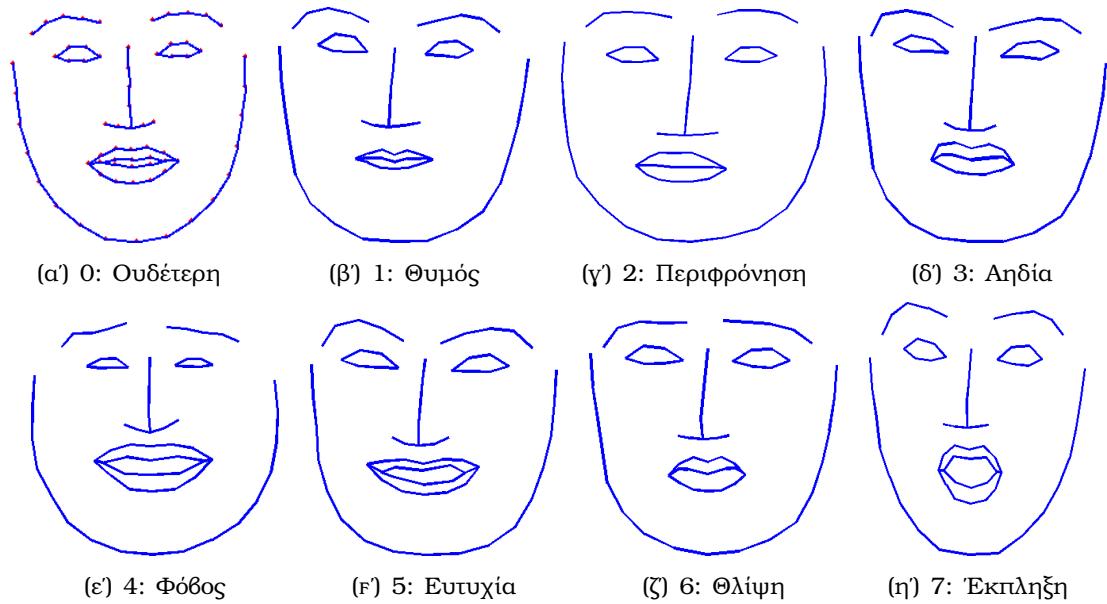
Πίνακας 11.2: Αναλυτική κωδικοποίηση συναισθημάτων με τις δράσεις AUs

Ωστόσο για λόγους απλότητας του πειράματος που ακολουθεί, θα χρησιμοποιηθεί μια απλούστερη δομή συνδυασμών για την κωδικοποίηση των 7 βασικών συναισθημάτων, όπως αυτή αναφέρεται στο [Luc+10]. Η απλούστερη κωδικοποίηση φαίνεται στον Πίνακα 11.3.

Συναισθημα	Κριτήρια AUs
Θυμός	$23 + 24$
Αηδία	$9 \text{ ή } 10$
Φόβος	αν 5 έχει ένταση Ε τότε $(1 + 2)$ αλλιώς $(1 + 2 + 4)$
Ευτυχία	12
Θλίψη	$(1 + 4 + 15) \text{ ή } 11 \text{ με εξαίρεση το } (6 + 15)$
Έκπληξη	$(1 + 2) \text{ ή } 5 \text{ και η ένταση του } 5 \text{ να μην είναι μεγαλύτερη του } B$
Περιφρόνηση	$14 \text{ είτε από τη μια είτε και από τις δύο}$

Πίνακας 11.3: Απλοποιημένη κωδικοποίηση συναισθημάτων με τις δράσεις AUs

Στο Σχήμα 11.1 φαίνονται τυχαία επιλεγμένα παραδείγματα των 7 βασικών συναισθημάτων από τη βάση δεδομένων CK+. Φυσικά, τα συναισθήματα αναγνωρίστηκαν με βάση των πίνακα 11.3.



Σχήμα 11.1: Παραδείγματα των 7 βασικών συναισθημάτων από τη βάση δεδομένων CK+

Κεφάλαιο 12

Αυτόματη Αναγνώριση Εκφράσεων Προσώπου και Συναισθημάτων

Στο κεφάλαιο αυτό αναλύεται μια μέθοδος αναγνώρισης εκφράσεων προσώπου και συναισθημάτων χρησιμοποιώντας ταξινομητή Support Vector Machine (SVM). Τα διανύσματα χαρακτηριστικών εξάγονται αυτόματα από το σύστημα οπτικής μοντελοποίησης προσώπου με Active Appearance Models που παρουσιάστηκε στο Μέρος II.

12.1 Εξαγωγή Χαρακτηριστικών

Από το κεφάλαιο της οπτικής μοντελοποίησης του ανθρώπινου προσώπου στο Μέρος II και συγκεκριμένα από τη μέθοδος του ταιριάσματος του μοντέλου σε νέα πρόσωπα, γνωρίζουμε ότι το αποτέλεσμα είναι η εύρεση του διανύσματος των παραμέτρων σχήματος προσώπου

$$\tilde{\mathbf{p}} = [\mathbf{t}_{1:4}, \mathbf{p}_{1:n}]^T$$

μήκους $4 + n$, όπου \mathbf{t} οι παράμετροι του μετασχηματισμού ομοιότητας και \mathbf{p} οι παράμετροι του μοντέλου σχήματος. Σε περίπτωση που χρησιμοποιούμε τον αλγόριθμο τάξης $r > 0$ βρίσκουμε και το συνολικό διάνυσμα παραμέτρων υφής προσώπου

$$\tilde{\boldsymbol{\lambda}} = [\mathbf{u}_{1:2}, \boldsymbol{\lambda}_{1:m}]^T$$

μήκους $2 + m$, όπου \mathbf{u} οι παράμετροι του ομοπαραλληλικού μετασχηματισμού υφής και $\boldsymbol{\lambda}$ οι παράμετροι του μοντέλου υφής προσώπου.

Από το διάνυσμα παραμέτρων σχήματος αγνοούνται οι παράμετροι του μετασχηματισμού ομοιότητας και με βάση τις παραμέτρους \mathbf{p} δημιουργείται το σπιγμιότυπο του μοντέλου AAM, από τη γνωστή εξίσωση

$$\mathbf{s}_n = \mathbf{s}_0 + \sum_{i=1}^n p_i \mathbf{s}_i$$

το οποίο είναι ένα διάνυσμα μήκους $2L$, όπου L ο αριθμός των σημείων landmark που περιγράφουν το σχήμα προσώπου.

Αυτό το διάνυσμα s_n είναι και το χαρακτηριστικό (feature) με το οποίο γίνεται η εκπαίδευση του SVM και φυσικά η αυτόματη αναγνώριση των εκφράσεων και των συναισθημάτων. Είναι προφανές ότι το χαρακτηριστικό αυτό εξάγεται αυτόματα από το σύστημα οπτικής μοντελοποίησης προσώπου που έχει αναπτυχθεί. Για την καλύτερη μορφή των χαρακτηριστικών και κατ' επέκταση την καλύτερη λειτουργία του SVM σε κάθε διάνυσμα χαρακτηριστικών που δημιουργείται, εφαρμόζεται και μια κανονικοποίηση ως προς AU0. Αυτό επιτυγχάνεται αφαιρώντας από το διάνυσμα \mathbf{m}_{fs_n} το διάνυσμα του μέσου σχήματος προσώπου s_0 το οποίο τυπικά αντιπροσωπεύει και τη δράση AU0, η οποία δεν υπάρχει στον πίνακα των δράσεων FACS. Συνεπώς, το βασικό διάνυσμα χαρακτηριστικό που εξάγεται αυτόματα έχει τη μορφή

$$\mathbf{s}_f = \mathbf{s}_n - \mathbf{s}_0 \quad (12.1)$$

όπου s_n η έξοδος του αλγορίθμου ταιριάσματος της οπτικής μοντελοποίησης.

12.2 Ταξινόμηση με Support Vector Machine

Ο ταξινομητής SVM (Support Vector Machine) είναι μια πολύ διαδεδομένη μέθοδος που με βάση ένα σύνολο παραδειγμάτων εκπαίδευσης με κατάλληλες ετικέτες που δείχνουν αν το εκάστοτε παράδειγμα ανήκει σε μια από τις δυο πιθανές κατηγορίες, δημιουργεί ένα μοντέλο που καταφέρνει να ταξινομήσει νέα στιγμιότυπα σε μια από τις δυο πιθανές κατηγορίες. Πιο συγκεκριμένα η μέθοδος SVM προσπαθεί να κατασκευάσει το υπερεπίπεδο ή ένα σύνολο από υπερεπίπεδα σε ένα χώρο πολλών διαστάσεων έτσι ώστε να μεγιστοποιείται το διάστημα μεταξύ των θετικών και των αρνητικών παρατηρήσεων μιας συγκεκριμένης κατηγορίας.

Στο πείραμα της συγκεκριμένης διπλωματικής εργασίας χρησιμοποιείται SVM με γραμμικό πυρήνα, με τον οποίο η απόφαση της ταξινόμησης για ένα νέο διάνυσμα χαρακτηριστικών \mathbf{x} χωρίς ετικέτα, γίνεται με τη σχέση

$$\mathbf{w} \cdot \mathbf{x} = \mathbf{w}^T \mathbf{x} \begin{cases} >^{true} b \\ <_{false} b \end{cases} \quad (12.2)$$

όπου w είναι το διάνυσμα του υπερεπιπέδου και b είναι το κατώφλι που καθορίζει την απόφαση. Στα πειράματα που ακολουθούν χρησιμοποιείται γραμμικός πυρήνας για το SVM λόγω της ικανότητάς του να δημιουργεί περισσότερο γενικευμένο μοντέλο που αποκρίνεται καλύτερα σε νέα μη ταξινομημένα παραδείγματα. Η συνάρτηση SVM που χρησιμοποιείται είναι αυτή που περιλαμβάνεται στη γλώσσα προγραμματισμού MATLAB.

12.3 Πειραματικά Αποτελέσματα Αναγνώρισης Εκφράσεων Προσώπου

Η αυτόματη αναγνώριση των εκφράσεων του προσώπου ουσιαστικά γίνεται με την ανίχνευση και ταξινόμηση των δράσεων AU του συστήματος FACS. Η συγκεκριμένη

ταξινόμηση έγινε με ένα γραμμικό ένα-εναντίον-όλων μοντέλο SVM (linear one-vs-all) δυο κατηγοριών, δηλαδή του AU υπό μελέτη εναντίον των AUs που δεν είναι υπό μελέτη. Η εκπαίδευση των SVM για κάθε δράση AU έγινε χρησιμοποιώντας όλες τις ουδέτερες εκφράσεις και τις εικόνες κορύφωσης των εκφράσεων όλων των 593 ακολουθιών εικόνων. Οι εικόνες που που στην ετικέτα τους περιλαμβάνεται η συγκεκριμένη AU θεωρούνταν θετικά παραδείγματα και όλες οι υπόλοιπες από τις 1186 θεωρούνταν αρνητικά παραδείγματα. Αυτή η επιλογή έγινε ανεξάρτητα από το κατά πόσο η δράση AU ήταν η μοναδική που εμφανιζόταν στην εικόνα ή ήταν μέρος ενός συνδυασμού δράσεων.

Η τελική επιλογή της ταξινόμησης SVM γίνεται σύμφωνα με τα σκορ που επιτυγχάνονται. Ωστόσο, όπως προτείνεται στο [Luc+10] για να μπορούν τα σκορ να είναι συγκρίσιμα από τα διαφορετικά SVM, θα πρέπει να βαθμονομηθούν με τη μέθοδο της Λογιστικής Γραμμικής Παλινδρόμησης (Logistic Linear Regression). Η βαθμονόμηση έγινε χρησιμοποιώντας την αντίστοιχη συνάρτηση του MATLAB.

Για μεγιστοποίηση του μεγέθους των δεδομένων εκπαίδευσης και δοκιμής χρησιμοποιήθηκε η διαδικασία του Cross Validation με επιλογή του leave-one-subject-out. Εφαρμόζοντας το cross-validation ανά άτομο και όχι ανά ακολουθία εικόνων, είχε ως αποτέλεσμα τη χρήση 123 διαφορετικών συνόλων εκπαίδευσης και δοκιμής. Τονίζεται ότι στο σύνολο των εικόνων προς δοκιμή η εξαγωγή του διανύσματος των χαρακτηριστικών έγινε αυτόματα από το σύστημα οπτικής μοντελοποίησης και το ταίριασμα του AAM. Ακόμη, το ποσοστό επιτυχίας υπολογίζεται ως η περιοχή κάτω από την καμπύλη ROC (Receiver-Operator Characteristic) που είναι η καμπύλη των σωστά-θετικών έναντι των λάθος-θετικών παραδειγμάτων. Το τελικό ποσοστό υπολογίστηκε ως ο μέσος όρος όλων αυτών των καμπυλών. Ο τελικός πίνακας αποτελεσμάτων χρησιμοποιώντας τα διανύσματα χαρακτηριστικών που αναλύθηκαν στην παράγραφο 12.1 φαίνεται στον Πίνακα 12.1.

Τα αποτελέσματα κρίνονται πολύ καλά. Τα ποσοστά είναι μεν χαμηλότερα από αυτά που αναφέρονται στο [Luc+10], ωστόσο κάτι τέτοιο εξηγείται από το γεγονός ότι το πείραμα της συγκεκριμένης δημοσίευσης χρησιμοποιεί σαν διανύσματα χαρακτηριστικών αυτά που έχουν σημειωθεί χειροκίνητα στις εικόνες, ενώ σε αυτή τη διπλωματική εργασία τα διανύσματα χαρακτηριστικών εξάγονται από το ταίριασμα του Active Appearance Model στις εικόνες εισόδου. Επιπλέον, δεν εκτελέστηκαν τα αντίστοιχα πειράματα χρησιμοποιώντας σαν διανύσματα χαρακτηριστικών την υφή του προσώπου, κάτι που αναμένεται να βελτιώσει τα ποσοστά επιτυχίας κάποιων δράσεων AUs.

12.4 Πειραματικά Αποτελέσματα Αναγνώρισης Συνασθημάτων

Η κωδικοποίηση των 7 βασικών συνασθημάτων (θυμός, περιφρόνηση, αηδία, φόβος, ευτυχία, θλίψη, έκπληξη) έγινε σύμφωνα με τον Πίνακα 11.1. Η ταξινόμηση γίνεται με ένα one-vs-all SVM, για παράδειγμα ύπαρξη θυμού ένταντι μη ύπαρξης θυμού (anger vs not anger), ύπαρξη ευτυχίας ένταντι μη ύπαρξης ευτυχίας (happiness vs

AU	Ποσοστό (%)
1	92.2
2	96.3
4	81.7
5	92.1
6	90.8
7	75.5
9	95.9
11	69.8
12	88.9
15	78.1
17	84.4
20	91.0
23	90.8
24	85.0
25	96.1
26	73.6
27	98.3
M.T.	87.1

Πίνακας 12.1: Ποσοστά επιτυχίας αναγνώρισης εκφράσεων προσώπου με χρήση σχήματος προσώπου ως διάνυσμα χαρακτηριστικών

not happiness) κτλ. Όλες οι εικόνες της βάσης δεδομένων CK+ που έχουν την ετικέτα του εκάστοτε συναισθήματος χρησιμοποιούνται ως θετικά παραδείγματα, ενώ όλες οι υπόλοιπες εικόνες ως αρνητικά παραδείγματα.

Ο τρόπος διεξαγωγής του πειράματος είναι ίδιος με αυτόν που εξηγήθηκε στην αναγνώριση εκφράσεων προσώπου με AUs. Χρησιμοποιείται leave-one-subject-out Cross Validation. Ωστόσο ο αριθμός των πειραμάτων είναι μακρύτερος, διότι μόνο 118 από τα 123 άτομα της βάσης δεδομένων έχουν ετικέτες συναισθημάτων. Αυτό οφείλεται στο γεγονός του μη καθορισμούς ετικετών σε περιπτώσεις ακολουθιών εικόνων που η τελική έκφραση συναισθήματος δεν συμφωνούσε με τις προδιαγραφές του συστήματος FACS, παρά το ότι είχε ζητηθεί το συγκεκριμένο συναίσθημα από τα άτομα που βιντεοσκοπούνται. Συνεπώς έχουμε 118 διαφορετικά σύνολα εικόνων εκπαίδευσης και δοκιμής. Ακόμη τονίζεται ότι και πάλι τα χαρακτηριστικά για την αναγνώριση των δράσεων AUs εξάγονται αυτόματα με το σύστημα οπτικής μοντελοποίησης AAM. Τα τελικά ποσοστά επιτυχίας παρουσιάζονται σε μορφή πίνακα confusion για λόγους εποπτείας και φαίνονται στον Πίνακα 12.2.

Για λόγους πληρότητας αναφέρεται ότι οι συντομογραφίες του Πίνακα 12.2 έχουν τις εξής ερμηνείες: (ΘΥΜ)ός, (ΠΕΡ)ιφρόνηση, (ΑΗΔ)ία, (ΦΟΒ)ος, (ΕΥΤ)υχία, (ΘΛΙ)ψη, (ΕΚΠ)ληξη. Παρατηρούμε ότι τα συναισθήματα της Ευτυχίας και της Έκπληξης έχουν πολύ μεγαλύτερη επιτυχία σε σχέση με τα υπόλοιπα. Αυτό εξηγείται από το γεγονός ότι ως συναισθήματα και με τα AUs που τους αντιστοιχήθηκαν, προκαλούν πολύ με-

	<i>ΘΥΜ</i>	<i>ΠΕΡ</i>	<i>ΑΗΔ</i>	<i>ΦΟΒ</i>	<i>ΕΥΤ</i>	<i>ΘΛΙ</i>	<i>ΕΚΠ</i>
<i>ΘΥΜ</i>	32.6	0.0	42.6	0.0	5.5	5.0	3.7
<i>ΠΕΡ</i>	13.6	24.9	4.6	6.7	0.0	11.2	37.8
<i>ΑΗΔ</i>	6.7	2.6	65.3	0.3	10.2	3.3	1.5
<i>ΦΟΒ</i>	5.6	12.9	3.0	21.7	19.8	8.3	9.6
<i>ΕΥΤ</i>	00.0	00.0	00.0	00.0	93.7	0.9	00.0
<i>ΘΛΙ</i>	13.3	21.0	4.0	8.0	0.3	46.1	0.0
<i>ΕΚΠ</i>	0.0	0.0	1.0	0.0	0.0	0.0	94.3

Πίνακας 12.2: Ποσοστά επιτυχίας αναγνώρισης συναισθημάτων με χρήση σχήματος προσώπου ως διάνυσμα χαρακτηριστικών

γάλες παραμορφώσεις στο σχήμα του προσώπου που είναι εύκολο να ανιχνευτούν. Το ίδιο θα μπορούσε να τονιστεί, αν και σε μικρότερο βαθμό για την Αηδία και τη Θλίψη. Αντίθετα, ο Θυμός, η Περιφρόνηση και ο Φόβος έχουν πολύ χαμηλά ποσοστά. Αυτό εξηγείται, κατ' αντιστοιχία, από το γεγονός ότι τα AUs που έχουν αντιστοιχηθεί δεν έχουν μεγάλη σχηματική παραμόρφωση. Οι περιπτώσεις μάλιστα αυτών των συναισθημάτων επηρεάζονται αρκετά από την υφή του προσώπου, πράγμα που σημαίνει ότι αναμένονταν καλύτερα αποτελέσματα σε περίπτωση που χρησιμοποιούνταν ως διανύσματα χαρακτηριστικών οι παράμετροι της υφής του AAM. Εν κατακλείδι, τονίζεται ότι τα ποσοστά του πειράματος είναι χαμηλότερα από αυτά που καταγράφονται στο [Luc+10] διότι δεν χρησιμοποιούνται οι παράμετροι σχήματος των ετικετών, αλλά αυτοί που προκύπτουν από το ταίριασμα του μοντέλου AAM.

Κεφάλαιο 13

Συμπεράσματα

Στην παρούσα διπλωματική εργασία αναλύθηκε η Οπτική Μοντελοποίηση του ανθρώπινου προσώπου σε Πραγματικό Χρόνο και παρουσιάστηκε μια εφαρμογή της για Αυτόματη Αναγνώριση Εκφράσεων Προσώπου και Συναισθημάτων. Το θέμα, που σχετίζεται με τους τομείς της Όρασης Υπολογιστών και της Αναγνώρισης Προτύπων, έχει μεγάλη ανάπτυξη τα τελευταία χρόνια και πληθώρα εφαρμογών. Πιο αναλυτικά:

- Υλοποιήθηκαν δυο μέθοδοι **Ανίχνευσης Δέρματος**: μια πιθανοτική μέθοδος με χρήση μίγματος οκτών κατανομών Gausσιαν και μια απλή μεέθοδος σύγκρισης του σε χρωματικό χώρο HSV. Και οι δυο μέθοδοι εκτελούνται με πολύ μικρό χρονικό κόστος. Ωστόσο επηρεάζονται πολύ από τις συνθήκες φωτισμού με αποτέλεσμα να μην είναι ιδιαίτερα αξιόπιστες. Η πιθανοτική μέθοδος είναι η πιο αποτελεσματική και μπορεί να χρησιμοποιηθεί βοηθητικά για την τελική ανίχνευση προσώπου. Ακόμη υλοποιήθηκε αλγορίθμος **Ανίχνευσης Προσώπου** με χρήση Haar χαρακτηριστικών και ενός Cascade ταξινομητή. Κατόπιν πειραματικών αποτελεσμάτων δείχαμε ότι η απόδοσή του είναι πολύ καλή. Σε περιπτώσεις που το παρασκήνιο της εικόνας είναι πολύπλοκο, αποδείχθηκε ότι η αρχική χρήση της μεθόδου ανίχνευσης περιοχών ανθρώπινου δέρματος βελτιώνει πολύ την αποτελεσματικότητα.
- Αναλύθηκε η **Οπτική Μοντελοποίηση** του ανθρώπινου προσώπου με χρήση Active Appearance Models. Παρουσιάστηκε και αποδείχθηκε μαθηματικά η διαδικασία εκπαίδευσης δυο μοντέλων: ένα μοντέλο σχήματος και ένα μοντέλο υφής προσώπου. Η εκπαίδευση έγινε σε δυο βάσεις δεδομένων, την IMM και την CK+, αν και τα αποτελέσματα που παρουσιάζονται είναι από την IMM. Αποδείχθηκε αλγόριθμος ταιριάσματος του μοντέλου AAM σε νέα πρόσωπα που δεν ανήκουν στη βάση δεδομένων. Τα ποσοστά επιτυχίας είναι ιδιαίτερα υψηλά και χάρη στην παράμετρο της τάξης του αλγορίθμου και στην παράμετρο της ανάλυσης σε pixels του μοντέλου, μπορούν τα αποτελέσματα να διατηρηθούν σε υψηλό επίπεδο καθώς ο χρόνος εκτέλεσης του αλγορίθμου μειώνεται. Επιτεύχθηκε ο στόχος του ταιριάσματος του μοντέλου σε πραγματικό χρόνο και δόθηκε η δυνατότητα στο χρήστη να ρυθμίζει την ακρίβεια των αποτελεσμάτων ταιριάσματος έναντι της χρονικής πολυπλοκότητας.

- Παρουσιάστηκε η σημασία της Οπτικής Μοντελοποίησης σε εφαρμογές αναγνώρισης. Εκτελέστηκαν δυο πειράματα **Αναγνώρισης Εκφράσεων Προσώπου και Συναισθημάτων** με χρήση Support Vector Machine και της βάσης δεδομένων CK+. Τα διανύσματα των χαρακτηριστικών που χρησιμοποιήθηκαν είναι τα διανύσματα σχήματος προσώπου που εξάγονται αυτόματα κατά το ταίριασμα του μοντέλου AAM. Τα αποτελέσματα της αναγνώρισης εκφράσεων προσώπου, που βασίζονται στο σύστημα κωδικοποίησης δράσεων προσώπου Facial Action Coding System (FACS), είναι ικανοποιητικά υψηλά. Ωστόσο τα αποτελέσματα της αναγνώρισης συναισθημάτων με βάση τις εκφράσεις του προσώπου κρίνονται ως μέτρια. Αυτό οφείλεται στη χρήση μόνο των διανυσμάτων σχήματος και όχι και των διανυσμάτων υφής προσώπου και στη μη ακριβή περιγραφή των συναισθημάτων με βάση τις μονάδες δράσεις (Action Units, AU). Τα πειράματα ωστόσο ήταν ενδεικτικά της χρησιμότητας της οπτικής μοντελοποίησης για την αυτόματη εξαγωγή χαρακτηριστικών.
- Αναπτύχθηκε εφαρμογή για την παρούσα διπλωματική εργασία σε MATLAB και C/C++. Η εφαρμογή σχεδιάστηκε με τέτοιο τρόπο ώστε να είναι ολοκληρωμένη, σωστά δομημένη και φιλική προς το χρήστη δίνοντάς του μεγάλο εύρος επιλογών ώστε να ρυθμίζεται την ποιότητα και την ταχύτητα λειτουργίας. Όλα τα απαιτούμενα στάδια εκπαίδευσης των μοντέλων εκτελέστηκαν εκ των προτέρων και αποθηκεύτηκαν σε κατάλληλη μορφή αρχείων, ώστε να μην απαιτείται από το χρήστη η συνεχομένη εκτέλεση της λειτουργίας της εκπαίδευσης, αν και του δίνεται η δυνατότητα. Όλα τα γραφικά του GUI (Graphical User Interface) σχεδιάστηκαν στο MATLAB, δίνοντας έμφαση στη σωστή εμφάνιση. Τέλος, στην παρούσα έκδοση η εφαρμογή υποστηρίζει δυο βάσεις δεδομένων εικόνων, μια για τα Active Appearance Models και μια για την Ανγνώριση Συναισθημάτων, με τη δυνατότητα ωστόσο να προστεθούν περισσότερες.

13.1 Μελλοντική Εργασία

Το θέμα της παρούσας διπλωματικής εργασίας και κυρίως οι εφαρμογές που αυτό έχει σε θέματα ανγνώρισης, το καθιστούν ολοένα και πιο σημαντικό ερευνητικό και επιστημονικό τομέα. Κάποιες προεκτάσεις της διπλωματικής εργασίας που θα αναπτυχθούν στο μέλλον είναι:

- **Υλοποίησης συστήματος αυτόματης αναγνώρισης συναισθημάτων μέσω των εκφράσεων του προσώπου σε πραγματικό χρόνο:** Συγκεκριμένα, τα πειράματα που παρουσιάστηκαν στο τρίτο μέρος θα ενσωματωθούν στο σύστημα οπτικής μοντελοποίησης σε πραγματικό χρόνο. Βέβαια, θα πρέπει να εφαρμοστούν περισσότερες τεχνικές και να χρησιμοποιηθούν επιπλέον διανύσματα χαρακτηριστικών ώστε να βελτιωθούν τα ποσοστά αναγνώρισης. Όμως, η χρονική πολυπλοκότητα της εφαρμογής οπτικής μοντελοποίησης με AAM, δίνει περιθώρια να εισαχθεί και η αναγνώριση συναισθημάτων χωρίς να χαθεί το στοιχείο του πραγματικού χρόνου.

- **Χρησιμοποίηση περισσότερων βάσεων δεδομένων και δοκιμή συνδυασμού δυο ή περισσότερων βάσεων δεδομένων:** Η παρούσα μορφή της διπλωματικής εργασίας περιλαμβάνει δυο βάσεις δεδομένων: την IMM για οπική μοντελοποίηση προσώπου και ταίριασμα AAM και την CK+ για αναγνώριση συναισθημάτων. Ωστόσο, υπάρχουν αρκετές ακόμα βάσεις δεδομένων που μπορούν να δοκιμαστούν. Κάποιες από αυτές παρουσιάζουν πλεονεκτήματα όπως ο μεγάλος αριθμός εικόνων και η καλή ποιότητα εικόνων, αλλά και μειονεκτήματα, όπως η ειδίκευση σε συγκεκριμένες φυλετικές κατηγορίες ανθρώπων ή οι κακές συνθήκες φωτισμού. Συνεπώς, θα δοκιμαστεί ο συνδυασμός δυο ή περισσότερων βάσεων δεδομένων, ώστε το τελικό σύστημα να εκμεταλλεύεται τα πλεονεκτήματα της καθεμιάς που πιθανόν να εξουδετερώνουν τα μειονεκτήματα μιας άλλης.
- **Υλοποίηση τελικού προγράμματος εξ' ολοκλήρου σε C/C++:** Η παρούσα έκδοση του προγράμματος είναι βασισμένη σε MATLAB εκτός από τις πιο σημαντικές συναρτήσεις των αλγορίθμων που έχουν υλοποιηθεί σε C/C++. Το κυριότερο στοιχείο όμως είναι η υλοποίηση του GUI σε MATLAB, που παρά το πλεονέκτημα της ευκολίας στη σχεδίαση, έχει το μειονέκτημα της χαμηλής ταχύτητας λειτουργίας και του συχνού παγώματος. Στόχος λοιπόν είναι η αναβάθμιση του προγράμματος και η μεταφορά του εξ' ολοκλήρου σε C/C++ για ακόμα μεγαλύτερη εξοικονόμηση χρόνου και επίτευξη μεγαλύτερου ρυθμού frames/sec.

Βιβλιογραφία

- [BaHa05] Batur, A. and Hayes, M., "Adaptive active appearance models", IEEE Tr. on Image Proc., 14(11):1707-1721, 2005.
- [BaMa01] Baker, S. and Matthews, I., "Equivalence and efficiency of image alignment algorithms", in Proc. CVPR, volume 1, pages 1090-1097, 2001.
- [BGM04] Baker, S. and Gross, R. and Matthews, I., "Lucas-Kanade 20 years on: A unifying framework - Part 4", Technical Report CMU-R1-TR-04-14, Robotics Institute, CMU, 2004.
- [CET01] Cootes, T. F. and Edwards, G. J. and Taylor, C. J., "Active Appearance Models", IEEE Trans. on Pattern Analysis and Machine Intelligence, vol.23, no.6, pp. 681-685, 2001.
- [CoTa01] Cootes, T. F. and Taylor, C. J., "Constrained active appearance models", in Proc. ICCV, volume 1, pages 748-754, 2001.
- [CoTa06] Cootes, T. F. and Taylor, C. J., "An algorithm for tuning an active appearance model to new data", in Proc. BMVC, 2006.
- [Don+99] Donato, G. and Bartlett, M. and Hager, J. and Ekman, P. and Sejnowski, T., "Classifying Facial Actions", IEEE Trans. on Pattern Analysis and Machine Intelligence, 21(10):974-989, Oct. 1999.
- [EFH02] Ekman, P. and Friesen, W. and Hager, J., "Facial action coding system (FACS)", A Human Face, Salt Lake City, 2002.
- [EkFr78] Ekman, P. and Friesen, W., "Facial action coding system", Palo Alto, CA, USA, 1978.
- [GMB05] Gross, R. and Matthews, I. and Baker, S., "Generic vs. person specific active appearance models", Image and Vision Comp., 23:1080-1093, 2005.
- [Gow75] Gower, J.C., "Generalized procrustes analysis", Psychometrika, 40, 33-51, 1975.
- [HaBe98] Hager, G. and Belhumeur, P., "Efficient region tracking with parametric models of geometry and illumination", IEEE Tr. on PAMI, 20(10):1025-1039, 1998.

- [HRB07] Huang, G. B. and Ramesh, M. and Berg T., "*Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments*", University of Massachusetts, Amherst, Technical Report 07-49, <http://vis-www.cs.umass.edu/lfw/index.html>, October 2007.
- [JKF01] Jesorsky, O. and Kirchberg, K. and Frischholz, R., "*Robust face detection using the hausdorff distance*", Audio-and Video-Based Biometric Person Authentication, pages 90-95, Springer, 2001.
- [Jol89] Jolliffe, I. T., "*Principal component analysis*", Springer-Verlag, New York, 1989.
- [JoRe99] Jones, M. J. and Rehg, J. M., "*Statistical color models with application to skin detection*", In Proc. of the CVPR 99, Vol. 1, 274280, 1999.
- [Ka87] Kass, M., "*Snakes: Active contour models*", International Journal of Computer Vision, 1987.
- [Kap97] Kapur, P., "*Face Detection in Color Images*", University of Washington, Department of Electrical Engineering, 1997.
- [Luc+10] Lucey, P. and Cohn, J.F. and Kanade, T. and Saragih, J. and Ambadar, Z. and Matthews, I., "*The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression*", Computer Vision and Pattern Recognition Workshops (CVPRW), pages 94-101, IEEE, 2010.
- [MaBa04] Matthews, I. and Baker, S., "*Active appearance models revisited*", Int. J. of Comp. Vision, 60(2):135-164, 2004.
- [MeMu99] Menser, B. and Muller, F., "*Face detection in color images using principal components analysis*", Image Processing and Its Applications, Seventh International Conference on (Conf. Publ. No. 465), pages 620-624, 1999.
- [MIB04] Matthews, I. and Ishikawa, T. and Baker S., "*The template update problem*", IEEE Tr. on PAMI, 26(6):810-815, 2004.
- [MMKLM99] Messer, K. and Matas, J. and Kittler, J. and Luettin, J. and Maitre, G., "*Xm2vtsdb: The extended m2vts database*", in AVBPA, 1999.
- [NLSS04] Nordstrøm, M. M. and Larsen, M. and Sierakowski, J. and Stegmann, M. B., "*The IMM Face Database - An Annotated Dataset of 240 Face Images*", <http://www2.imm.dtu.dk/pubdb/p.php?3160>, May 2004.
- [PaMa08] Papandreou, G. and Maragos, P., "*Adaptive and Constrained Algorithms for Inverse Compositional Active Appearance Model Fitting*", in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR-2008), Anchorage, AL, U.S.A., June 2008.

- [PaRo00] Pantic, M. and Rothkrantz L. J. M., "*Expert system for automatic analysis of facial expressions*", Image Vis. Comput., vol. 18, no. 11, pp. 881-905, Aug. 2000.
- [RBK98] Rowley, H.A. and Baluja, S. and Kanade, T., "*Neural network-based face detection*", Pattern Analysis and Machine Intelligence, IEEE Transactions on Vol. 20, pages 23-38, 1998.
- [ScIs03] Sclaroff, S. and Isidoro, J., "*Active blobs: region-based deformable appearance models*", Comput. Vis. Image Underst., 89:197-225, 2003.
- [TaCo92] Taylor, C. J. and Cootes, T. F., "*Active shape models - smart snakes*", in British Machine Vision Conference, 1992.
- [ViJo01a] Viola, P. and Jones, M., "*Robust real-time face detection*", International conference on computer vision, 2001.
- [ViJo01b] Viola, P. and Jones, M., "*Rapid object detection using a boosted cascade of simple features*", Published by the IEEE Computer Society, 2001.