



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ
ΥΠΟΛΟΓΙΣΤΩΝ

Παρακολούθηση Χαρακτηριστικών Προσώπου με Αυτόματη Χρήση Μοντέλων Ενεργού Περιγράμματος

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Κωνσταντίνα Πανταγάκη

Επιβλέπων : Στέφανος Κόλλιας
Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούλιος 2014



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ
ΥΠΟΛΟΓΙΣΤΩΝ

Παρακολούθηση Χαρακτηριστικών Προσώπου με Αυτόματη Χρήση Μοντέλων Ενεργού Περιγράμματος

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Κωνσταντίνα Πανταγάκη

Επιβλέπων : Στέφανος Κόλλιας
Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 18^η Ιουλίου 2014.
Αθήνα, Ιούλιος 2014

.....
Στέφανος Κόλλιας
Καθηγητής ΕΜΠ

.....
Ανδρέας Γεώργιος Σταφυλοπάτης
Καθηγητής ΕΜΠ

.....
Γεώργιος Στάμου
Επίκουρος Καθηγητής ΕΜΠ

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

Copyright © Κωνσταντίνα Πανταγάκη, 2014
Με επιφύλαξη παντός δικαιώματος. All rights reserved.

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

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

Περίληψη

Ο υπέρτατος στόχος της όρασης των υπολογιστών είναι η κατανόηση των εικόνων, η οποία περιλαμβάνει εκτός από την αναγνώριση της δομής τους, και την γνώση του τι αυτή αναπαριστά. Η αναγνώριση των εκφράσεων του ανθρώπινου προσώπου, είναι μία έμφυτη ικανότητα των ανθρώπων, η οποία μάλιστα αναπτύσσεται από πολύ νωρίς. Μία τέτοια ικανότητα είναι πολύ χρήσιμη, ακόμη και απαραίτητη, να την κατέχουν και τα υπολογιστικά συστήματα. Μέσω των εκφράσεων του προσώπου μπορούν να εξαχθούν πολύτιμα συμπεράσματα για τη συναισθηματική κατάσταση του χρήστη, τα οποία αν αξιοποιηθούν σωστά, μπορούν να βελτιώσουν σημαντικά τον τρόπο και την ποιότητα της αλληλεπίδρασης του υπολογιστή με τον χρήστη. Στην αναγνώριση των εκφράσεων, και κατ' επέκταση των συναισθημάτων, βασίζεται πληθώρα εφαρμογών όπως τα video games, η προσωποποιημένη επιλογή μουσικής και τηλεοπτικών προγραμμάτων, τα συστήματα ασφαλείας των αυτοκινήτων κ.α.

Σκοπός αυτής της διπλωματικής εργασίας είναι η ανάπτυξη ενός συστήματος αυτόματης αναγνώρισης των εκφράσεων του προσώπου. Για την αναγνώριση των εκφράσεων, παρακολουθούνται τα χαρακτηριστικά του προσώπου σε ακολουθίες εικόνων. Ο εντοπισμός του προσώπου και των χαρακτηριστικών του γίνεται με τη χρήση στατιστικών μοντέλων εμφάνισης, και ειδικότερα με τα Μοντέλα Ενεργού Περιγράμματος. Η υλοποίηση του συστήματος έγινε σε Matlab και χρησιμοποιήθηκε ήδη υπάρχουσα υλοποίηση των Μοντέλων Ενεργού Περιγράμματος σε C++. Επίσης παρουσιάζεται η σχετική βιβλιογραφία, που αφορά τις εκφράσεις του προσώπου. Αναφέρονται τα χαρακτηριστικά των εκφράσεων, η σημασία τους στην έκφραση των συναισθημάτων, οι εφαρμογές που έχουν κ.α. Τέλος περιγράφονται τα στατιστικά μοντέλα σχήματος, υψής και εμφάνισης καθώς και τα μοντέλα Ενεργού Περιγράμματος.

Λέξεις κλειδιά: Μοντέλα ενεργού περιγράμματος, στατιστικά μοντέλα, όραση υπολογιστή, χαρακτηριστικά προσώπου, εκφράσεις προσώπου, παρακολούθηση προσώπου, αναγνώριση εκφράσεων, αναγνώριση συναισθημάτων.

Abstract

The ultimate goal of machine vision is image-understanding, the ability not only to recover image structure but also to know what it represents. Recognizing the expressions of the human face is an innate ability to the humans that is developed at an early age. This ability can be very useful, for the computers to have it. Face expressions contain a lot of useful information for the user's emotional state. If this information is used correctly, the way the computer interacts with the user, as well as the quality of this interaction, can be significantly improved. A lot of applications are based on expression and emotion recognition, such as video games, personalized music and TV programs, safety systems in the cars, etc.

The goal of this diploma thesis is the development of an automated face expressions recognition system. For the recognition of the expressions, the features of the face are tracked, through a sequence of images. The tracking of the face and its features is performed with statistical models of appearance, and in particular with Active Appearance Models. The system is implemented in Matlab. A previous implementation of the Active Appearance Models in C++ is also used. In the diploma thesis is also presented the literature that is relevant with face expressions. Among others, are presented the characteristics of face expressions, their role in emotion expression, their applications etc. Finally the statistical models of shape, texture, appearance and the Active Appearance Models are described.

Keywords: Active Appearance Models, statistical models, computer vision, face features, face expressions, face tracking, expression recognition, emotion recognition

Ευχαριστίες

Με την ευκαιρία που μου δίνεται, θα ήθελα να ευχαριστήσω τον καθηγητή κ. Στέφανο Κόλλια για την ανάθεση της παρούσας διπλωματικής.

Ευχαριστώ επίσης θερμά τον ερευνητή κ. Κωνσταντίνο Καρπούζη για τη βοήθεια και την καθοδήγηση που μου προσέφερε κατά τη διάρκεια της εκπόνησης της διπλωματικής μου εργασίας.

Τέλος οφείλω ένα μεγάλο ευχαριστώ στην οικογένειά μου για την αγάπη τους και την υποστήριξη που μου παρέχουν πάντα και κυρίως κατά τη διάρκεια των σπουδών μου.

Περιεχόμενα

Περίληψη.....	5
Abstract	6
Ευχαριστίες.....	7
Περιεχόμενα	8
Ευρετήριο Εικόνων.....	10
1. Εισαγωγή.....	11
1.1 Διάρθρωση διπλωματικής εργασίας.....	11
2. Εκφράσεις του Προσώπου	12
2.1 Ιστορική αναδρομή	12
2.2 Γενικά για τις εκφράσεις του προσώπου	13
2.3 Εφαρμογές.....	15
2.4 Παραμετροποίηση του προσώπου	16
2.4.1 Facial Action Coding System (FACS).....	16
2.4.2 Facial Animation Parameters (FAPs)	18
2.5 Παγκόσμιος χαρακτήρας των εκφράσεων του προσώπου και των συναισθημάτων	20
2.6 Οι εκφράσεις του προσώπου ως μέρος ενός συνόλου συναισθηματικών αντιδράσεων.....	22
2.7 Προσωπικότητα και εκφράσεις συναισθημάτων	22
2.8 Περιοχές του εγκεφάλου που ενεργοποιούνται από την παρατήρηση των εκφράσεων του προσώπου.....	23
2.9 Ψυχοπαθολογία και εκφράσεις του συναισθήματος.....	23
2.10 Σημαντικές κοινωνικές λειτουργίες που εξυπηρετούν οι εκφράσεις του προσώπου	23
2.10.1 Πρόκληση αντιδράσεων.....	24
2.10.2 Περιγραφή της φύσης των σχέσεων	24
2.10.3 Ρυθμιστής κοινωνικών συναναστροφών	24
2.10.4 Η σημασία του να κρίνει κανείς με ακρίβεια τα συναισθήματα.....	24
2.11 Χαρακτηριστικά ενός καλού συστήματος αναγνώρισης των εκφράσεων.....	25
2.12 Συμπεράσματα	26
3. Στατιστικά Μοντέλα	27
3.1 Εισαγωγή.....	27
3.2 Στατιστικά Μοντέλα Σχήματος.....	28

3.3 Στατιστικά Μοντέλα Εμφάνισης	33
3.3.1 Στατιστικά Μοντέλα Υφής	33
3.3.2 Συνδυαστικό μοντέλο εμφάνισης	35
3.3.3 Παράδειγμα: Μοντέλο εμφάνισης προσώπου	36
3.3.4 Προσεγγίζοντας ένα νέο παράδειγμα.....	38
3.4 Μοντέλα Ενεργού Περιγράμματος(Active Appearance Models-AAM)	39
3.4.1 Αναζήτηση AAM.....	39
3.4.2 Διόρθωση των παραμέτρων του μοντέλου	39
3.4.3 Διόρθωση των παραμέτρων του μοντέλου επαναληπτικά.....	40
3.4.4 Παραδείγματα της αναζήτησης των μοντέλων ενεργού περιγράμματος σε εικόνες προσώπων.	41
3.4.5 Υλοποίηση	42
4. Ανάπτυξη Συστήματος Παρακολούθησης των Χαρακτηριστικών του Προσώπου.....	43
4.1 Υλοποίηση	43
4.2 Πειραματικά αποτελέσματα.....	49
5. Προκλήσεις και Μελλοντική Εργασία	52
Παράρτημα Α	54
Παράρτημα Β	55
Βιβλιογραφία.....	57

Ευρετήριο Εικόνων

2.1: AUs άνω προσώπου και συνδυασμοί αυτών.	17
2.2: AUs κάτω προσώπου και συνδυασμοί αυτών.	17
2.3: Τα FPs του MPEG-4.	19
3.1: Τα 58 landmark points του προσώπου.....	29
3.2: Καλά ορισμένα landmark points.....	29
3.3: Αριστερά: Κανονικοποιημένη εικόνα με βάση το μέσο σχήμα. Πάνω: Τα σημεία του μοντέλου της κανονικοποιημένης εικόνας. Κάτω: Το κανονικοποιημένο τμήμα της εικόνας	34
3.4: Μεταβολή στο σχήμα	37
3.5: Μεταβολή στην υφή	37
3.6: Μεταβολή της εμφάνισης	38
3.7: Αναπαράσταση προσώπου από συνδυαστικό μοντέλο εμφάνισης.....	38
3.8: Καλύτερο ταίριασμα του μοντέλου	41
3.9: Απεικόνιση σταδίων της αναζήτησης.....	41
4.1: Σημεία που περιγράφουν τη μύτη.....	45
4.2: Σημεία που περιγράφουν το σαγόνι	46
4.3: Σημεία που περιγράφουν τα μάτια.....	46
4.4: Σημεία που περιγράφουν τα φρύδια	47
4.5: Σημεία που περιγράφουν το στόμα.....	47
4.6: Ουδέτερη έκφραση, Χαμόγελο, Ανοιχτό στόμα και μάτια και ανασηκωμένα φρύδια, Συνοφρυωμένος.	50
4.7: Ουδέτερη έκφραση, Κλειστά Μάτια, Συνοφρυωμένος, Χαμόγελο.....	50
4.8: Ουδέτερη έκφραση, Ανοιχτό στόμα, Πιεσμένα χείλη.	51
4.9: Ουδέτερη έκφραση, Ανοιχτό στόμα και μάτια και ανασηκωμένα φρύδια, Κλειστό στόμα, Συνοφρυωμένος.	51
4.10: Ουδέτερη έκφραση, Συνοφρυωμένος και ανοιχτό στόμα, Χαμόγελο, Χαμηλωμένα φρύδια	51
4.11: Λανθασμένη αναγνώριση ως χαμόγελο.....	52
4.12: Λανθασμένη αναγνώριση ως ανοιχτό στόμα.....	52

1. Εισαγωγή

Οι εικόνες του προσώπου παρουσιάζουν ιδιαίτερο ενδιαφέρον όσον αφορά την ανάλυση και την ερμηνεία τους. Εκτός από τις πολλές και ενδιαφέρουσες εφαρμογές τους (βιομετρική ταυτοποίηση, αναγνώριση ομιλίας με τη χρήση βίντεο, παρακολούθηση κινήσεων των ματιών κλπ) οι εικόνες του ανθρώπινου προσώπου έχουν ένα χαρακτηριστικό που τις κάνει να ξεχωρίζουν από τις υπόλοιπες κλάσεις αντικειμένων: όλοι μας είμαστε ειδικοί στην ερμηνεία τους. Μας είναι εύκολο να εντοπίσουμε ακόμα και πολύ μικρές αλλαγές σε ένα πρόσωπο. Πολύ σύντομα μετά τη γέννησή μας είμαστε σε θέση να αναγνωρίζουμε τα πρόσωπα της μητέρας και του πατέρα μας και να τα συσχετίζουμε με αισθήματα ικανοποίησης και ασφάλειας. Λίγο αργότερα αναπτύσσουμε την ικανότητα να αξιολογούμε τις εκφράσεις του προσώπου. Αυτός είναι και ο λόγος που μας ενδιαφέρει τόσο πολύ η ανάλυση των εικόνων του προσώπου.

Στο ανθρώπινο πρόσωπο κωδικοποιείται πληροφορία σχετικά με τα συναισθήματα του ατόμου. Η ερμηνεία λοιπόν των εκφράσεων μπορεί να αποδειχθεί πολύ χρήσιμη σε πολλές περιπτώσεις, κατά τη διάρκεια κοινωνικών συναναστροφών, για την επίλυση προβλημάτων, για την επικοινωνία μεταξύ ατόμων διαφορετικών κουλτουρών, για τον εντοπισμό ψεμάτων κλπ. Λόγω της μεγάλης σημασίας των εκφράσεων και του ρόλου που παίζουν στην καθημερινή μας ζωή έχουν γίνει εκτεταμένες σχετικές έρευνες. Ένα σύστημα το οποίο μπορεί να αναγνωρίζει τις εκφράσεις του χρήστη, αποκτά πιο ανθρώπινο χαρακτήρα, γίνεται λιγότερο απρόσωπο και αλληλεπιδρά με τον χρήστη πιο άμεσα.

1.1 Διάρθρωση διπλωματικής εργασίας

Η εργασία αυτή οργανώνεται σε 5 κεφάλαια. Στο 1^ο κεφάλαιο αναφέρεται συνοπτικά το πρόβλημα της αναγνώρισης των εκφράσεων του προσώπου.

Στο 2^ο κεφάλαιο περιγράφονται τα χαρακτηριστικά των εκφράσεων και η σημασία που αυτές έχουν για τον άνθρωπο. Επίσης γίνεται ανασκόπηση της σχετικής βιβλιογραφίας από τότε που ξεκίνησε η σχετική έρευνα μέχρι και σήμερα.

Το 3^ο κεφάλαιο ασχολείται με τα στατιστικά μοντέλα εμφάνισης και ειδικότερα τα Μοντέλα Ενεργού Περιγράμματος. Μελετάται κυρίως η χρήση τους για τον εντοπισμό του ανθρώπινου προσώπου και των χαρακτηριστικών του σε εικόνες.

Στο 4^ο κεφάλαιο περιγράφεται το σύστημα αυτόματης αναγνώρισης εκφράσεων που αναπτύχθηκε για τους σκοπούς αυτής της εργασίας.

Τέλος στο 5^ο κεφάλαιο αναφέρονται οι προκλήσεις και προτείνεται η μελλοντική εργασία που μπορεί να πραγματοποιηθεί στον τομέα της αυτόματης αναγνώρισης εκφράσεων.

2. Εκφράσεις του Προσώπου

2.1 Ιστορική αναδρομή

Η μελέτη των εκφράσεων του προσώπου και της φυσιολογικής ξεκίνησε ήδη από την εποχή του Αριστοτέλη τον 4^ο αιώνα π.Χ. Η φυσιολογική ορίζεται ως η αξιολόγηση του χαρακτήρα ή της προσωπικότητας ενός ατόμου, από την εξωτερική του εμφάνιση, και κυρίως από το πρόσωπό του. Με το πέρασμα του χρόνου το ενδιαφέρον για την φυσιολογική ελαττώθηκε. Παρόλα αυτά η μελέτη των εκφράσεων του προσώπου συνεχίζει να είναι ένα θέμα με ιδιαίτερο ενδιαφέρον. Κάποιες από τις σημαντικότερες μελέτες σχετικά με τις εκφράσεις του προσώπου, που αποτέλεσαν τη βάση της σημερινής έρευνας, τοποθετούνται χρονικά στον 17^ο αιώνα. Χαρακτηριστικά αναφέρεται ο John Bulwer που στο βιβλίο του Pathomyotomia το 1649 περιγράφει με λεπτομέρεια τις διάφορες εκφράσεις και κινήσεις των μυών του κεφαλιού. Ενδιαφέρον επίσης παρουσιάζει το έργο του γάλλου ακαδημαϊκού και ζωγράφου Le Brun στο οποίο βασίστηκαν οι ηθοποιοί και καλλιτέχνες του 18^{ου} αιώνα προκειμένου να πετύχουν την ‘τέλεια μίμηση των αυθεντικών εκφράσεων του προσώπου’.

Ένα από τα σημαντικότερα έργα της ανάλυσης των εκφράσεων του προσώπου, που επηρεάζει ακόμα και σήμερα την επιστήμη της αυτόματης αναγνώρισης των εκφράσεων, χρονολογείται από τον 19^ο αιώνα. Πρόκειται για το έργο του Καρόλου Δαρβίνου (Charles Darwin). Η πραγματεία που έγραψε ο Δαρβίνος το 1872 θεμελίωσε τις γενικές αρχές σχετικά με τις εκφράσεις του προσώπου και γενικά τους τρόπους με τους οποίους εκφράζονται οι άνθρωποι και τα ζώα. Επίσης ομαδοποίησε τα διάφορα είδη εκφράσεων σε κατηγορίες. Η κατηγοριοποίηση είναι η εξής:

- ακεφιά, άγχος, θλίψη, κατήφεια, απογοήτευση
- χαρά, κέφι, αγάπη, τρυφερότητα, αφοσίωση
- συλλογισμός, διαλογισμός, μελαγχολία, αποφασιστικότητα
- μίσος, θυμός
- περιφρόνηση, καταφρόνηση, αηδία, ενοχή, περηφάνια
- έκπληξη, κατάπληξη, φόβος, τρόμος
- ντροπή, συστολή, μετριοφροσύνη

Επιπλέον κατέγραψε τις παραμορφώσεις του προσώπου που παρατηρούνται για κάθε μία από τις παραπάνω κλάσεις εκφράσεων. Για παράδειγμα ‘η σύσπαση των μυών γύρω από τα μάτια όταν κανείς είναι θλιμμένος’, ‘το χαμήλωμα των γωνιών του στόματος όταν κανείς είναι άκεφος’, κλπ.

Ένα ακόμα σημαντικό ορόσημο στη μελέτη των εκφράσεων και των ανθρώπινων συναισθημάτων είναι το έργο του ψυχολόγου Paul Ekman και των συναδέλφων του από τη δεκαετία του 70. Το έργο τους είναι μεγάλης σημασίας και επηρέασε πολύ τα σημερινά συστήματα αυτόματης αναγνώρισης εκφράσεων.

Παραδοσιακά οι εκφράσεις του προσώπου μελετώνται από ψυχολόγους, γιατρούς, ηθοποιούς και καλλιτέχνες. Όμως από το τέλος του 20^{ου} αιώνα, με τις εξελίξεις που επήλθαν στους τομείς της ρομποτικής, των γραφικών με υπολογιστή και της όρασης των υπολογιστών, εκδηλώθηκε ενδιαφέρον και από σχεδιαστές κινουμένων σχεδίων (animators) και μηχανικούς υπολογιστών για τη μελέτη των εκφράσεων του ανθρώπινου προσώπου.

Το πρώτο βήμα προς την αυτόματη αναγνώριση των εκφράσεων έγινε το 1978 από τον Suwa και τους συνεργάτες του οι οποίοι παρουσίασαν ένα σύστημα για την ανάλυση των εκφράσεων μίας ακολουθίας εικόνων, χρησιμοποιώντας 20 σημεία εντοπισμού. Παρ' ότι το

σύστημα προτάθηκε το 1978, οι ερευνητές δεν ακολούθησαν αυτή τη κατεύθυνση μέχρι και την αρχή της δεκαετίας του 90. Αυτό πιθανότατα οφείλεται στο γεγονός ότι τα συστήματα για τον εντοπισμό του προσώπου που απαιτούνται για την αυτόματη αναγνώριση των εκφράσεων βρίσκονταν ακόμη σε στάδιο ανάπτυξης, μέχρι το 1980.

Στο τέλος της δεκαετίας του 80 και στις αρχές του 90 έγινε διαθέσιμη φτηνή υπολογιστική ισχύς με αποτέλεσμα να αναπτυχθούν συστήματα εντοπισμού του προσώπου. Παράλληλα οι τομείς της *Αλληλεπίδρασης Ανθρώπου- Υπολογιστή (Human-Computer Interaction)* και της *Συναισθηματικής Υπολογιστικής (Affective Computing)* άρχισαν να γίνονται πιο δημοφιλείς. Οι ερευνητές που εργάζονταν σε αυτούς τους τομείς συνειδητοποίησαν ότι χωρίς την δυνατότητα της αυτόματης αναγνώρισης εκφράσεων και συναισθημάτων, οι υπολογιστές θα παρέμεναν 'ψυχροί' και αδιάφοροι για την συναισθηματική κατάσταση του χρήστη. Όλοι οι παραπάνω παράγοντες οδήγησαν στο να ανανεωθεί το ενδιαφέρον για την ανάπτυξη συστημάτων αυτόματης αναγνώρισης εκφράσεων.

2.2 Γενικά για τις εκφράσεις του προσώπου

Ένας από τους τρόπους με τους οποίους δείχνουμε τα συναισθήματά μας είναι μέσω των αλλαγών στις εκφράσεις του προσώπου μας. Μέχρι το τέλος της δεκαετίας του 60, υπήρχε η πεποίθηση ότι οι εκφράσεις του προσώπου δεν ήταν αξιόπιστες και έφεραν μικρή έως καθόλου επικοινωνιακή αξία. Οι σχετικοί συγγραφείς εκείνη την εποχή ανέφεραν πληθώρα παραδειγμάτων (για παράδειγμα ανέφεραν άτομα τα οποία χαμογελούσαν ενώ παρακολουθούσαν τον ακρωτηριασμό ενός αρουραίου), τα οποία αμφισβητούσαν την ένα προς ένα αντιστοίχιση ανάμεσα στις εκφράσεις του προσώπου και στο συναίσθημα που ένιωθε εκείνη τη στιγμή ο άνθρωπος. Σήμερα όμως υπάρχουν αρκετές έρευνες που αποδεικνύουν τη σχέση μεταξύ των εκφράσεων και των συναισθημάτων.

Σύμφωνα με έρευνες, τα χαμόγελα που συνοδεύονται από το σήκωμα των μάγουλων(χαμόγελα Duchenne), συνδέονται με θετικά συναισθήματα σε νέους και μεγαλύτερης ηλικίας ενήλικους. Οι εκφράσεις της αμηχανίας και της διασκέδασης, το αυθόρμητο γέλιο και το χαμόγελο, έχει επίσης αποδειχθεί ότι εκφράζουν διακριτά συναισθήματα. Τέλος η ένταση του γέλιου και του χαμόγελου έχει αποδειχθεί ότι εξαρτάται από το πόσο αστειό ήταν το ερέθισμα που τα προκάλεσε.

Ο Δαρβίνος υποστήριζε ότι οι εκφράσεις του προσώπου αποτελούν μέρος πιο σύνθετων αντιδράσεων και παρουσιάζονται σε συνδυασμό με άλλες αντιδράσεις του σώματος, όπως είναι οι κινήσεις, η στάση του σώματος, η φωνή και οι διάφορες αντιδράσεις του νευρικού συστήματος.

Οι υποστηρικτές της θεωρίας της εξέλιξης επίσης υποστηρίζουν ότι οι εκφράσεις του προσώπου είναι πολλά περισσότερα από απλές ενδείξεις της εσωτερικής κατάστασης του ατόμου. Συντονίζουν τις κοινωνικές μας συναναστροφές, παρέχουν πληροφορίες για την συναισθηματική μας κατάσταση, τις προθέσεις μας, τη σχέση μας με τον αποδέκτη της έκφρασης. Τα παραπάνω συμπεράσματα προέκυψαν από μελέτες στις οποίες συμμετείχαν γονείς που αλληλεπιδρούσαν με τα παιδιά τους καθώς και από παρατήρηση συμπεριφορών όπως το φλερτ, η συμφιλίωση, η άσκηση βίας και το παιχνίδι.

Σημαντικό ρόλο στις εκφράσεις παίζουν τα χαρακτηριστικά του προσώπου. Αυτά μπορούν να χωριστούν σε μόνιμα και παροδικά. Τα μόνιμα χαρακτηριστικά περιλαμβάνουν τα μάτια, τη μύτη, τα χείλη, τα μάγουλα κλπ. Ως παροδικά θεωρούμε τις γραμμές και τις ρυτίδες που δημιουργούνται κατά τη διάρκεια μίας έκφρασης και εξαφανίζονται όταν το πρόσωπο είναι

στην ουδέτερη κατάσταση (πχ οι γραμμές ανάμεσα στα φρύδια όταν κάποιος είναι συνοφρυωμένος).

Το μεγαλύτερο μέρος της πληροφορίας σχετικά με την έκφραση του προσώπου περιέχεται στα φρύδια και στο στόμα. Σύμφωνα με έρευνες, η έκπληξη, η χαρά, και η αηδία έχουν υψηλά ποσοστά επιτυχίας όσον αφορά την αναγνώρισή τους (100%, 93.4% και 97.3% αντίστοιχα) επειδή παρατηρείται έντονη κίνηση στα φρύδια και το στόμα. Ενδιαφέρον επίσης έχει η παρατήρηση ότι το στόμα φέρει περισσότερη πληροφορία από τα φρύδια. Διεξήχθησαν πειράματα με εμφανές μόνο το στόμα, τα οποία είχαν ποσοστό επιτυχίας 78% ενώ όταν ήταν μόνο τα φρύδια εμφανή, το ποσοστό έπεσε στο 50%. Σύμφωνα με διαφορετική μελέτη η πληροφορία για την έκφραση της λύπης περιέχεται κυρίως στο στόμα. Οι μελέτες που διεξάγονται με μόνο κάποια χαρακτηριστικά του προσώπου εμφανή είναι ιδιαίτερα σημαντικές, αφού σε πραγματικές εφαρμογές συνήθως έχουμε διαθέσιμο μόνο μέρος του προσώπου. Καθημερινά αντικείμενα όπως γυαλιά, μαντήλια, γένια, σκιές κλπ τροποποιούν την εικόνα του προσώπου. Ενδιαφέρον επίσης παρουσιάζει το γεγονός ότι αν έχουμε διαθέσιμο μόνο το δεξί ή το αριστερό μισό του προσώπου τα αποτελέσματα δεν επηρεάζονται αφού οι εκφράσεις είναι συμμετρικές στον κάθετο άξονα.

Στις έρευνές τους οι Ekman και Friesen χρησιμοποίησαν μόνο έξι συναισθήματα : χαρά, λύπη, θυμός, έκπληξη, αηδία και φόβος. Αυτά θεωρούνται ως τα βασικά ή αρχέτυπα συναισθήματα. Εκτός από αυτά βέβαια τα 6 συναισθήματα το ανθρώπινο πρόσωπο έχει τη δυνατότητα να εκφράσει μία πληθώρα συναισθημάτων. Οι ερευνητές εξετάζουν και άλλες εκφράσεις που μπορούν να περιγράφουν κάποιο συναίσθημα, μελετώντας και την χρονική εξέλιξη των εκφράσεων, την κατεύθυνση του βλέμματος, το κεφάλι και τη στάση του σώματος.

Έχουν καταγραφεί διακριτές εκφράσεις για την ντροπή και την αμηχανία (Keltner 1995, Keltner & Buswell 1997, Keltner & Harker 1998) για την συμπόνια (Eisenberg & al. 1989), καθώς και διαφορές ανάμεσα στο γέλιο και το χαμόγελο (Keltner & Bonanno 1997). Επίσης έχουν γίνει έρευνες σχετικές με το κοκκίνισμα (Leary, Britt, Cutlip, & Templeton 1992, Shearn et al. 1990) και το βγάλσιμο της γλώσσας (Haidt & Keltner 1999) τα οποία μπορούν επίσης να εκφράζουν συναίσθημα. Το 2000, ο Parrot εντόπισε 136 συναισθηματικές καταστάσεις που εκφράζουν οι άνθρωποι, τις οποίες ταξινόμησε σε διαφορετικές κλάσεις και υποκλάσεις.

Στις περισσότερες περιπτώσεις τα συστήματα που αναπτύσσονται επικεντρώνονται στην αναγνώριση των βασικών και κάποιες φορές και των μη-βασικών εκφράσεων. Λίγες όμως είναι οι προσπάθειες που αφορούν την μελέτη της εξέλιξης των εκφράσεων στο χρόνο. Πιο συγκεκριμένα μας ενδιαφέρει να παρατηρήσουμε την έναρξη, τη κορύφωση και τη λήξη της κάθε έκφρασης.

Μία σημαντική κατηγοριοποίηση των εκφράσεων είναι οι προμελετημένες και οι αυθόρμητες εκφράσεις. Οι προμελετημένες είναι οι τεχνητές εκφράσεις που κάποιος παίρνει όταν αυτό του ζητηθεί. Αυτό συνήθως συμβαίνει όταν το άτομο το οποίο συμμετέχει στην έρευνα βρίσκεται υπό παρατήρηση σε κάποιο εργαστήριο. Αντίθετα οι αυθόρμητες εκφράσεις είναι αυτές που παίρνουν οι άνθρωποι χωρίς να το σκεφτούν. Είναι αυτές που παρατηρούμε στη καθημερινότητά μας, όταν συμμετέχουμε σε συζητήσεις, παρακολουθούμε μία ταινία κλπ. Τα περισσότερα από τα συστήματα αυτόματης αναγνώρισης εκφράσεων, επικεντρώνονται στην αναγνώριση προμελετημένων εκφράσεων επειδή είναι εύκολο να καταγραφούν και να αναγνωριστούν. Είναι πολύ δύσκολο να δημιουργηθεί μία βάση δεδομένων που να περιέχει εικόνες και βίντεο με αυθόρμητες εκφράσεις. Σύμφωνα με έρευνες οι προμελετημένες εκφράσεις μπορούν να περιγράψουν σε ικανοποιητικό βαθμό συναισθήματα όπως η αμηχανία, η ντροπή, το γέλιο και η συμπόνια όχι όμως τόσο καλά όσο περιγράφουν τα βασικά

συναισθήματα. Έχει αποδειχθεί ότι οι αυθόρμητες εκφράσεις διαφέρουν από τις προμελετημένες τόσο στην εμφάνιση όσο και στα χρονικά χαρακτηριστικά. Επίσης οι προμελετημένες εκφράσεις τείνουν να είναι υπερβολικές και να φαίνονται ψεύτικες σε αντίθεση με τις αυθόρμητες που συνήθως είναι πιο διακριτικές.

Μία άλλη κατηγορία είναι οι μικροεκφράσεις (microexpressions). Οι μικροεκφράσεις είναι οι εκφράσεις του προσώπου των ατόμων όταν αυτοί προσπαθούν να κρύψουν τα πραγματικά τους συναισθήματα. Σε αντίθεση με τις αυθόρμητες εκφράσεις, οι μικροεκφράσεις διαρκούν μόνο 1/25^ο του δευτερολέπτου. Λόγω της δυσκολίας που υπάρχει στην καταγραφή των μικροεκφράσεων δεν έχει γίνει σημαντική έρευνα προς αυτή την κατεύθυνση. Ο Ekman έχει μελετήσει τις μικροεκφράσεις και τη χρήση τους για τον εντοπισμό ψεμάτων και εξαπάτησης. Επίσης αναφέρεται στις εκφράσεις που ξεκινούν να φαίνονται στο πρόσωπο κάποιου, πριν αυτός προλάβει να τις κρύψει αλλάζοντας την έκφρασή του. Αυτές διαρκούν περισσότερο από τις μικροεκφράσεις όμως δεν είναι ολοκληρωμένες όσον αφορά τα χρονικά τους χαρακτηριστικά.

Έχει αποδειχθεί ότι η πληροφορία για το αν κάποιος λέει ψέματα, βρίσκεται στις εκφράσεις του προσώπου. Το πρόβλημα είναι ότι δεν μπορούν όλοι οι άνθρωποι να την εντοπίσουν. Οι Ekman, Friesen, O'Sullivan και Scherer το 1980 συνέκριναν τις απαντήσεις παρατηρητών όταν έρχονταν αντιμέτωποι με ειλικρινείς και μη ειλικρινείς εκφράσεις. Όσον αφορά τις ειλικρινείς εκφράσεις, τα συμπεράσματα των παρατηρητών ήταν ακριβή σε αντίθεση με τις μη-ειλικρινείς. Κάποιοι όμως μπόρεσαν να ερμηνεύσουν με ακρίβεια, ακόμα και τις μη-ειλικρινείς εκφράσεις πράγμα που οδηγεί στο συμπέρασμα ότι γενικά υπάρχει δυνατότητα να καταλάβουμε πότε κάποιος λέει ψέματα με απλή παρατήρηση του προσώπου του.

2.3 Εφαρμογές

Τα συστήματα αυτόματης αναγνώρισης εκφράσεων έχουν πολλές ενδιαφέρουσες εφαρμογές. Με τις πρόσφατες εξελίξεις στην ρομποτική, και ιδιαίτερα στα ανθρωποειδή ρομπότ, είναι απαραίτητη η ύπαρξη τέτοιων συστημάτων. Καθώς τα ρομπότ αλληλεπιδρούν όλο και περισσότερο με τους ανθρώπους και γίνονται μέρος των χώρων στους οποίους κατοικούμε και εργαζόμαστε, είναι ανάγκη να γίνουν πιο έξυπνα όσον αφορά την κατανόηση της διάθεσης και των συναισθημάτων των ανθρώπων. Τα συστήματα αναγνώρισης εκφράσεων βοηθούν, παρέχοντας την οπτική διεπαφή μεταξύ των ανθρώπων και του υπολογιστή.

Οι άνθρωποι επικοινωνούν μεταξύ τους και αντιδρούν στην συναισθηματική κατάσταση των ατόμων με τα οποία αλληλεπιδρούν. Αυτή την ικανότητα πρέπει να αποκτήσουν και οι υπολογιστές. Σε αυτό ακριβώς εστιάζει η ερευνητική κοινότητα της Αλληλεπίδρασης Ανθρώπου-Υπολογιστή, στην Συναισθηματική Υπολογιστική. Η αναγνώριση των εκφράσεων παίζει σημαντικό ρόλο στην κατανόηση των συναισθημάτων και βοηθάει στη δημιουργία μίας αποτελεσματικής διεπαφής για την αλληλεπίδραση ανθρώπου-υπολογιστή.

Εκτός από την Ρομποτική και την Συναισθηματική Υπολογιστική τα συστήματα αναγνώρισης εκφράσεων βρίσκουν εφαρμογή και σε άλλους τομείς όπως στις Τηλεπικοινωνίες, στα Video Games, στα Animations, στη Ψυχιατρική, στα συστήματα ασφάλειας των αυτοκινήτων, σε εκπαιδευτικά λογισμικά, στη τηλεόραση και στο ραδιόφωνο για επιλογή του προγράμματος ανάλογα με τη διάθεση του θεατή/ακροατή κλπ.

Παράλληλα έχουν παρουσιαστεί ενδιαφέρουσες και αποτελεσματικές εφαρμογές πραγματικού χρόνου. Για παράδειγμα έχουν δημιουργηθεί συστήματα με τα οποία ένας ψηφιακός χαρακτήρας μπορεί να μιμηθεί τις εκφράσεις του χρήστη, καθώς και συστήματα τα οποία κατά τη διάρκεια του chat μπορούν να εισάγουν στη συνομιλία emoticons ανάλογα με την έκφραση του προσώπου του χρήστη.

Στο μέλλον αναμένεται να παρουσιαστούν πολλές νέες και καινοτόμες εφαρμογές καθώς τα συστήματα αναγνώρισης εκφράσεων γίνονται όλο και πιο αποτελεσματικά ειδικά όσον αφορά τις εφαρμογές πραγματικού χρόνου.

2.4 Παραμετροποίηση του προσώπου

Οι διάφορες συμπεριφορές του προσώπου μπορούν να παραμετροποιηθούν με βάση τις κινήσεις των μυών. Στη συνέχεια αυτές οι παράμετροι μπορούν να χρησιμοποιηθούν για την αναπαράσταση των εκφράσεων του προσώπου. Μέχρι σήμερα υπάρχουν δύο σημαντικές και επιτυχείς τεχνικές για την επίτευξη της παραμετροποίησης:

1. Το Facial Action Coding System(FACS) που αναπτύχθηκε από τους Ekman και Friesen το 1977 και
2. Οι Facial Animation Parameters(FAPs) που αποτελούν μέρος του προτύπου MPEG-4 Synthetic/Natural Hybrid Coding (SNHC), 1998

2.4.1 Facial Action Coding System (FACS)

Πριν τη θεμελίωση του FACS, οι περισσότεροι από τους ερευνητές της συμπεριφοράς του προσώπου βασίζονταν σε ανθρώπινους παρατηρητές, οι οποίοι παρατηρούσαν το υπό εξέταση πρόσωπο και παρουσίαζαν τα συμπεράσματά τους. Κάτι τέτοιο όμως δεν μπορεί να θεωρηθεί επιστημονική μέθοδος καθώς οι παρατηρητές δεν είναι πάντα αξιόπιστοι και ακριβείς. Είναι πολύ εύκολο ένας παρατηρητής να επηρεαστεί από άλλου είδους πληροφορίες όπως η φωνή του ατόμου και επιπλέον τα συμπεράσματα των παρατηρητών είναι στενά συνδεδεμένα με το πολιτισμικό τους υπόβαθρο.

Οι περιορισμοί που παρουσιάζονται από τη χρήση παρατηρητών μπορούν να παρακαμφθούν αναπαριστώντας τις εκφράσεις με ένα σύνολο σταθερών παραμέτρων. Έχοντας ένα τέτοιο πλαίσιο, αρκεί να παρατηρήσουμε ανεξάρτητα τις παραμέτρους χωρίς να λάβουμε υπόψη τη συμπεριφορά του προσώπου στο σύνολό της.

Το FACS είναι μία προσέγγιση που επικεντρώνεται στους μύες. Εντοπίζονται οι διάφοροι μύες του προσώπου οι οποίοι, ανεξάρτητα μεταξύ τους, ή σε ομάδες, προκαλούν αλλαγές στην εμφάνιση του προσώπου. Οι αλλαγές αυτές καθώς και οι μύες(ένας ή περισσότεροι) που προκάλεσαν τις αλλαγές καλούνται Action Units(AU). Το FACS αποτελείται από αρκετές τέτοιες Action Units. Αναφέρονται κάποιες ενδεικτικά:

- AU 1 σήκωμα του εσωτερικού τμήματος του φρυδιού. Προκαλείται από τους μύες Frontalis και Pars Medialis
- AU2 σήκωμα του εξωτερικού τμήματος του φρυδιού. Προκαλείται από τους μύες Frontalis και Pars Lateralis
- AU 26 κατέβασμα του σαγονιού. Προκαλείται από τους μύες Massetter ,Temporal και Internal Pterygoid
















Υπάρχουν όμως και AUs οι οποίες δεν προκαλούνται από μύες του προσώπου. Για παράδειγμα:

- AU 19 γλώσσα έξω
- AU 33 φούσκωμα του μάγουλου κλπ




















Οι AUs μπορούν να είναι προσθετικές ή μη-προσθετικές. Προσθετικές θεωρούνται όταν η εμφάνιση της κάθε AU είναι ανεξάρτητη, και μη-προσθετικές όταν η κάθε μία επηρεάζει την εμφάνιση την άλλης. Γνωρίζοντας τα παραπάνω, είναι εύκολο να αναπαραστήσουμε τις εκφράσεις του προσώπου. Κάθε έκφραση μπορεί να αναπαρασταθεί ως ένας συνδυασμός μίας ή

και περισσότερων, προσθετικών ή μη AUs. Για παράδειγμα, ο φόβος μπορεί να αναπαρασταθεί ως ο συνδυασμός των AUs 1,2 και 26.

Παρακάτω φαίνονται δύο εικόνες με παραδείγματα των AUs του άνω και κάτω τμήματος του προσώπου και των κινήσεων που έχει ως αποτέλεσμα ο συνδυασμός τους.

				
Eyes, brow, and cheek are relaxed.	Inner portion of the brows is raised.	Outer portion of the brows is raised.	Brows lowered and drawn together	Upper eyelids are raised.
AU 6	AU 7	AU 1+2	AU 1+4	AU 4+5
				
Cheeks are raised.	Lower eyelids are raised.	Inner and outer portions of the brows are raised.	Medial portion of the brows is raised and pulled together.	Brows lowered and drawn together and upper eyelids are raised.
AU 1+2+4	AU 1+2+5	AU 1+6	AU 6+7	AU 1+2+5+6+7
				
Brows are pulled together and upward.	Brows and upper eyelids are raised.	Inner portion of brows and cheeks are raised.	Lower eyelids cheeks are raised.	Brows, eyelids, and cheeks are raised.

Εικόνα 2.1: AUs άνω προσώπου και συνδυασμοί αυτών.

				
Lips relaxed and closed.	The infraorbital triangle and center of the upper lip are pulled upwards. Nasal root wrinkling is present.	The infraorbital triangle is pushed upwards. Upper lip is raised. Causes angular bend in shape of upper lip. Nasal root wrinkle is absent.	Lip corners are pulled obliquely.	The lips and the lower portion of the nasolabial furrow are pulled back laterally. The mouth is elongated.
AU15	AU 17	AU 25	AU 26	AU 27
				
The corners of the lips are pulled down.	The chin boss is pushed upwards.	Lips are relaxed and parted.	Lips are relaxed and parted; mandible is lowered.	Mouth stretched open and the mandible pulled downwards.
AU 23+24	AU 9+17	AU 9+25	AU 9+17+23+24	AU 10+17
				
Lips tightened, narrowed, and pressed together.				
AU 10+25	AU 10+15+17	AU 12+25	AU 12+26	AU 15+17
				
AU 17+23+24	AU 20+25			
				

Εικόνα 2.2: AUs κάτω προσώπου και συνδυασμοί αυτών.

Έχουν αναπτυχθεί και άλλες μέθοδοι κωδικοποίησης που είναι παραλλαγές του FACS. Αυτές είναι οι: FAST(Facial Action Scoring Technique), EMFACS(Emotional Facial Action

Coding System), MAX(Maximally Discriminative Facial Movement Coding System), EMG(Facial Electromyography), AFFEX(Affect Expressions by Holistic Judgment), Mondic Phases, FACSaid(FACS Affect Interpretation Database) and Infant/Baby FACS.

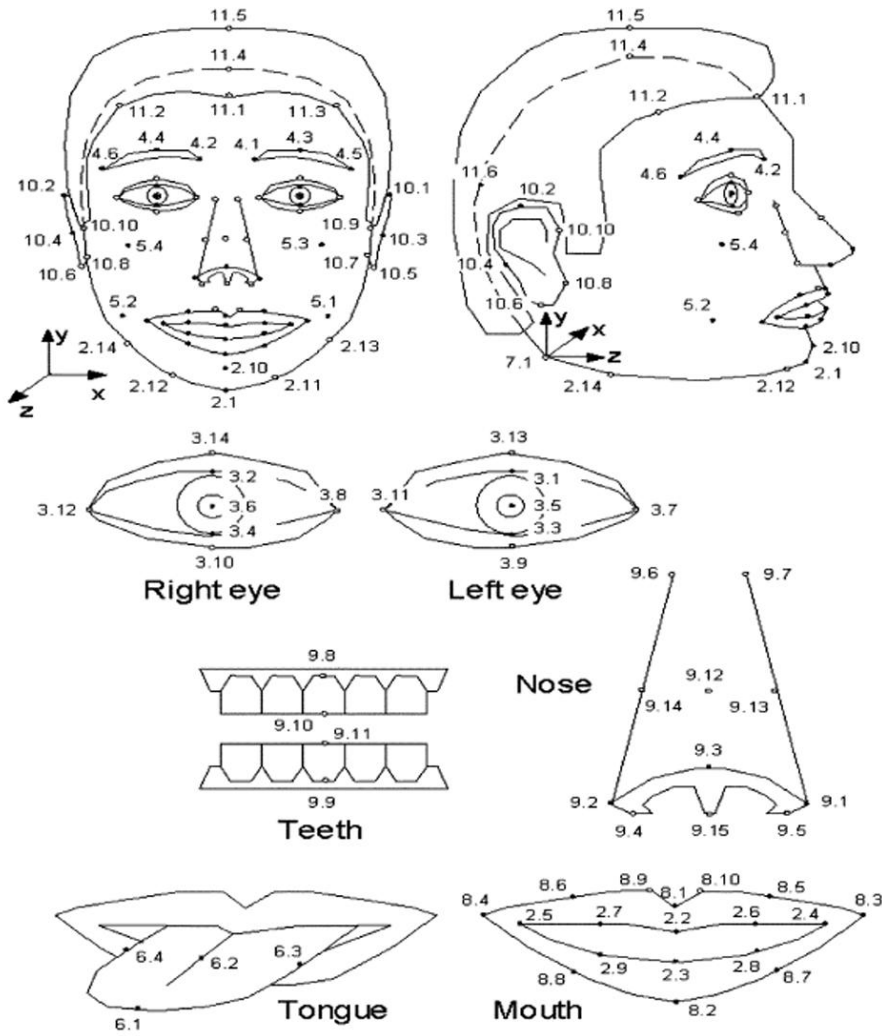
Η αυτόματη αναγνώριση των AUs για δοσμένη εικόνα ή βίντεο είναι ένα δύσκολο εγχείρημα. Αναγνωρίζοντας τις AUs μπορούμε να καθορίσουμε την έκφραση και κατ' επέκταση τη συναισθηματική κατάσταση του ατόμου. Κάποιες φορές για την αναγνώριση ορισμένων AUs (πχ σπρώξιμο σαγονιού προς τα έξω) είναι απαραίτητη η παρατήρηση του προσώπου από το πλάι. Για αυτό το λόγο έχει πραγματοποιηθεί σχετική έρευνα για την κωδικοποίηση εικόνων που έχουν ληφθεί από το πλάι του κεφαλιού(προφίλ).

2.4.2 Facial Animation Parameters (FAPs)

Στη δεκαετία του 90, η ερευνητική κοινότητα του computer animation αντιμετώπισε παρόμοια προβλήματα με αυτά που αντιμετώπιζαν οι ερευνητές της αυτόματης αναγνώρισης εκφράσεων, πριν τη θεμελίωση των FACS. Δεν υπήρχε κάποιο κοινό πρότυπο και σχεδόν κάθε σύστημα που αναπτυσσόταν για τους σκοπούς της σχεδίασης κινουμένων σχεδίων (animation), είχε τις δικές του παραμέτρους. Προκειμένου να αντιμετωπιστεί αυτό το πρόβλημα το Moving Pictures Experts Group(MPEG) εισήγαγε τις προδιαγραφές του Facial Animation(FA) στο πρότυπο MPEG-4. Το MPEG-4 παρέχει τις Facial Animation Parameters(FAPs) οι οποίες είναι στενά συνδεδεμένες με τις Action Units του FACS.

Σύμφωνα με το MPEG-4 ένα μοντέλο του προσώπου στην ουδέτερη κατάστασή, έχει ένα συγκεκριμένο σύνολο ιδιοτήτων όπως α) όλοι οι μύες του προσώπου είναι χαλαροί β) τα βλέφαρα εφάπτονται στην ίριδα γ) η κόρη του ματιού έχει διάμετρο ίση με το 1/3 της διαμέτρου της ίριδας κλπ.

Το πρότυπο επίσης ορίζει 84 σημεία (FPs) στα σημαντικά χαρακτηριστικά του προσώπου όταν αυτό βρίσκεται σε ουδέτερη κατάσταση. Η μετακίνηση των σημείων αυτών χρησιμοποιείται για την κατανόηση και την αναγνώριση των κινήσεων του προσώπου(εκφράσεων) καθώς και για το animation των προσώπων. Στην παρακάτω εικόνα φαίνονται οι θέσεις των 84 σημείων στο πρόσωπο όπως αυτά ορίζονται από το πρότυπο MPEG-4.



Εικόνα 2.3: Τα FPs του MPEG-4.

Οι FAPs είναι ένα σύνολο παραμέτρων οι οποίες αναπαριστούν ένα ολοκληρωμένο σύνολο κινήσεων του προσώπου καθώς και τις κινήσεις του κεφαλιού, της γλώσσας, των ματιών και του στόματος. Με άλλα λόγια κάθε FAP είναι μία κίνηση του προσώπου η οποία το παραμορφώνει από την ουδέτερη κατάσταση. Η τιμή της FAP υποδεικνύει το μέγεθος της κίνησης το οποίο με τη σειρά του υποδεικνύει το μέγεθος της παραμόρφωσης που προκαλείται (πχ ένα μικρό χαμόγελο σε σχέση με ένα πλατύ χαμόγελο). Το MPEG-4 ορίζει 68 FAPs. Στον παρακάτω πίνακα φαίνονται κάποιες από τις FAPs και η περιγραφή τους.

FAP No.	FAP Name	FAP Description
3	open_jaw	Vertical jaw displacement
4	lower_t_midlip	Vertical top middle inner lip displacement
5	raise_b_midlip	Vertical bottom middle inner lip displacement
6	stretch_l_cornerlip	Horizontal displacement of left inner lip corner
7	stretch_r_cornerlip	Horizontal displacement of right inner lip corner

Πίνακας 2.1: Κάποιες από τις FAPs.

Το MPEG-4 ορίζει έξι κύριες εκφράσεις: χαρά, θυμό, λύπη, φόβο, αηδία και έκπληξη. Παρόλο που οι FAPs αρχικά δημιουργήθηκαν για να εξυπηρετήσουν τους σκοπούς του animation, τα τελευταία χρόνια, οι ερευνητές τις χρησιμοποιούν και για την αναγνώριση εκφράσεων και συναισθημάτων. Για παράδειγμα η λύπη μπορεί να εκφραστεί με τις FAPs ως εξής: *close_t_l_eyelid (FAP 19), close_t_r_eyelid(FAP 20), close_b_l_eyelid(FAP 21), close_b_r_eyelid(FAP 22), raise_l_i_eyebrow(FAP 31), raise_r_i_eyebrow(FAP 32), raise_l_m_eyebrow(FAP 33), raise_r_m_eyebrow(FAP 34), raise_l_o_eyebrow(FAP 35), raise_r_o_eyebrow(FAP 36)*.

Σύμφωνα με τα παραπάνω είναι προφανές ότι υπάρχει στενή σχέση μεταξύ των Action Units του FACS και των FAPs. Έχουν γίνει προσπάθειες αντιστοίχισης και ενοποίησης των αρχών της ανάλυσης και της σύνθεσης των εκφράσεων του προσώπου. Κάποια αποτελέσματα της αντιστοίχισης φαίνονται στον παρακάτω πίνακα.

Action Units (FACS AUs)	Facial Action Parameters (MPEG-4 FAPs)
AU1 (Inner Brow Raise)	raise_l_i_eyebrow + raise_r_i_eyebrow
AU2 (Outer Brow Raise)	raise_l_o_eyebrow + raise_r_o_eyebrow
AU9 (Nose Wrinkle)	lower_t_midlip + raise_nose + stretch_l_nose + stretch_r_nose
AU15 (Lip Corner Depressor)	lower_l_cornerlip + lower_r_cornerlip

Πίνακας 2.2: Αντιστοίχιση FACS AUs με FAPs.

2.5 Παγκόσμιος χαρακτήρας των εκφράσεων του προσώπου και των συναισθημάτων

Το ερώτημα που τίθεται σε αυτό το σημείο είναι εάν οι εκφράσεις που περιγράφουν τα συναισθήματα, επηρεάζονται από την κουλτούρα και το πολιτισμικό υπόβαθρο του κάθε ατόμου.

Με βάση τη θεωρία της εξέλιξης ο Δαρβίνος υποστήριζε ότι οι εκφράσεις του προσώπου είναι παγκόσμιες. Για να στηρίξει την πεποίθησή του αυτή, μοίρασε ερωτηματολόγια σε ιεραπόστολους σε διάφορα μέρη του κόσμου, ζητώντας τους να απαντήσουν εάν πίστευαν ότι οι άνθρωποι στα μακρινά αυτά μέρη εξέφραζαν τα συναισθήματά τους με τον ίδιο τρόπο.

Το 1971 οι Ekman και Friesen διεξήγαγαν μελέτες σε άτομα από δυτικές και ανατολικές κουλτούρες και ανέφεραν ότι οι εκφράσεις του προσώπου, που σχετίζονται με τα συναισθήματα,

είναι πράγματι σταθερές και ανεξάρτητες από την κουλτούρα του ατόμου. Από τότε παρά τις κάποιες διαφωνίες έχει επικρατήσει αυτή η άποψη.

Κατά τη διάρκεια ερευνών που διεξήχθησαν σε πολλές διαφορετικές κουλτούρες, παρουσιάστηκαν στους συμμετέχοντες εικόνες που περιείχαν διάφορες εκφράσεις του προσώπου και τους ζητήθηκε να επιλέξουν από μία λίστα, το συναίσθημα εκείνο που περιγράφει καλύτερα την κάθε εικόνα. Όπως αποδείχθηκε, οι άνθρωποι ερμηνεύουν τα συναισθήματα με ακρίβεια μεταξύ του 60% και 80%. Αυτά τα αποτελέσματα οδήγησαν τους ερευνητές στο συμπέρασμα ότι οι άνθρωποι απ' όλο τον κόσμο αντιλαμβάνονται τις εκφράσεις του θυμού, της περιφρόνησης, της αηδίας, του φόβου, της λύπης και της έκπληξης με παρόμοιο τρόπο.

Η ακρίβεια της έρευνας αμφισβητήθηκε λόγω του ότι υπήρχε λίστα συναισθημάτων από την οποία οι συμμετέχοντες επέλεγαν την απάντηση τους και του ότι τα συναισθήματα που παρουσιάζονταν δεν ήταν αυθόρμητα. Παρόλα αυτά μελλοντικές έρευνες έδειξαν ότι τα αποτελέσματα δεν αλλάζουν σημαντικά όταν οι συμμετέχοντες ερμηνεύουν τις εικόνες με δικά τους λόγια ή όταν τους παρουσιάστηκαν εικόνες αυθόρμητων συναισθημάτων.

Ο Ekman το 1973 κατέγραψε τις αντιδράσεις φοιτητών από τις ΗΠΑ και την Ιαπωνία ενώ αυτοί παρακολουθούσαν ένα βίντεο. Οι συμμετέχοντες δεν γνώριζαν ότι οι αντιδράσεις τους καταγράφονταν και έδειξαν παρόμοιες αρνητικές εκφράσεις καθώς παρακολουθούσαν το βίντεο, το οποίο ήταν σχεδιασμένο με στόχο να προκαλεί άγχος. Το 1992 μετά από έρευνα των Camras, Oster, Campos, Miyake, & Bradshaw φάνηκε ότι μωρά 5 και 12 μηνών από τις ΗΠΑ και την Ιαπωνία αντέδρασαν με παρόμοιες εκφράσεις, ήχους και στάση σώματος, καθώς εξέφραζαν το θυμό τους προς ένα αντικείμενο που κρατούσε περιορισμένο το χέρι τους, χωρίς να τους προκαλεί πόνο.

Πιο πρόσφατες έρευνες από τον Horstmann το 2003 και από τους Lawrence et al. το 2005 επιβεβαιώνουν το συμπέρασμα ότι οι άνθρωποι αναγνωρίζουν τα συναισθήματα που αντιστοιχούν στις διαφορετικές εκφράσεις του προσώπου με τον ίδιο τρόπο, ανεξάρτητα από την καταγωγή τους.

Διαφορετική έρευνα συνέλεξε τις αναφορές διαφορετικών ανθρώπων σχετικά με τις εκφράσεις που σχετίζονται με κάθε συναίσθημα. Το συμπέρασμα που προέκυψε είναι ότι άνθρωποι απ' όλο τον κόσμο συμφωνούν στον τρόπο με τον οποίο εκφράζονται τα συναισθήματα. Η αμηχανία, για παράδειγμα, εκφράζεται με ένα νευρικό χαμόγελο και με αποτροπή του βλέμματος.

Μπορεί οι εκφράσεις του προσώπου να είναι παγκόσμια ίδιες όσον αφορά την έκφραση συναισθημάτων, παρόλα αυτά υπάρχουν διαφορές από κουλτούρα σε κουλτούρα στα συναισθήματα ως γενικότερη έννοια. Οι εκφράσεις είναι παγκόσμιες λόγω της σημασίας τους στην αντιμετώπιση διάφορων κοινωνικών δυσκολιών που παρατηρούνται, όπως η ανάπτυξη δεσμών, οι διαπραγματεύσεις κλπ. Άλλες πτυχές των συναισθημάτων, όπως για παράδειγμα η περιγραφή που δίνουν οι άνθρωποι για τα προσωπικά τους συναισθήματα παρουσιάζουν περισσότερη ποικιλομορφία με βάση την κουλτούρα.

Μελέτες που έχουν πραγματοποιηθεί με άτομα διαφορετικής καταγωγής έχουν δείξει ότι παρόλο που η ερμηνεία των συναισθημάτων είναι η ίδια σε όλο τον κόσμο, η έκφραση κάποιων συναισθημάτων, μέσω αλλαγών στην έκφραση του προσώπου, επηρεάζεται από τον κοινωνικό περίγυρο. Για παράδειγμα όταν Αμερικάνοι και Ιάπωνες παρακολούθησαν κάποια βίντεο που τους προκαλούσαν συγκεκριμένα συναισθήματα, έδειξαν παρόμοιες εκφράσεις. Όμως όταν οι Ιάπωνες βρίσκονταν στον ίδιο χώρο με άτομα που έφεραν εξουσία, δίσταζαν να δείξουν τα πραγματικά τους συναισθήματα μέσω αλλαγών στην έκφραση του προσώπου.

2.6 Οι εκφράσεις του προσώπου ως μέρος ενός συνόλου συναισθηματικών αντιδράσεων

Όπως αναφέρθηκε και παραπάνω, σύμφωνα με τον Δαρβίνο, οι εκφράσεις του προσώπου αποτελούν μέρος σύνθετων συναισθηματικών αντιδράσεων. Οι εκφράσεις του προσώπου συνδέονται με την συναισθηματική μας εμπειρία και είναι συντονισμένες με άλλες αντιδράσεις, όπως νευρο-ενδοκρινικές αλλαγές. Οι ερευνητές αναφέρονται στο γεγονός αυτό ως «συναισθηματικό πακέτο» (emotion package).

Οι εκφράσεις του προσώπου σύμφωνα με τη θεωρία αυτή σχετίζονται και με φυσιολογικές αλλαγές του οργανισμού. Για παράδειγμα, σύμφωνα με έρευνες, ο φόβος συνοδεύεται από αντιδράσεις όπως χλόμιασμα, εφίδρωση, διαστολή των κορών των ματιών, ανατριχίλα κλπ. Η λύπη από κλάμα και η αμηχανία από κοκκίνισμα. Στις έρευνες αυτές συμμετείχαν άτομα διαφορετικής καταγωγής (ΗΠΑ, Ινδονησία).

Έντονη σχέση επίσης παρατηρείται μεταξύ των εκφράσεων του προσώπου και της συμπεριφοράς του ατόμου. Στην πρώτη έρευνα που διεξήχθη για τον συγκεκριμένο σκοπό, οι Ekman, Liebert et al. το 1972, εξέτασαν τη σχέση μεταξύ των εκφράσεων και της συμπεριφοράς παιδιών καθώς παρακολουθούσαν τηλεόραση. Κατεγράφησαν οι αντιδράσεις τους καθώς παρακολουθούσαν βίαιες σκηνές από ταινία ή σκηνές από κάποιο ανταγωνιστικό άθλημα. Στη συνέχεια τα παιδιά αυτά βρέθηκαν αντιμέτωπα με μία κατάσταση κατά την οποία είχαν την επιλογή να βλάψουν ή να βοηθήσουν ένα άλλο παιδί. Τα παιδιά που χαμογέλασαν κατά τη διάρκεια των βίαιων σκηνών, φάνηκε στη συνέχεια, να τείνουν προς βίαιη συμπεριφορά, ενώ εκείνα που έδειξαν να στεναχωριούνται επέλεξαν να βοηθήσουν το άλλο παιδί. Αντίστοιχες έρευνες έχουν γίνει και από τους Matsumoto, Haan, Gary, Theodorou και Cooke-Carney σε παιδιά προσχολικής ηλικίας το 1986 και από τους Keltner, Moffit και Stouthamer-Loeber σε έφηβους το 1995 και κατέληξαν σε παρόμοια συμπεράσματα.

2.7 Προσωπικότητα και εκφράσεις συναισθημάτων

Οι Malatesta και Pervin το 1993 υποστήριξαν την άποψη ότι διαφορές στην προσωπικότητα σχετίζονται με διαφορές στα συναισθήματα. Οι Keltner το 1996, Larsen & Ketelaar το 1991 και Watson & Clark το 1992 συμφωνούν με την άποψη αυτή. Οι έρευνες τους έχουν δείξει ότι η εξωστρέφεια και οι νευρώσεις σχετίζονται με εκφράσεις θετικών και αρνητικών συναισθημάτων αντίστοιχα. Άλλες έρευνες έχουν δείξει ότι το κάθε άτομο διαφέρει στην έκφραση των συναισθημάτων του.

Οι Harkel και Keltner το 2001 ανέλυσαν τις φωτογραφίες γυναικών από ένα σχολικό άλμπουμ χρησιμοποιώντας την παραμετροποίηση FACS. Σύμφωνα με τα αποτελέσματα, όσες γυναίκες χαμογέλουσαν έντονα στις φωτογραφίες (χαμόγελο Duchenne) είχαν θετικά χαρακτηριστικά στην προσωπικότητά τους, προσωπική ευτυχία και ικανοποιητικό έγγαμο βίο σε διάφορες χρονικές περιόδους τα επόμενα 30 χρόνια. Αντίστοιχη έρευνα πραγματοποίησαν και οι Abe και Izard το 1999 σε μωρά ηλικίας 18 μηνών καταλήγοντας σε παρόμοια συμπεράσματα. Οι εκφράσεις του προσώπου που περιγράφουν κάποιο διακριτό συναίσθημα, μπορούν να συσχετιστούν με τον τρόπο σκέψης και τις δράσεις που αντιστοιχούν στην προσωπικότητα του ατόμου.

Ενδιαφέρον έχει να σκεφτούμε πότε ξεκινούν αυτές οι αλλαγές να είναι εμφανείς για κάθε άτομο. Κάποιες μελέτες έχουν δείξει αλλαγές στις εκφράσεις του προσώπου μόλις στους 7 πρώτους μήνες της ζωής μας.

2.8 Περιοχές του εγκεφάλου που ενεργοποιούνται από την παρατήρηση των εκφράσεων του προσώπου

Από μελέτες εγκεφαλογραφημάτων και απεικονίσεων του εγκεφάλου προκύπτει ότι η αναγνώριση των εκφράσεων του προσώπου ενεργοποιεί διαφορετικές περιοχές του εγκεφάλου ανάλογα με το συναίσθημα που αντιστοιχεί στην έκφραση.

Η παρατήρηση εκφράσεων φόβου ενεργοποιεί περιοχές στην αριστερή αμυγδαλή, ακόμα και όταν αμέσως μετά την έκφραση φόβου ακολουθήσει μία ουδέτερη έκφραση. Οι θλιμμένες εκφράσεις ενεργοποιούν την αριστερή αμυγδαλή και τον δεξιό κροταφικό λοβό. Οι εκφράσεις θυμού ενεργοποιούν τον δεξιό προμετωπιαίο φλοιό και τον φλοιό του προσαγωγίου. Οι εκφράσεις αηδίας ενεργοποιούν τα βασικά γάγγλια, τον πρόσθιο διάμεσο εγκέφαλο και τον μετωπιαίο λοβό. Υπάρχουν ενδείξεις ότι οι περιοχές που ευθύνονται για την πρόκληση συναισθημάτων είναι οι ίδιες που ενεργοποιούνται και κατά την αναγνώριση των αντίστοιχων συναισθημάτων.

Το ότι το κάθε συναίσθημα αναγνωρίζεται από διαφορετική περιοχή του εγκεφάλου φαίνεται και από έρευνες που έχουν γίνει σχετικά με τα αποτελέσματα τραυματισμών και κακώσεων του εγκεφάλου. Κακώσεις στην περιοχή της αμυγδαλής αφαιρούν από τον ασθενή την ικανότητα να αναγνωρίζει τις εκφράσεις φόβου, ακόμα και όταν αυτός εκφράζεται με επιφωνήματα. Αναγνωρίζουν όμως τις εκφράσεις θλίψης, αηδίας και χαράς. Ασθενείς που πάσχουν από τη νόσο Huntington, η οποία επηρεάζει τα βασικά γάγγλια, αδυνατούν να αναγνωρίσουν εκφράσεις αηδίας αλλά αναγνωρίζουν με ακρίβεια τις εκφράσεις άλλων αρνητικών συναισθημάτων. Το ίδιο πρόβλημα αντιμετωπίζουν ακόμα και οι φορείς της νόσου.

2.9 Ψυχοπαθολογία και εκφράσεις του συναισθήματος

Υπάρχει έντονη συσχέτιση μεταξύ του συναισθήματος και των ψυχολογικών διαταραχών. Έχουν γίνει αρκετές έρευνες σχετικά με το πώς διάφορες διαταραχές σχετίζονται με διάφορες εκφράσεις. Για παράδειγμα όσοι πάσχουν από κατάθλιψη έχουν περιορισμένες εκφράσεις συναισθημάτων, ιδιαίτερα όσον αφορά τις εκφράσεις που σχετίζονται με θετικά συναισθήματα. Όσοι πάσχουν από σχιζοφρένεια δεν επιδεικνύουν τόσο έντονες εκφράσεις στο πρόσωπό τους, σε σχέση με υγιή άτομα, όταν παρακολουθούν συναισθηματικά βίντεο, κινούμενα σχέδια και κατά τη διάρκεια κοινωνικών συναναστροφών. Παρόλα αυτά υποστηρίζουν ότι αισθάνονται το ίδιο ή και περισσότερο έντονα τα ανάλογα συναισθήματα. Έφηβα αγόρια επιρρεπή σε βίαιη και εγκληματική συμπεριφορά εκφράζουν λιγότερο την αμηχανία και περισσότερο τον θυμό.

Τέτοια ευρήματα θεωρούνται εξαιρετικά σημαντικά. Γνωρίζοντας τη σχέση που έχει κάθε συναίσθημα με το κεντρικό νευρικό σύστημα μπορούμε να ανακαλύψουμε μηχανισμούς οι οποίοι συνεισφέρουν σε κάθε διαταραχή. Τέλος οι εκφράσεις του προσώπου μπορούν να χρησιμοποιηθούν για να εκτιμηθεί η πρόοδος του ασθενή σε σχέση με το πώς αντιδρά στη θεραπεία στην οποία έχει υποβληθεί.

2.10 Σημαντικές κοινωνικές λειτουργίες που εξυπηρετούν οι εκφράσεις του προσώπου

Σύμφωνα με τη θεωρία της εξέλιξης, τα συναισθήματα εξελίχθηκαν με τον καιρό, συμβάλλοντας στην επίλυση κοινωνικών προβλημάτων. Οι εκφράσεις είναι χρήσιμες για το σκοπό αυτό για 3 λόγους. Πρώτον παρέχουν πληροφορία για τα συναισθήματα και τις προθέσεις

του ατόμου, για τη σχέση του με το άτομο στο οποίο απευθύνεται η έκφραση και για τη σχέση του με το περιβάλλον. Δεύτερον προκαλούν αντιδράσεις ,κυρίως συναισθήματα, από τους άλλους και τέλος λειτουργούν ως ρυθμιστές για την επιθυμητή κοινωνική συμπεριφορά. Μέρος του πρώτου από τους τρεις αυτούς λόγους έχει αναπτυχθεί παραπάνω. Στη συνέχεια αναπτύσσονται και οι υπόλοιποι.

2.10.1 Πρόκληση αντιδράσεων

Λόγω του καθολικού χαρακτήρα των εκφράσεων του προσώπου ,αυτές περιέχουν πληροφορία όχι μόνο για τις προθέσεις και την μελλοντική συμπεριφορά του ατόμου που τις εκφράζει, αλλά και πληροφορία για το πώς πιθανόν θα αντιδράσει το άτομο το οποίο παρατηρεί την έκφραση. Οι Marsh, Ambady και Klerk το 2005 έδειξαν σε παρατηρητές πρόσωπα που εξέφραζαν φόβο και θυμό και τους ζήτησαν να σπρώξουν ή να τραβήξουν ένα μοχλό όταν έβλεπαν το κάθε πρόσωπο. Ο θυμός οδηγούσε σε συμπεριφορές αποφυγής (σπρώξιμο μοχλού) ενώ ο φόβος σε συμπεριφορές προσέγγισης (τράβηγμα μοχλού). Οι Winkielman, Berridge και Wilbanger το 2005 έδειξαν ότι παρουσιάζοντας στους συμμετέχοντες χαμογελαστά πρόσωπα με τρόπο υποσυνείδητο, αυξανόταν η ποσότητα του ποτού που σέρβιραν και καταναλώναν καθώς και το ποσό που ήταν διατεθειμένοι να πληρώσουν για το ποτό. Παρουσιάζοντας τους αντίθετα, θυμωμένα πρόσωπα , προκαλούταν μείωση των παραπάνω. Επίσης η επίδειξη συναισθημάτων προκαλεί συγκεκριμένες συναισθηματικές αντιδράσεις από τους παρατηρητές. Για παράδειγμα, ο θυμός προκαλεί φόβο, ενώ η θλίψη συμπόνια.

2.10.2 Περιγραφή της φύσης των σχέσεων

Τα πιο σημαντικά ευρήματα στην περιοχή των εκφράσεων του προσώπου ως μέσο έκφρασης συναισθημάτων, προέρχονται από τις μελέτες των Gottman και Levenson που αφορούσαν παντρεμένα ζευγάρια. Για τους σκοπούς της έρευνας, τα ζευγάρια επισκέπτονταν το εργαστήριο, αφού προηγουμένως οι σύζυγοι δεν είχαν συναντηθεί για 24 ώρες, και συζητούσαν καθημερινά θέματα, αντιπαραθέσεις κλπ. Εκφράσεις περιφρόνησης, κυρίως από τους άντρες και αηδίας, κυρίως από τις γυναίκες, προέβλεπαν προβλήματα στο γάμο και κάποιες φορές ακόμα και διαζύγιο.

2.10.3 Ρυθμιστής κοινωνικών συναναστροφών

Η έκφραση των συναισθημάτων μέσω των εκφράσεων του προσώπου, αποτελεί σημαντικό ρυθμιστή των κοινωνικών συναναστροφών. Η άποψη αυτή έχει ερευνηθεί στα πλαίσια της κοινωνικής αναφοράς, όταν δηλαδή τα μικρά παιδιά αναζητούν συναισθηματική πληροφορία από τους γύρω τους για την ερμηνεία ασαφών αντικειμένων και γεγονότων και στη συνέχεια χρησιμοποιούν τη πληροφορία αυτή για να δράσουν.

2.10.4 Η σημασία του να κρίνει κανείς με ακρίβεια τα συναισθήματα

Επειδή οι εκφράσεις του προσώπου είναι αξιόπιστοι δείκτες του συναισθήματος, και επειδή εξυπηρετούν σημαντικές κοινωνικές λειτουργίες , η ικανότητα να τις κρίνει κάποιος με ακρίβεια συνδέεται με σημαντικές διαπροσωπικές διεργασίες. Η ικανότητα της αναγνώρισης των συναισθημάτων έχει συνδεθεί με την συναισθηματική ευφυΐα. Έχει συσχετιστεί με την

επίλυση προβλημάτων, την ικανότητα να θέτει κανείς στόχους και με την αποτελεσματικότητα συνολικά.

Ο Matsumoto μαζί με τους συνεργάτες του, δημιούργησε το Japanese and Caucasian Brief Affect Recognition Test(JACBART) το οποίο βασίζεται στις παρατηρήσεις των Ekman και Friesen σχετικά με τις μικροεκφράσεις. Στο JACBART παρουσιάζονται πολύ σύντομα (0.20 sec) επτά εκφράσεις και στη συνέχεια για 1 δευτερόλεπτο το κάθε πρόσωπο στην ουδέτερη έκφραση.

Οι Yoo, Matsumoto και LeRoux το 2006 χρησιμοποίησαν το JACBART για να δείξουν ότι η ικανότητα διεθνών φοιτητών να αναγνωρίζουν με ακρίβεια τα συναισθήματα, όπως αυτή μετρήθηκε στην αρχή της ακαδημαϊκής χρονιάς, σχετίζεται με πολλούς παράγοντες όπως το άγχος, το πολιτισμικό σοκ, και το ότι τους έλειπε το σπίτι τους. Η αναγνώριση των συναισθημάτων έχει συσχετισθεί επίσης με το μέσο όρο των μαθητών στο τέλος της χρονιάς.

Η υπόθεση ότι οι μικροεκφράσεις σχετίζονται με το ψέμα και την απάτη, αποδεικνύει την σημασία του να μπορεί κάποιος να αναγνωρίζει με ακρίβεια τα συναισθήματα. Ο Ekman το 2003 ανέπτυξε το Micro Expression Training Tool(METT), ένα εργαλείο που επιτρέπει στον καθένα να ελέγξει την ικανότητά του να αναγνωρίζει σωστά τις βασικές εκφράσεις του θυμού, του φόβου, της θλίψης, της αηδίας, της χαράς, της έκπληξης και της περιφρόνησης. Αν κάποιος χρησιμοποιήσει αυτή την τεχνική για να εκπαιδευθεί, αυξάνεται η ακρίβεια του στην αναγνώριση ψεμάτων, κατά τη διάρκεια των οποίων είναι παρούσες οι μικροεκφράσεις. Τα αποτελέσματα αυτά ισχύουν ανεξάρτητα από την καταγωγή των ατόμων

2.11 Χαρακτηριστικά ενός καλού συστήματος αναγνώρισης των εκφράσεων

Ένα καλό σύστημα αναγνώρισης εκφράσεων πρέπει να έχει τα εξής χαρακτηριστικά:

- Να είναι πλήρως αυτοματοποιημένο.
- Να μπορεί να δουλεύει τόσο με εικόνες όσο και με βίντεο.
- Να δουλεύει σε πραγματικό χρόνο.
- Να μπορεί να αναγνωρίζει τις αυθόρμητες εκφράσεις.
- Εκτός από τις βασικές πρέπει να μπορεί να αναγνωρίζει και πολλές άλλες εκφράσεις (ενδεχομένως αναγνωρίζοντας όλες τις Action Units του προσώπου).
- Να είναι αποδοτικό παρά τις αλλαγές στο φωτισμό.
- Να δουλεύει αρκετά καλά ακόμα και όταν είναι ορατό μόνο ένα μέρος του προσώπου.
- Οι εικόνες και το βίντεο που δέχεται σαν είσοδο να μην χρειάζονται προηγούμενη επεξεργασία.
- Τα αποτελέσματα να μην εξαρτώνται από το άτομο του οποίου το πρόσωπο εξετάζεται.
- Να δουλεύει για άτομα διαφορετικής καταγωγής και με διαφορετικό χρώμα δέρματος. Να μην επηρεάζεται από την ηλικία, να μπορεί να αναγνωρίσει τις εκφράσεις μωρών, ενηλίκων και ηλικιωμένων.
- Να μην επηρεάζεται από γένια, γυαλιά, makeup κλπ.
- Να δουλεύει με εικόνες και βίντεο διάφορων αναλύσεων.
- Να αναγνωρίζει εκφράσεις από λήψεις en face, προφίλ και από ενδιάμεσες γωνίες.

Μέχρι τώρα οι διάφορες έρευνες που έχουν γίνει στο συγκεκριμένο τομέα έχουν εστιάσει σε ένα η περισσότερα από τα παραπάνω σημεία. Κανένα σύστημα όμως δεν έχει όλα τα

επιθυμητά χαρακτηριστικά. Προς αυτή την κατεύθυνση πρέπει να κινηθούν οι μελλοντικές έρευνες για την επίτευξη του ιδανικού συστήματος.

2.12 Συμπεράσματα

Κλείνοντας, και σύμφωνα με όσα αναφέρθηκαν παραπάνω, μπορούμε να οδηγηθούμε στα εξής συμπεράσματα για τις εκφράσεις του προσώπου.

- 1) Αποτελούν παγκόσμιους και αξιόπιστους δείκτες των διακριτών συναισθημάτων, όταν δεν γίνεται προσπάθεια από το άτομο να ελέγξει την έκφρασή του.
- 2) Μεταβάλλονται σύμφωνα με την συναισθηματική εμπειρία του ατόμου.
- 3) Είναι μέρος ενός σύνθετου πακέτου συναισθηματικών αντιδράσεων
- 4) Τα συναισθήματα που εκφράζουν αναγνωρίζονται με ακρίβεια από άτομα διαφορετικών культурών.
- 5) Εξυπηρετούν πολλές διαπροσωπικές και ρυθμιστικές κοινωνικές λειτουργίες.

3. Στατιστικά Μοντέλα

3.1 Εισαγωγή

Ο απώτερος στόχος της όρασης των υπολογιστών είναι η κατανόηση της δομής των εικόνων αλλά και τι αυτές αναπαριστούν. Για την επίτευξη του στόχου αυτού είναι απαραίτητη η χρήση μοντέλων. Με τα μοντέλα μπορεί να δοθεί μία περιγραφή και μία ετικέτα στις δομές που αναπαριστώνται στις εικόνες και περιγράφουν αντικείμενα που συναντώνται στον κόσμο μας.

Η διαδικασία αυτή εφαρμόζεται εύκολα σε αντικείμενα με συγκεκριμένη δομή, κυρίως αυτά τα οποία είναι κατασκευασμένα από τον άνθρωπο. Όταν όμως πρόκειται για εικόνες πολύπλοκων και ευμετάβλητων αντικειμένων η ερμηνεία τους με μεθόδους που βασίζονται στη χρήση μοντέλων είναι σαφώς δυσκολότερη. Σε αυτή την κατηγορία εμπίπτουν και οι εικόνες του ανθρώπινου προσώπου. Το πρόβλημα έγκειται στο γεγονός ότι για να είναι ένα μοντέλο επιτυχημένο και χρήσιμο πρέπει να έχει δύο χαρακτηριστικά τα οποία είναι σχεδόν αμοιβαίως αποκλειόμενα μεταξύ τους. Το μοντέλο πρέπει να μπορεί να αναπαριστά μόνο αποδεκτά στιγμιότυπα του αντικειμένου. Ταυτόχρονα όμως είναι αναγκαίο να επιτρέπει ως ένα βαθμό τη μεταβλητότητα (variation) που περιγράφει όλα τα φυσικά αντικείμενα.

Το πρόβλημα αυτό αντιμετωπίζεται από τα στατιστικά μοντέλα τα οποία μπορούν να χρησιμοποιηθούν απευθείας για την ερμηνεία εικόνων. Στα στατιστικά μοντέλα ανήκουν και τα Μοντέλα Ενεργού Περιγράμματος (Active Appearance Models) τα οποία θα μελετηθούν σε αυτή την εργασία. Η παρακάτω περιγραφή των μοντέλων ενεργού περιγράμματος βασίζεται σε εφαρμογές αναγνώρισης προσώπου. Οι ίδιες αρχές όμως έχουν εφαρμογή και σε πολλούς άλλους τομείς.

Υπό πραγματικές συνθήκες οι δομές τις οποίες καλείται να αναγνωρίσει ένα μοντέλο είναι πολύπλοκες και ευμετάβλητες. Πολλές φορές επίσης οι εικόνες που εξετάζονται περιέχουν θόρυβο και μη επαρκή πληροφορία με αποτέλεσμα να είναι σχεδόν αδύνατο να ερμηνευθεί μία εικόνα χωρίς να υπάρχει προηγουμένως γνώση, σχετική με το υπό εξέταση αντικείμενο (πχ ιατρικές απεικονίσεις όταν πρόκειται για όργανα του ανθρώπινου οργανισμού). Οι δυσκολίες αυτές μπορούν να αντιμετωπιστούν με τη χρήση μεθόδων που βασίζονται στη χρήση μοντέλων.

Ιδιαίτερο ενδιαφέρον έχουν τα γενετικά μοντέλα (generative models). Τα μοντέλα αυτά είναι ικανά να παράγουν ρεαλιστικές εικόνες των υπό εξέταση αντικειμένων. Για παράδειγμα ένα γενετικό μοντέλο θα μπορούσε να δημιουργήσει εικόνες του προσώπου οποιουδήποτε ανθρώπου αλλάζοντας την έκφρασή του κλπ. Με τη χρήση ενός τέτοιου μοντέλου η ερμηνεία των εικόνων μετατρέπεται αυτόματα σε ένα πρόβλημα ταιριάσματος. Μεταβάλλοντας τις παραμέτρους του μοντέλου επιχειρούμε να δημιουργήσουμε μία εικόνα η οποία να είναι όσο το δυνατόν πιο όμοια με την πραγματική.

Για να αντιμετωπιστεί το πρόβλημα της μεταβλητότητας που παρουσιάζουν τα αντικείμενα όπως το ανθρώπινο πρόσωπο, ιδιαίτερα χρήσιμα είναι τα παραμορφώσιμα μοντέλα (deformable models). Τα μοντέλα αυτά έχουν την ικανότητα να διατηρούν τα βασικά χαρακτηριστικά της κλάσης των αντικειμένων που αναπαριστούν. Ταυτόχρονα όμως μπορούν να παραμορφωθούν αρκετά ώστε να ταιριάζουν σε ένα ευρύ φάσμα στιγμιότυπων. Αυτού του είδους τα μοντέλα θέλουμε να έχουν δύο κυρίως χαρακτηριστικά. Πρώτον πρέπει να είναι γενικά, δηλαδή να μπορούν να παράγουν κάθε πιθανό στιγμιότυπο της κλάσης την οποία αντιπροσωπεύουν. Δεύτερον πρέπει να είναι ειδικά, δηλαδή τα στιγμιότυπα τα οποία παράγουν να είναι αποδεκτά στιγμιότυπα της κλάσης. Αυτή η δεύτερη ιδιότητα είναι και η πιο σημαντική

καθώς όλο το νόημα της χρήσης τεχνικών που βασίζονται στα μοντέλα είναι να επικεντρωθεί το σύστημά μας σε αποδεκτά στιγμιότυπα.

Για τη δημιουργία ενός στατιστικού μοντέλου αναλύεται η δομή ενός συνόλου παραδειγμάτων, που ονομάζεται σύνολο εκπαίδευσης. Τα παραδείγματα παρουσιάζουν μεταξύ τους διαφορές στο σχήμα και την υφή με αποτέλεσμα να μπορεί το σύστημα μας να εκπαιδευτεί σχετικά με το ποιες μεταβολές θεωρούνται αποδεκτές και ποιες όχι.

Μία τέτοια μέθοδος έχει αρκετά πλεονεκτήματα:

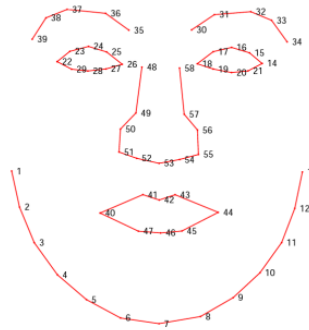
- Έχει ευρεία εφαρμογή. Ο ίδιος αλγόριθμος μπορεί να εφαρμοστεί σε πολλά διαφορετικά προβλήματα χρησιμοποιώντας απλά, διαφορετικά παραδείγματα για τη δημιουργία του μοντέλου.
- Το σύστημά μας εμπεριέχει γνώση σχετική με την εκάστοτε εφαρμογή μέσω των παραδειγμάτων με τα οποία το εκπαιδεύουμε.
- Τα μοντέλα μας δίνουν μία αναπαράσταση των επιτρεπτών μεταβολών, ταυτόχρονα όμως είναι αρκετά ειδικά ώστε να μην επιτρέπουν αυθαίρετες μεταβολές που δεν εμπεριέχονταν στα παραδείγματα του συνόλου εκπαίδευσης.
- Το σύστημα μας δεν χρειάζεται να κάνει σχεδόν καμία αρχική υπόθεση σχετική με τη φύση των αντικειμένων που μοντελοποιούνται, πέρα από αυτά που 'μαθαίνει' από τα παραδείγματα του συνόλου εκπαίδευσης.

Απαραίτητη είναι η ύπαρξη ενός χρήστη ο οποίος θα σημειώσει σε κάθε παράδειγμα του συνόλου εκπαίδευσης κάποια χαρακτηριστικά σημεία (landmark points). Τα σημεία αυτά πρέπει να είναι υπαρκτά σε όλα τα παραδείγματα του συνόλου εκπαίδευσης και να αποτελούν μέρος της δομής του αντικειμένου που εξετάζουμε. Για παράδειγμα, αν εξετάζουμε το ανθρώπινο πρόσωπο, καλά landmark points είναι οι δύο γωνίες του ματιού. Είναι προφανές ότι αυτή η τεχνική περιορίζει το εύρος των αντικειμένων για τα οποία μπορούμε να δημιουργήσουμε μοντέλα. Η τοπολογία των αντικειμένων πρέπει να παραμένει σχετικά αμετάβλητη και η δομή του αντικειμένου να προσφέρεται για την εύρεση κατάλληλων landmark points.

Τα μοντέλα ενεργού περιγράμματος χειρίζονται ένα μοντέλο ικανό να παράγει νέες εικόνες του αντικειμένου που εξετάζουμε. Ο αλγόριθμος των μοντέλων ενεργού περιγράμματος αναζητά τις παραμέτρους εκείνες του μοντέλου οι οποίες παράγουν μία συνθετική εικόνα όσο το δυνατόν πιο όμοια με την αρχική.

3.2 Στατιστικά Μοντέλα Σχήματος

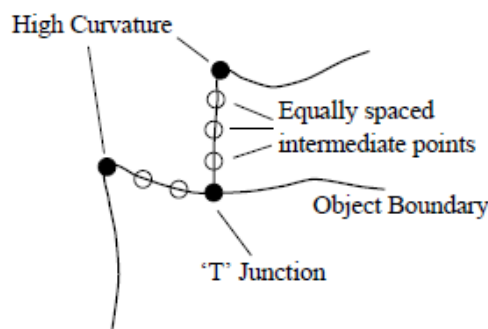
Τα στατιστικά μοντέλα σχήματος (statistical shape models) χρησιμοποιούνται για την αναπαράσταση αντικειμένων σε εικόνες. Η βάση των μοντέλων σχήματος είναι τα 58 landmark points του προσώπου όπως αυτά φαίνονται στην Εικόνα 3.1. Το σχήμα ενός αντικειμένου αναπαρίσταται από ένα σύνολο n σημείων στο χώρο. Το σχήμα συχνά ορίζεται ως η ιδιότητα ενός συνόλου σημείων η οποία δε μεταβάλλεται εάν εφαρμοστεί κάποιος μετασχηματισμός. Στις δύο και τρεις διαστάσεις συνήθως θεωρούμε τον μετασχηματισμό ομοιότητας (μετατόπιση, περιστροφή, κλίμακα).



Εικόνα 3.1: Τα 58 landmark points του προσώπου.

Για τη δημιουργία του μοντέλου το πρώτο πράγμα που πρέπει να γίνει είναι η δημιουργία του συνόλου εκπαίδευσης, σημειώνοντας τα landmark points στα παραδείγματα που θα χρησιμοποιηθούν για την εκπαίδευση. Ο απλούστερος τρόπος για να γίνει αυτό είναι κάποιος άνθρωπος να σημειώσει τα σημεία αυτά σε όλα τα παραδείγματα. Η διαδικασία αυτή είναι χρονοβόρα και επίπονη. Γι αυτό το λόγο έχουν αναπτυχθεί αυτόματες και ημιαυτόματες μέθοδοι ορισμού των landmark points.

Στις δύο διαστάσεις μπορούν να επιλεγούν σημεία που βρίσκονται στις γωνίες, στις ‘διασταυρώσεις’ των ορίων, ή σε ευδιάκριτα χαρακτηριστικά σημεία της δομής των αντικειμένων. Εάν τα παραπάνω σημεία δεν είναι αρκετά για την περιγραφή του σχήματος του αντικειμένου, προσθέτουμε και σημεία που ανήκουν στο περίγραμμα και τα οποία είναι ομοιόμορφα κατανομημένα μεταξύ των καλά ορισμένων σημείων που επιλέξαμε αρχικά (Εικόνα 3.2). Αν επιλέξουμε συνολικά n σημεία σε d διαστάσεις προκύπτει ένα διάνυσμα \mathbf{x} μεγέθους nd . Για παράδειγμα $\mathbf{x} = (x_1, \dots, x_n, y_1, \dots, y_n)^T$, στις 2 διαστάσεις. Αν το σύνολο εκπαίδευσης αποτελείται από s παραδείγματα, έχουμε τελικά s τέτοια διανύσματα \mathbf{x}_j . Τα διανύσματα αυτά θα χρησιμοποιήσουμε στη στατιστική ανάλυση από την οποία θα προκύψει το μοντέλο μας.



Εικόνα 3.2: Καλά ορισμένα landmark points.

Πριν προχωρήσουμε στη στατιστική ανάλυση των διανυσμάτων πρέπει τα σχήματα να ευθυγραμμιστούν. Με αυτό τον τρόπο αφαιρείται η μεταβλητότητα που ενδεχομένως να είναι παρούσα ως αποτέλεσμα μετασχηματισμών. Υπάρχουν πολλοί μέθοδοι με τις οποίες μπορούμε να ευθυγραμμίσουμε τα σχήματα σε κοινό σύστημα συντεταγμένων. Η πιο συνηθισμένη προσέγγιση είναι η ανάλυση του Προκρούστη. Σύμφωνα με αυτή όλα τα σχήματα

ευθυγραμμίζονται έτσι ώστε το άθροισμα των αποστάσεων κάθε σχήματος από το μέσο σχήμα ($D = \sum |x_i - \bar{x}|^2$) να ελαχιστοποιείται. Μια απλή επαναληπτική προσέγγιση είναι η εξής:

1. Μετατόπιση κάθε παραδείγματος ώστε το κέντρο βάρους του να βρίσκεται στην αρχή των αξόνων.
2. Επιλογή ενός παραδείγματος ως μία αρχική εκτίμηση του μέσου σχήματος και κλίμακας ώστε $|\bar{x}|=1$.
3. Καταγραφή της πρώτης εκτίμησης ως \bar{x}_0 για να οριστεί το καρέ αναφοράς.
4. Ευθυγράμμιση όλων των σχημάτων με την τρέχουσα εκτίμηση του μέσου σχήματος.
5. Επανεκτίμηση του μέσου για τα ευθυγραμμισμένα σχήματα.
6. Εφαρμογή περιορισμών στην τρέχουσα εκτίμηση του μέσου, ευθυγραμμίζοντάς το με το \bar{x}_0 έτσι ώστε $|\bar{x}|=1$.
7. Επιστροφή στο βήμα 4 μέχρι να υπάρξει σύγκλιση (σύγκλιση θεωρούμε ότι υπάρχει όταν η εκτίμηση του μέσου δεν αλλάζει σημαντικά μετά από μία επανάληψη).

Στην συνέχεια επιθυμούμε να μοντελοποιήσουμε την μεταβλητότητα των σχημάτων. Έχουμε από τα προηγούμενα βήματα ένα σύνολο s διανυσμάτων τα οποία είναι κατανομημένα στον χώρο διαστάσεων nd . Μοντελοποιώντας την κατανομή αυτή, μπορούμε να παράγουμε νέα παραδείγματα, παρόμοια με αυτά του συνόλου εκπαίδευσης αλλά και να εξετάσουμε νέα σχήματα για να αποφανθούμε αν αυτά είναι αποδεκτά. Αναζητούμε δηλαδή ένα μοντέλο της μορφής $\mathbf{x} = \mathbf{M}(\mathbf{b})$, όπου \mathbf{b} είναι το διάνυσμα των παραμέτρων του μοντέλου. Μεταβάλλοντας τις παραμέτρους \mathbf{b} μπορούμε να παράγουμε νέα διανύσματα \mathbf{x} . Μοντελοποιώντας την κατανομή των παραμέτρων $p(\mathbf{b})$ μπορούμε να τις περιορίσουμε με τέτοιο τρόπο ώστε τα διανύσματα \mathbf{x} που παράγονται να είναι παρόμοια με αυτά του συνόλου εκπαίδευσης.

Για λόγους απλότητας επιθυμούμε να μειώσουμε τις διαστάσεις του παραπάνω προβλήματος. Μία προσέγγιση είναι να εφαρμόσουμε Principal Component Analysis (PCA) στα δεδομένα μας τα οποία αποτελούν ένα σύννεφο σημείων nd -διάστατο χώρο. Η PCA υπολογίζει τους κύριους άξονες του σύννεφου επιτρέποντάς μας να προσεγγίσουμε οποιοδήποτε από τα αρχικά σημεία χρησιμοποιώντας ένα μοντέλο με λιγότερες από nd παραμέτρους. Η προσέγγιση αυτή φαίνεται στη συνέχεια.

1. Υπολογισμός του μέσου όρου των δεδομένων ,

$$\bar{\mathbf{x}} = \frac{1}{s} \sum_{i=1}^s \mathbf{x}_i \quad (3.1)$$

2. Υπολογισμός της συμμεταβλητότητας (covariance) των δεδομένων ,

$$\mathbf{S} = \frac{1}{s-1} \sum_{i=1}^s (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T \quad (3.2)$$

3. Υπολογισμός των ιδιοδιανυσμάτων ϕ_i και των αντίστοιχων ιδιοτιμών λ_i του \mathbf{S} ταξινομημένες έτσι ώστε $\lambda_i \geq \lambda_{i+1}$.

Αν το Φ περιέχει τα t ιδιοδιανύσματα που αντιστοιχούν στις μεγαλύτερες ιδιοτιμές, μπορούμε να προσεγγίσουμε οποιοδήποτε παράδειγμα \mathbf{x} από το σύνολο εκπαίδευσης ως εξής:

$$\mathbf{x} \approx \bar{\mathbf{x}} + \Phi \mathbf{b} \quad (3.3)$$

όπου $\Phi = (\phi_1 | \phi_2 | \dots | \phi_t)$ και \mathbf{b} είναι ένα διάνυσμα διάστασης t που δίνεται από τη σχέση

$$\mathbf{b} = \Phi^T (\mathbf{x} - \bar{\mathbf{x}}) \quad (3.4)$$

Μεταβάλλοντας τα στοιχεία του \mathbf{b} μπορούμε να μεταβάλλουμε τα σχήματα \mathbf{x} μέσω της εξίσωσης 3.3. Η μεταβλητότητα της i -οστής παραμέτρου b_i δίνεται από το λ_i . Εφαρμόζοντας το όριο $\pm 3\sqrt{\lambda_i}$ στην παράμετρο b_i εξασφαλίζουμε ότι το σχήμα που παράγεται είναι παρόμοιο με αυτά του συνόλου εκπαίδευσης.

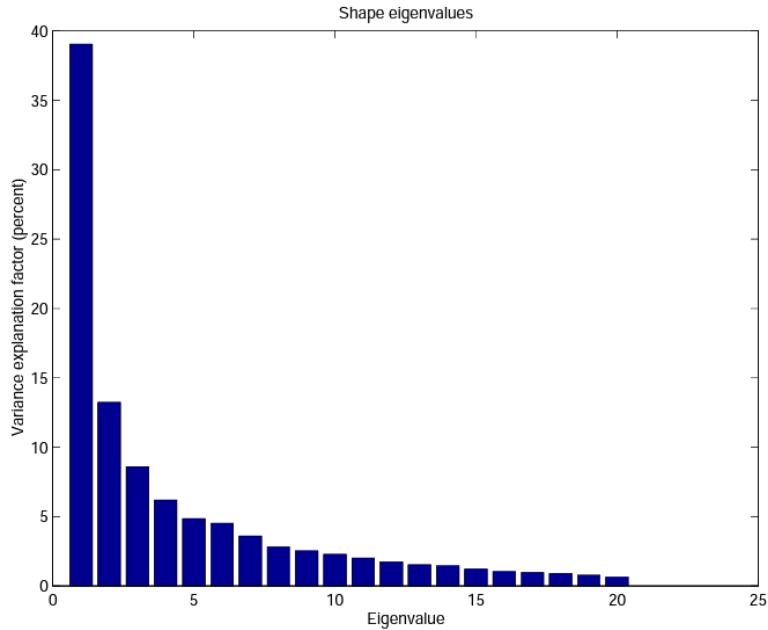
Ο αριθμός t των ιδιοδιανυσμάτων που θα κρατήσουμε μπορεί να επιλεγεί ώστε το μοντέλο τελικά να αντιπροσωπεύει ένα μέρος (πχ 98%) της συνολικής μεταβλητότητας των δεδομένων. Έστω ότι λ_i είναι οι ιδιοτιμές του πίνακα συμμεταβλητότητας του συνόλου εκπαίδευσης. Κάθε ιδιοτιμή δίνει τη διακύμανση των δεδομένων προς την κατεύθυνση του αντίστοιχου ιδιοδιανύσματος. Η συνολική διακύμανση του συνόλου εκπαίδευσης είναι το άθροισμα όλων των ιδιοτιμών, $V_T = \sum \lambda_i$. Μπορούμε επομένως να επιλέξουμε τις t μεγαλύτερες ιδιοτιμές ώστε $\sum_{i=1}^t \lambda_i \geq \frac{p}{100} V_T$ όπου p είναι το ποσοστό της συνολικής μεταβλητότητας που επιθυμούμε να διατηρήσουμε. Συνήθως υποθέτουμε ότι ένα 5% αντιστοιχεί σε θόρυβο. Η επιλογή μίας μεγάλης τιμής για το p θα έχει ως αποτέλεσμα το μοντέλο να είναι πολύ ευέλικτο και να λαμβάνει υπόψη του και τον θόρυβο που υπάρχει στο σύνολο εκπαίδευσης, άρα να μην γενικεύει όπως θα έπρεπε (υψηλή μεταβλητότητα). Από την άλλη μεριά, μία χαμηλή τιμή του p θα έχει ως αποτέλεσμα ένα πολύ περιοριστικό μοντέλο το οποίο δε θα ταιριάζει επαρκώς στο σύνολο εκπαίδευσης.

Κατασκευάζουμε ένα μοντέλο χρησιμοποιώντας ένα dataset που αποτελείται από 37 διαφορετικές εικόνες προσώπων, χωρίς γυαλιά και με ουδέτερη έκφραση. Πρόκειται για 7 γυναίκες και 30 άνδρες. Οι εικόνες είναι έγχρωμες, αρχικά 640x480 JPEG που στη συνέχεια μετατράπηκαν σε BMP. Στις εικόνες αυτές σημειώθηκαν με το χέρι τα 58 landmark points στα φρύδια, μάτια, μύτη, στόμα και σαγόι. Συνολικά χρησιμοποιήθηκαν 7 μονοπάτια, τρία κλειστά και τέσσερα ανοιχτά (Εικόνα 3.1).

Για το μοντέλο που δημιουργήθηκε φαίνεται παρακάτω ο πίνακας που περιλαμβάνει τα ποσοστά της μεταβλητότητας που αναπαριστάται από τις 10 μεγαλύτερες ιδιοτιμές. Στο διάγραμμα φαίνονται οι 20 μεγαλύτερες ιδιοτιμές που συνολικά αναπαριστούν το 95% της συνολικής μεταβλητότητας.

Mode	Variance	Acc. variance
1	37.34%	37.34%
2	12.66%	50.00%
3	8.22%	58.22%
4	5.92%	64.14%
5	4.64%	68.77%
6	4.32%	73.10%
7	3.45%	76.55%
8	2.69%	79.24%
9	2.43%	81.67%
10	2.18%	83.85%

Πίνακας 3.1: Μεταβλητότητα που αναπαρίσταται από τις 10 μεγαλύτερες ιδιοτιμές.



Διάγραμμα 3.1: Μεταβλητότητα που αναπαρίσταται από τις 20 μεγαλύτερες ιδιοτιμές.

Τα μοντέλα σχήματος χρησιμοποιούν γραμμικούς συνδυασμούς των σχημάτων του συνόλου εκπαίδευσης. Προκειμένου να μπορούν να ταιριάζουν καλά σε ένα νέο σχήμα πρέπει το σύνολο εκπαίδευσης να αναπαριστά όλη την αναμενόμενη μεταβλητότητα της κλάσης των σχημάτων που μοντελοποιούνται. Διαφορετικά το μοντέλο δεν θα είναι αρκετά γενικό ώστε να ταιριάζει με όλα τα αποδεκτά παραδείγματα. Για παράδειγμα, αν ένα μοντέλο έχει εκπαιδευτεί μόνο πάνω σε τετράγωνα δε θα μπορεί να γενικευθεί σε ένα ορθογώνιο αφού αυτό δεν αντιπροσωπεύεται στο το σύνολο εκπαίδευσης.

Ένας τρόπος για να ελέγξουμε πόσο γενικό είναι το μοντέλο μας, είναι διεξάγοντας miss-one-out πειράματα. Έστω ότι το σύνολο εκπαίδευσης αποτελείται από s παραδείγματα. Κατασκευάζουμε ένα μοντέλο από τα $s-1$ παραδείγματα και στη συνέχεια προσπαθούμε να ταιριάζουμε το μοντέλο στο παράδειγμα που δεν χρησιμοποιήσαμε για την κατασκευή του, υπολογίζοντας το σφάλμα. Επαναλαμβάνουμε την παραπάνω διαδικασία για όλα τα παραδείγματα του συνόλου εκπαίδευσης. Αν κάποιο από τα σφάλματα είναι μεγάλο, πιθανόν να χρειάζονται περισσότερα παραδείγματα στο σύνολο εκπαίδευσης. Εάν το σφάλμα είναι μικρότερο για όλα τα παραδείγματα σημαίνει ότι υπάρχουν περισσότερα από ένα παραδείγματα για κάθε τύπο μεταβλητότητας, όχι όμως και ότι όλοι οι τύποι αντιπροσωπεύονται επαρκώς.

Όταν το σύνολο εκπαίδευσης αποτελείται από λίγα μόνο παραδείγματα είναι φυσικό, μέρος της μεταβλητότητας να μην αναπαρίσταται σε αυτό. Σε αυτή τη περίπτωση μπορούμε με κάποιες τεχνικές να προσθέσουμε την επιπλέον μεταβλητότητα που θα κάνει το μοντέλο μας πιο ευέλικτο. Οι τεχνικές αυτές είναι συνοπτικά οι εξής:

- **Finite Element Models (FEM):** Αντιμετωπίζουμε ένα σχήμα σαν να ήταν κατασκευασμένο από ελαστικό υλικό. Παραμορφώνοντας το σχήμα αυτό μπορούμε να δημιουργήσουμε με τεχνητό τρόπο τη μεταβλητότητα που λείπει από το σύνολο εκπαίδευσης. Ο τρόπος που παραμορφώνουμε το σχήμα όμως είναι αυθαίρετος, επομένως η μεταβλητότητα που προκύπτει μπορεί να μην αναπαριστά αποδεκτά παραδείγματα της κλάσης.

- **Συνδυασμός Στατιστικών Μοντέλων και FEM:** Τόσο τα στατιστικά μοντέλα όσο και τα FEM είναι γραμμικά μοντέλα επομένως είναι εύκολο να τα συνδυάσουμε. Εάν έχουμε ένα μόνο παράδειγμα τότε αναγκαστικά χρησιμοποιούμε τα FEM. Εάν έχουμε δύο παραδείγματα, μπορούμε να κατασκευάσουμε ένα στατιστικό μοντέλο το οποίο όμως θα είναι η γραμμική παρεμβολή των δύο παραδειγμάτων. Μία λύση είναι να δημιουργήσουμε νέα παραδείγματα χρησιμοποιώντας τα FEM και με τα παραδείγματα αυτά και τα αρχικά, να κατασκευάσουμε ένα στατιστικό μοντέλο. Και πάλι πρέπει να είμαστε προσεκτικοί όσον αφορά τη χρήση των FEM, ώστε να μην συμπεριληφθεί στο μοντέλο μας μη αποδεκτή μεταβλητότητα.

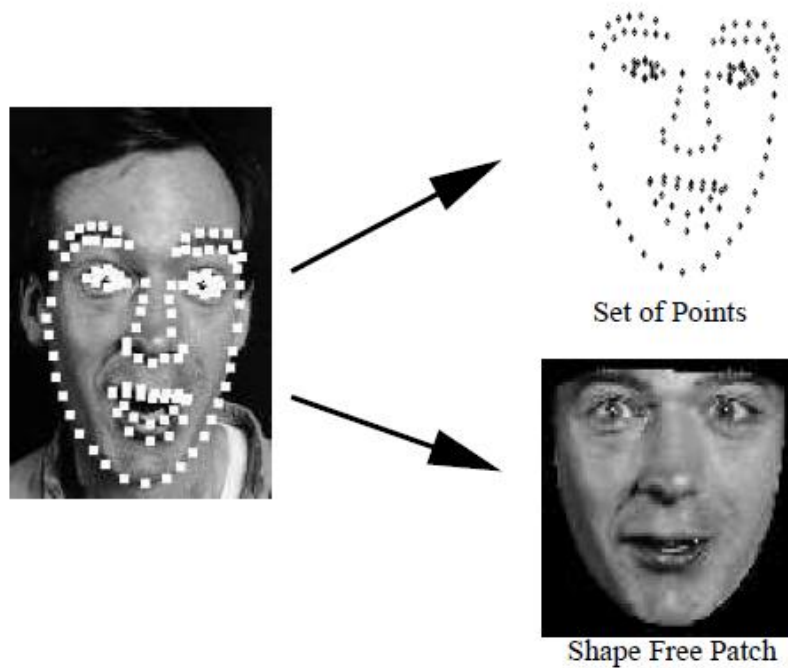
3.3 Στατιστικά Μοντέλα Εμφάνισης

Για να συνθέσουμε μία ολοκληρωμένη εικόνα ενός αντικειμένου ή μίας δομής πρέπει να μοντελοποιήσουμε τόσο το σχήμα όσο και την υφή του(το χρώμα ή τη φωτεινότητα στην περιοχή του αντικειμένου). Στη συνέχεια θα περιγράψουμε πώς μπορούμε να δημιουργήσουμε στατιστικά μοντέλα που να αναπαριστούν τη μεταβλητότητα στο σχήμα, στην υφή και τη μεταξύ τους συσχέτιση.

Αρχικά από το σύνολο εκπαίδευσης δημιουργούμε ένα στατιστικό μοντέλο σχήματος όπως περιγράφηκε παραπάνω. Έπειτα δημιουργούμε ένα στατιστικό μοντέλο της μεταβλητότητας της υφής. Υπάρχει κάποια συσχέτιση μεταξύ των παραμέτρων του μοντέλου σχήματος και αυτών του μοντέλου υφής σε κάθε σύνολο εκπαίδευσης. Λαμβάνουμε τη συσχέτιση αυτή υπ' όψη δημιουργώντας ένα μοντέλο εμφάνισης το οποίο ελέγχει τόσο το σχήμα όσο και την υφή.

3.3.1 Στατιστικά Μοντέλα Υφής

Για να κατασκευάσουμε ένα στατιστικό μοντέλο υφής τροποποιούμε κάθε ένα από τα παραδείγματα του συνόλου εκπαίδευσης ούτως ώστε τα landmark points του να ταυτίζονται με αυτά του μέσου σχήματος. Έτσι αφαιρούμε από το παράδειγμα την πλασματική πληροφορία σχετικά με την υφή, που ενδεχομένως να υπάρχει λόγω των διαφορών στο σχήμα. Στη συνέχεια λαμβάνουμε δείγματα της υφής της κανονικοποιημένης εικόνας στην περιοχή που καλύπτεται από το μέσο σχήμα. Με αυτόν τον τρόπο δημιουργούμε ένα διάνυσμα υφής, το \mathbf{g}_{im} . Στην παρακάτω εικόνα φαίνεται η κανονικοποιημένη εικόνα που προκύπτει από την αρχική και το μέσο σχήμα, καθώς και η περιοχή από την οποία λαμβάνουμε τα δείγματα της υφής.



Εικόνα 3.3: Αριστερά: Κανονικοποιημένη εικόνα με βάση το μέσο σχήμα. Πάνω: Τα σημεία του μοντέλου της κανονικοποιημένης εικόνας. Κάτω: Το κανονικοποιημένο τμήμα της εικόνας από το οποίο θα ληφθούν τα δείγματα της φωτεινότητας.

Εφαρμόζοντας PCA στα κανονικοποιημένα δεδομένα προκύπτει το γραμμικό μοντέλο:

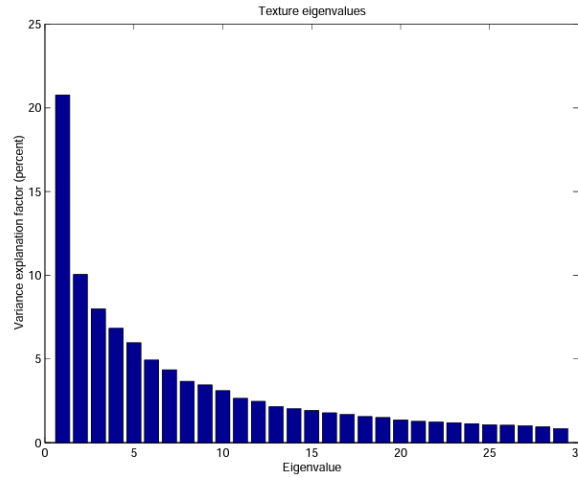
$$\mathbf{g} = \bar{\mathbf{g}} + \mathbf{P}_g \mathbf{b}_g \quad (3.5)$$

όπου $\bar{\mathbf{g}}$ είναι το μέσο κανονικοποιημένο διάνυσμα της φωτεινότητας, \mathbf{P}_g είναι ένα σύνολο ιδιοδιανυσμάτων και \mathbf{b}_g είναι ένα σύνολο παραμέτρων φωτεινότητας.

Χρησιμοποιώντας το ίδιο dataset που χρησιμοποιήθηκε για την κατασκευή του μοντέλου σχήματος, κατασκευάζουμε ένα μοντέλο υφής. Στον παρακάτω πίνακα φαίνεται η μεταβλητότητα που αναπαριστάται από τις 10 μεγαλύτερες ιδιοτιμές και στο διάγραμμα φαίνονται οι 29 μεγαλύτερες ιδιοτιμές που συνολικά αναπαριστούν το 95% της μεταβλητότητας.

Mode	Variance	Acc. variance
1	19.80%	19.80%
2	9.58%	29.38%
3	7.61%	36.99%
4	6.52%	43.51%
5	5.69%	49.20%
6	4.71%	53.91%
7	4.15%	58.06%
8	3.50%	61.56%
9	3.29%	64.84%
10	2.96%	67.81%

Πίνακας 3.2: Μεταβλητότητα που αναπαρίσταται από τις 10 μεγαλύτερες ιδιοτιμές.



Διάγραμμα 3.2: Μεταβλητότητα που αναπαρίσταται από τις 29 μεγαλύτερες ιδιοτιμές.

3.3.2 Συνδυαστικό μοντέλο εμφάνισης

Το σχήμα και η υφή οποιουδήποτε παραδείγματος μπορούν πια να δοθούν από τα διανύσματα παραμέτρων \mathbf{b}_s και \mathbf{b}_g . Λόγω των συσχετίσεων που ενδεχομένως να υπάρχουν μεταξύ της μεταβλητότητας στο σχήμα και στην υφή, εφαρμόζουμε ξανά στα δεδομένα PCA. Για κάθε παράδειγμα κατασκευάζουμε τα διάνυσμα

$$\mathbf{b} = \begin{pmatrix} \mathbf{W}_s \mathbf{b}_s \\ \mathbf{b}_g \end{pmatrix} = \begin{pmatrix} \mathbf{W}_s \mathbf{P}_s^T (\mathbf{x} - \bar{\mathbf{x}}) \\ \mathbf{P}_g^T (\mathbf{g} - \bar{\mathbf{g}}) \end{pmatrix} \quad (3.6)$$

όπου \mathbf{W}_s είναι ένας διαγώνιος πίνακας βαρών για κάθε παράμετρο σχήματος. Εφαρμόζουμε στη συνέχεια PCA σε αυτά τα διανύσματα και προκύπτει το μοντέλο

$$\mathbf{b} = \mathbf{P}_c \mathbf{c} \quad (3.7)$$

όπου \mathbf{P}_c είναι τα ιδιοδιανύσματα και \mathbf{c} είναι ένα διάνυσμα παραμέτρων εμφάνισης που ελέγχει τόσο το σχήμα όσο και τη φωτεινότητα του μοντέλου.

Λόγω της γραμμικής φύσης του μοντέλου, μπορούμε να εκφράσουμε το σχήμα και τη φωτεινότητα απ' ευθείας ως συναρτήσεις του \mathbf{c} :

$$\mathbf{x} = \bar{\mathbf{x}} + \mathbf{P}_s \mathbf{W}^{-1} \mathbf{P}_{cs} \mathbf{c}, \quad \mathbf{g} = \bar{\mathbf{g}} + \mathbf{P}_g \mathbf{P}_{cg} \mathbf{c} \quad (3.8)$$

όπου $\mathbf{P}_c = \begin{pmatrix} \mathbf{P}_{cs} \\ \mathbf{P}_{cg} \end{pmatrix}$ (3.9) ή συνοπτικά ως:

$$\begin{aligned} \mathbf{x} &= \bar{\mathbf{x}} + \mathbf{Q}_s \mathbf{c} \\ \mathbf{g} &= \bar{\mathbf{g}} + \mathbf{Q}_g \mathbf{c} \end{aligned} \quad (3.10)$$

όπου

$$\begin{aligned} \mathbf{Q}_s &= \mathbf{P}_s \mathbf{W}^{-1} \mathbf{P}_{cs} \\ \mathbf{Q}_g &= \mathbf{P}_g \mathbf{P}_{cg} \end{aligned} \quad (3.11)$$

Μπορούμε να συνθέσουμε την εικόνα ενός παραδείγματος για δοσμένο \mathbf{c} κατασκευάζοντας την κανονικοποιημένη εικόνα από το διάνυσμα \mathbf{g} και τροποποιώντας την σύμφωνα με τα landmark points που δίνονται από το \mathbf{x} .

Κατασκευάζουμε ένα συνδυαστικό μοντέλο εμφάνισης από τα παραπάνω μοντέλα σχήματος και υφής. Στον παρακάτω πίνακα φαίνονται οι 10 μεγαλύτερες ιδιοτιμές και το ποσοστό της μεταβλητότητας που αυτές αναπαριστούν.

Mode	Variance	Acc. variance
1	22.74%	22.74%
2	12.59%	35.33%
3	7.82%	43.16%
4	5.81%	48.96%
5	5.17%	54.13%
6	4.29%	58.42%
7	4.00%	62.42%
8	3.42%	65.84%
9	3.14%	68.98%
10	2.94%	71.91%

Πίνακας 3.3: Μεταβλητότητα που αναπαρίσταται από τις 10 μεγαλύτερες ιδιοτιμές.

3.3.3 Παράδειγμα: Μοντέλο εμφάνισης προσώπου

Η παραπάνω μέθοδος χρησιμοποιήθηκε για την κατασκευή ενός μοντέλου εμφάνισης. Το σύνολο εκπαίδευσης αποτελείται από 400 εικόνες του ανθρώπινου προσώπου. Στην καθεμία έχουν σημειωθεί 122 σημεία στα βασικά χαρακτηριστικά. Από αυτές κατασκευάστηκε ένα μοντέλο σχήματος 23^{ov} παραμέτρων, ένα μοντέλο υφής 114^{ov} παραμέτρων και ένα συνδυαστικό μοντέλο με μόνο 80 παραμέτρους το οποίο μπορεί να αναπαραστήσει το 98% της μεταβλητότητας.

Οι εικόνες 3.3 και 3.4 δείχνουν τα αποτελέσματα της μεταβολής των δύο πρώτων παραμέτρων του σχήματος και της υφής .



Εικόνα 3.4: Μεταβολή στο σχήμα.



Εικόνα 3.5: Μεταβολή στην υφή.

Η Εικόνα 3.5 δείχνει το αποτέλεσμα που προκύπτει από τη μεταβολή των τεσσάρων πρώτων παραμέτρων του μοντέλου εμφάνισης. Μπορούμε να παρατηρήσουμε αλλαγές στην ταυτότητα, τη θέση καθώς και στην έκφραση.



Εικόνα 3.6: Μεταβολή της εμφάνισης.

3.3.4 Προσεγγίζοντας ένα νέο παράδειγμα

Δοθείσας μίας νέας εικόνας, στην οποία έχουν σημειωθεί τα landmark points, μπορούμε να δημιουργήσουμε μία προσέγγιση χρησιμοποιώντας το μοντέλο. Ακολουθώντας τα βήματα που αναφέρθηκαν προηγουμένως κατασκευάζουμε το μοντέλο \mathbf{b} , συνδυάζοντας τις παραμέτρους σχήματος και υφής που ταιριάζουν με την εικόνα. Το \mathbf{P}_c είναι ορθογώνιο, άρα οι παράμετροι εμφάνισης \mathbf{c} δίνονται από την $\mathbf{c} = \mathbf{P}_c^T \mathbf{b}$. (3.12)

Για την επίτευξη του τελικού αποτελέσματος, χρησιμοποιούμε τις εξισώσεις 3.8, αντιστρέφουμε την κανονικοποίηση στη φωτεινότητα, εφαρμόζουμε την κατάλληλη θέση στα σημεία και εφαρμόζουμε το διάνυσμα της φωτεινότητας στην εικόνα.

Στην Εικόνα 3.6 φαίνεται μία προηγουμένως άγνωστη εικόνα (αριστερά) και η αναπαράσταση του προσώπου από το μοντέλο η οποία έχει τοποθετηθεί πάνω στην αρχική εικόνα(δεξιά).



Εικόνα 3.7: Αναπαράσταση προσώπου από συνδυαστικό μοντέλο εμφάνισης.

3.4 Μοντέλα Ενεργού Περιγράμματος(Active Appearance Models-AAM)

3.4.1 Αναζήτηση AAM

Αντιμετωπίζουμε την ερμηνεία μία εικόνας ως πρόβλημα βελτιστοποίησης κατά το οποίο ελαχιστοποιούμε τη διαφορά μεταξύ μια νέας εικόνας και μίας εικόνας που έχει δημιουργηθεί από το μοντέλο μας. Ορίζουμε ένα διάνυσμα διαφοράς:

$$\delta \mathbf{I} = \mathbf{I}_i - \mathbf{I}_m \quad (3.13)$$

όπου \mathbf{I}_i είναι το διάνυσμα της φωτεινότητας της νέας εικόνας και \mathbf{I}_m είναι το διάνυσμα της φωτεινότητας της εικόνας που προκύπτει από τις τρέχουσες παραμέτρους του μοντέλου.

Για να εντοπίσουμε το καλύτερο ταίριασμα μεταξύ μοντέλου και εικόνας, επιθυμούμε να ελαχιστοποιήσουμε το μέτρο του διανύσματος διαφοράς, $\Delta = |\delta \mathbf{I}|^2$, μεταβάλλοντας τις παραμέτρους του μοντέλου, \mathbf{c} . Αρχικά κάτι τέτοιο φαίνεται να είναι ένα δύσκολο πρόβλημα βελτιστοποίησης πολλών διαστάσεων, λόγω του αριθμού των παραμέτρων που μπορεί να έχει ένα μοντέλο. Παρατηρούμε όμως ότι κάθε προσπάθεια να ταιριάζει το μοντέλο σε μία νέα εικόνα είναι στην πραγματικότητα ένα παρόμοιο πρόβλημα βελτιστοποίησης. Παρέχοντας α-ρριογι γνώση σχετικά με το πώς να προσαρμόσουμε τις παραμέτρους του μοντέλου κατά τη διάρκεια της αναζήτησης, καταλήγουμε σε έναν αποδοτικό αλγόριθμο. Ειδικότερα η χωρική διασπορά του $\delta \mathbf{I}$ παρέχει πληροφορία για το πώς οι παράμετροι του μοντέλου πρέπει να μεταβληθούν προκειμένου να πετύχουμε καλύτερο ταίριασμα.

3.4.2 Διόρθωση των παραμέτρων του μοντέλου

Όπως αναφέρθηκε και παραπάνω, το μοντέλο έχει παραμέτρους \mathbf{c} , οι οποίες ρυθμίζουν το σχήμα και την υφή σύμφωνα με τις εξισώσεις 3.10. Ένα σχήμα στην εικόνα, \mathbf{X} , μπορεί να παραχθεί εφαρμόζοντας κατάλληλο μετασχηματισμό στα σημεία \mathbf{x} : $\mathbf{X} = \mathbf{S}_t(\mathbf{x})$. Συνήθως \mathbf{S}_t είναι ο μετασχηματισμός ομοιότητας, ο οποίος περιγράφεται από ένα συντελεστή κλίμακας, s , μία περιστροφή, θ , και μία μετατόπιση (t_x, t_y) . Για λόγους γραμμικότητας αναπαριστούμε την κλίμακα και την περιστροφή ως (s_x, s_y) όπου $s_x = (s \cos \theta - 1)$, $s_y = s \sin \theta$. Τότε το διάνυσμα $\mathbf{t} = (s_x, s_y, t_x, t_y)^T$ ισούται με μηδέν για τον ταυτοτικό μετασχηματισμό και $\mathbf{S}_{t+\delta t}(\mathbf{x}) \approx \mathbf{S}_t(\mathbf{S}_{\delta t}(\mathbf{x}))$

Η υφή της εικόνας μπορεί να παραχθεί εφαρμόζοντας κλίμακα και ένα offset στη φωτεινότητα, $\mathbf{g}_{im} = \mathbf{T}_u(\mathbf{g}) = (u_1 + 1)\mathbf{g}_{im} + u_2 \mathbf{1}$ όπου \mathbf{u} είναι το διάνυσμα των παραμέτρων του μετασχηματισμού, ορισμένο έτσι ώστε $\mathbf{u} = \mathbf{0}$ για τον ταυτοτικό μετασχηματισμό και $\mathbf{T}_{u+\delta u}(\mathbf{g}) \approx \mathbf{T}_u(\mathbf{T}_{\delta u}(\mathbf{g}))$

Οι παράμετροι του μοντέλου εμφάνισης, \mathbf{c} , και οι παράμετροι του μετασχηματισμού, \mathbf{t} , ορίζουν τη θέση των σημείων του μοντέλου στην εικόνα, \mathbf{X} , που μας δίνουν το σχήμα του τμήματος της εικόνας που θα αναπαρίσταται από το μοντέλο. Κατά τη διάρκεια του ταιριάσματος, παίρνουμε δείγματα από τα pixel σε αυτή την περιοχή της εικόνας, \mathbf{g}_{im} , και τα προβάλλουμε στο μοντέλο υφής, $\mathbf{g}_s = \mathbf{T}^{-1}(\mathbf{g}_{im})$. Η υφή δίνεται από: $\mathbf{g}_m = \bar{\mathbf{g}} + \mathbf{Q}_g \mathbf{c}$. Η διαφορά μεταξύ του μοντέλου και της εικόνας είναι $\mathbf{r}(\mathbf{p}) = \mathbf{g}_s - \mathbf{g}_m$ (3.14) όπου \mathbf{p} είναι οι παράμετροι του μοντέλου, $\mathbf{p}^T = (\mathbf{c}^T | \mathbf{t}^T | \mathbf{u}^T)$.

Ένα απλό βαθμωτό μέτρο της διαφοράς είναι το άθροισμα των τετραγώνων των στοιχείων του \mathbf{r} , $E(\mathbf{p}) = \mathbf{r}^T \mathbf{r}$.

Το πολυώνυμο Taylor πρώτης τάξης μας δίνει

$$\mathbf{r}(\mathbf{p} + \delta \mathbf{p}) = \mathbf{r}(\mathbf{p}) + \frac{\partial \mathbf{r}}{\partial \mathbf{p}} \delta \mathbf{p} \quad (3.15)$$

όπου το ij στοιχείο του πίνακα $\frac{\partial r}{\partial p}$ είναι $\frac{\partial r_i}{\partial p_j}$

Έστω ότι κατά το ταίριασμα το υπόλοιπό μας είναι \mathbf{r} . Θέλουμε να διαλέξουμε το $\delta \mathbf{p}$ με τέτοιο τρόπο ώστε να ελαχιστοποιηθεί το $|\mathbf{r}(\mathbf{p}+\delta \mathbf{p})|^2$. Εξισώνοντας την (3.15) με το 0 προκύπτει ότι $\delta \mathbf{p} = -\mathbf{R}\mathbf{r}(\mathbf{p})$ όπου $\mathbf{R} = \left(\frac{\partial \mathbf{r}}{\partial \mathbf{p}} \top \frac{\partial \mathbf{r}}{\partial \mathbf{p}}\right)^{-1} \frac{\partial \mathbf{r}}{\partial \mathbf{p}} \top$. (3.16)

Για να μην υπολογίζουμε το $\frac{\partial \mathbf{r}}{\partial \mathbf{p}}$ σε κάθε βήμα, υποθέτουμε ότι μπορεί να θεωρηθεί σχεδόν σταθερό λόγω της κανονικοποίησης που είχε γίνει στα δεδομένα πριν υπολογιστεί. Εκτιμούμε την τιμή του $\frac{\partial \mathbf{r}}{\partial \mathbf{p}}$ με αριθμητική παραγωγή. Για τον υπολογισμό μετατοπίζουμε μία-μία τις παραμέτρους από τη γνωστή βέλτιστη τιμή σε τυπικές εικόνες και υπολογίζουμε τον μέσο όρο στο σύνολο εκπαίδευσης. Μετράμε τα υπόλοιπα στις μετατοπίσεις και χρησιμοποιούμε ένα Γκαουσιανό πυρήνα.

$$\frac{dr_i}{dp_j} = \sum_k w(\delta c_{jk})(r_i(\mathbf{p} + \delta c_{jk}) - r_i(\mathbf{p})) \quad (3.17)$$

όπου $w(x)$ είναι μία κατάλληλα κανονικοποιημένη γκαουσιανή συνάρτηση βάρους.

Στη συνέχεια προ-υπολογίζουμε το \mathbf{R} και το χρησιμοποιούμε σε όλες τις επόμενες αναζητήσεις με το μοντέλο.

Οι εικόνες που χρησιμοποιούμε για τον υπολογισμό του $\frac{\partial \mathbf{r}}{\partial \mathbf{p}}$ μπορούν να είναι εικόνες από το σύνολο εκπαίδευσης ή συνθετικές εικόνες που παράγουμε από το ίδιο το μοντέλο.

Οι καλύτερες τιμές για τα δc , δt και δu που θα χρησιμοποιηθούν κατά την εκπαίδευση, προσδιορίζονται πειραματικά.

3.4.3 Διόρθωση των παραμέτρων του μοντέλου επαναληπτικά

Τώρα που έχουμε μία μέθοδο για την πρόβλεψη των διορθώσεων που απαιτούνται στις παραμέτρους του μοντέλου, μπορούμε να κατασκευάσουμε μία επαναληπτική μέθοδο η οποία θα λύσει αυτό το πρόβλημα βελτιστοποίησης.

Έχοντας την τρέχουσα πρόβλεψη των παραμέτρων του μοντέλου, \mathbf{c}_0 , και το κανονικοποιημένο δείγμα των φωτεινοτήτων που αντιστοιχεί στην εκτίμηση αυτή, \mathbf{g}_s , ακολουθούμε τα εξής βήματα:

- Υπολογισμός του διανύσματος σφάλματος $\delta \mathbf{g}_0 = \mathbf{g}_s - \mathbf{g}_m$.
- Υπολογισμός του τρέχοντος σφάλματος $E_0 = |\delta \mathbf{g}_0|^2$.
- Υπολογισμός της προβλεπόμενης μετατόπισης, $\delta \mathbf{c} = \mathbf{A} \delta \mathbf{g}_0$.
- Θέτουμε $k=1$.
- Έστω $\mathbf{c}_1 = \mathbf{c}_0 - k \delta \mathbf{c}$.
- Δειγματοληπτούμε την εικόνα σύμφωνα με τη νέα πρόβλεψη και υπολογίζουμε το νέο διάνυσμα σφάλματος, $\delta \mathbf{g}_1$.
- Εάν $|\delta \mathbf{g}_1|^2 < E_0$, δεχόμαστε τη νέα πρόβλεψη, \mathbf{c}_1 .
- Διαφορετικά δοκιμάζουμε για $k=1.5$, $k=0.5$, $k=0.25$ κλπ.

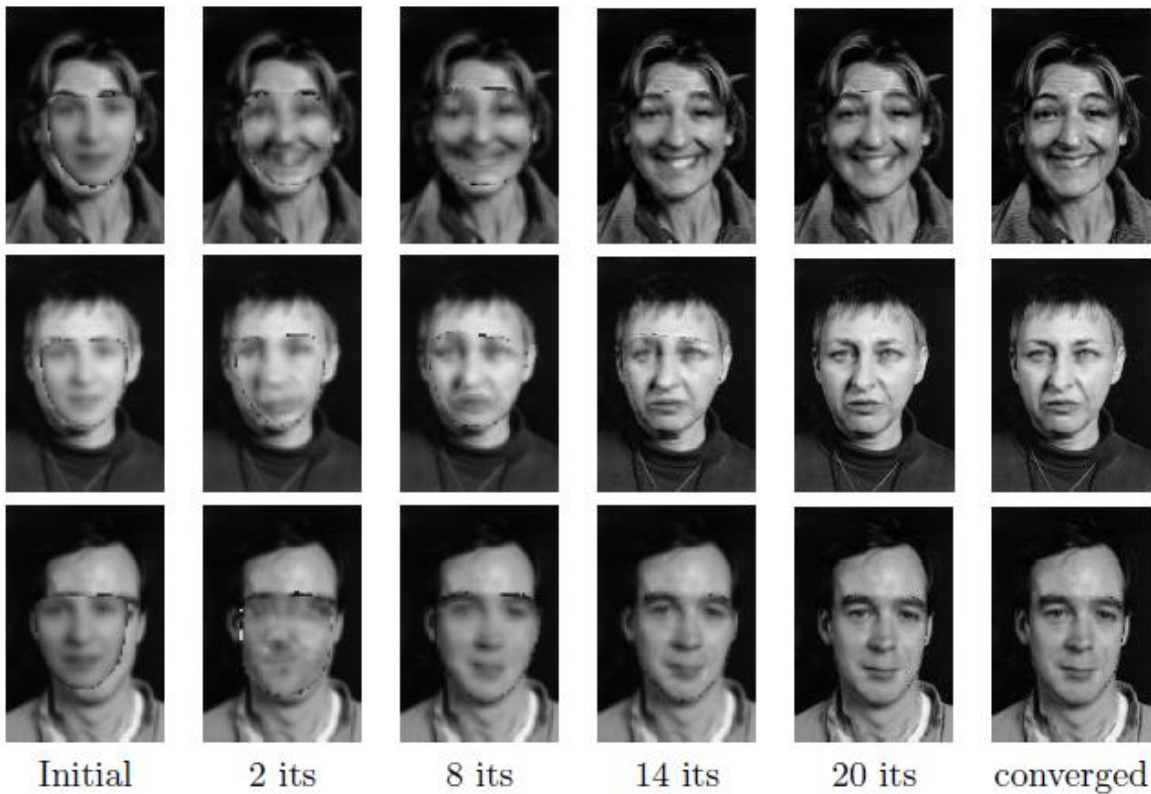
Αυτή η διαδικασία επαναλαμβάνεται μέχρις ότου να μην παρατηρείται καμία βελτίωση στο σφάλμα $|\delta \mathbf{g}|^2$.

3.4.4 Παραδείγματα της αναζήτησης των μοντέλων ενεργού περιγράμματος σε εικόνες προσώπων.

Χρησιμοποιούμε τα μοντέλα ενεργού περιγράμματος για την αναζήτηση προσώπων σε προηγουμένως άγνωστες εικόνες. Στις παρακάτω εικόνες φαίνεται αρχικά το καλύτερο ταίριασμα του μοντέλου για τρία πρόσωπα και στη συνέχεια φαίνονται κάποια από τα στάδια της αναζήτησης για το κάθε πρόσωπο. Η αναζήτηση ξεκινάει μετατοπίζοντας το μέσο σχήμα από τη θέση του.



Εικόνα 3.8: Καλύτερο ταίριασμα του μοντέλου.



Εικόνα 3.9: Απεικόνιση σταδίων της αναζήτησης.

3.4.5 Υλοποίηση

Όλα τα πειράματα διεξήχθησαν χρησιμοποιώντας μία υλοποίηση ανοιχτού κώδικα των Μοντέλων Ενεργού Περιγράμματος. Πρόκειται για το AAM-API, μία υλοποίηση σε C++ για τα Windows.

4. Ανάπτυξη Συστήματος Παρακολούθησης των Χαρακτηριστικών του Προσώπου.

Στα πλαίσια της εργασίας αυτής αναπτύχθηκε ένα σύστημα αυτόματης αναγνώρισης των εκφράσεων του προσώπου. Παρακολουθώντας τα χαρακτηριστικά του προσώπου και τη μεταβολή στη θέση και το σχήμα τους μπορούμε να αποφανθούμε για τις μεταβολές στην έκφραση του προσώπου. Για το σκοπό αυτό χρησιμοποιούμε τα Μοντέλα Ενεργού Περιγράμματος, για τον εντοπισμό του προσώπου και των χαρακτηριστικών του σε καρέ βίντεο.

4.1 Υλοποίηση

Η υλοποίηση έγινε στο Matlab. Χρησιμοποιήθηκαν κάποια από τα ήδη υπάρχοντα script εξ' ολοκλήρου ή τροποποιημένα, και δημιουργήθηκαν και νέα. Για την αναζήτηση AAM χρησιμοποιήθηκε το `aamcm.exe`. Ανάλογα με τις παραμέτρους με τις οποίες θα κληθεί μπορεί εκτός των άλλων να δημιουργήσει ένα μοντέλο από το data set που θα του δοθεί, ή να κάνει αναζήτηση στις εικόνες του φακέλου που θα του δοθεί. Παρακάτω γίνεται αναλυτική περιγραφή των scripts και των συναρτήσεων που χρησιμοποιήθηκαν και αναπτύχθηκαν.

tracking.m

Ξεκινάμε με την εκτέλεση του script *tracking.m*. Αρχικά το σύστημα ζητάει από το χρήστη να επιλέξει εάν θέλει να πραγματοποιήσει νέα αναζήτηση ή να χρησιμοποιήσει ήδη υπάρχοντα αποτελέσματα.

- Νέα Αναζήτηση: Σαν είσοδο το σύστημα δέχεται ένα αρχείο βίντεο. Το αρχείο αυτό μπορεί να είναι οποιασδήποτε μορφής και ανάλυσης. Στη συνέχεια καλούνται οι συναρτήσεις `frames_bmp.m` και `aam.m` οι οποίες πραγματοποιούν την εξαγωγή των frames του βίντεο και την αναζήτηση AAM αντίστοιχα.
- Ήδη υπάρχοντα αποτελέσματα: Ο χρήστης επιλέγει τον φάκελο που περιέχει τα αποτελέσματα τα οποία θέλει να χρησιμοποιήσει, τα οποία δημιουργήθηκαν από προηγούμενη αναζήτηση.

Έπειτα ο χρήστης καλείται να επιλέξει ποια frames θέλει να συγκρίνει για μεταβολές στην έκφραση. Αυτά είναι τα reference και test frames. Συνήθως σαν reference frame επιλέγεται το πρώτο frame (1.bmp) το οποίο περιέχει την ουδέτερη έκφραση. Αν όμως αυτό δεν ισχύει ή εάν ο χρήστης το επιθυμεί, μπορεί να επιλέξει κάποιο άλλο frame ως reference.

Μέσω του script αυτού εμφανίζονται τα δύο frames που επιλέχθηκαν στο προηγούμενο βήμα με σχεδιασμένο πάνω τους το σχήμα που προκύπτει από το μοντέλο όταν γίνει το καλύτερο δυνατό ταίριασμα.

Στη συνέχεια καλείται η συνάρτηση *features.m* η οποία ελέγχει τις μεταβολές στα χαρακτηριστικά του προσώπου.

Τέλος ο χρήστης μπορεί να επιλέξει για να συγκρίνει 2 νέα frames, ή να τερματίσει το πρόγραμμα.

frames_bmp.m

Αυτή η συνάρτηση παίρνει το όνομα του αρχείου βίντεο, δημιουργεί ένα νέο φάκελο και σε αυτόν αποθηκεύει τα frames σε μορφή bmp. Το όνομα του κάθε frame αντιστοιχεί στον αύξοντα αριθμό του μέσα στο βίντεο (1.bmp, 2.bmp κλπ). Σαν αποτέλεσμα επιστρέφει το όνομα του φακέλου (το ίδιο με το όνομα του βίντεο χωρίς το path και την επέκταση) .

aam.m

Η συνάρτηση αυτή δέχεται σαν είσοδο το directory στο οποίο θα πραγματοποιηθεί η αναζήτηση.

Αρχικά, εάν δεν έχει δημιουργηθεί προηγουμένως, δημιουργείται το μοντέλο από το σύνολο εκπαίδευσης. Αυτό πραγματοποιείται με την εντολή :

```
aamcm b <input dir> <out model> [acf file]
```

όπου <input dir> είναι το path του φακέλου που περιέχει το σύνολο εκπαίδευσης, <out model> είναι το path του μοντέλου που θα δημιουργηθεί και [acf file] το path για το αρχείο acf (Παράρτημα Α) .Εάν το μοντέλο έχει δημιουργηθεί σε προηγούμενη εκτέλεση, και εφόσον δεν υπάρχουν αλλαγές στο σύνολο εκπαίδευσης, η δημιουργία του μοντέλου εκ νέου είναι περιττή. Για λόγους συντομίας επομένως μπορούμε να σχολιάσουμε τη συγκεκριμένη γραμμή στον κώδικα.

Τέλος πραγματοποιείται η αναζήτηση με την εντολή:

```
aamcm s <input model.amf> <input dir> [movie filename]
```

όπου <input model.amf> είναι το path του μοντέλου, <input dir> είναι ο φάκελος που περιέχει τις εικόνες στις οποίες θα πραγματοποιηθεί η αναζήτηση και [movie filename] είναι το όνομα του βίντεο που δημιουργείται και δείχνει τα στάδια της αναζήτησης ([] για να μη δημιουργηθεί βίντεο).

Στις εικόνες που βρίσκονται στο input dir γίνεται αναζήτηση AAM και για κάθε εικόνα παράγεται ένα αρχείο .asf το οποίο περιέχει τις θέσεις των σημείων του μοντέλου (Παράρτημα Β) και αποθηκεύεται και αυτό στο input dir. Επειδή η διαδικασία της αναζήτησης είναι χρονοβόρα, ειδικά για μεγάλα βίντεο, τα αποτελέσματα της αναζήτησης μπορούν να χρησιμοποιηθούν απευθείας από το φάκελο στον οποίο αποθηκεύθηκαν χωρίς να προηγηθεί ξανά αναζήτηση.

Όπως φαίνεται παραπάνω το aamcm πραγματοποιεί διαφορετική λειτουργία ανάλογα με τις παραμέτρους που δέχεται. Στην περίπτωση μας, με b δημιουργείται το μοντέλο και με s πραγματοποιείται η αναζήτηση.

features.m

Η συνάρτηση *features.m* δέχεται σαν ορίσματα τα ονόματα των αρχείων asf που αντιστοιχούν στα reference και test frames που επέλεξε ο χρήστης. Σε ένα συνδυαστικό διάγραμμα σχεδιάζεται το σχήμα που προκύπτει από τα σημεία του μοντέλου μετά το τελευταίο βήμα της αναζήτησης. Με μπλε σχεδιάζεται το σχήμα που αντιστοιχεί στο reference frame και με κόκκινο αυτό που αντιστοιχεί στο test frame. Έτσι μπορούν να εκτιμηθούν ως ένα βαθμό, από απλή παρατήρηση του διαγράμματος, οι αλλαγές που έχουν γίνει. Η συνάρτηση *features.m* καλεί για κάθε χαρακτηριστικό του προσώπου μία συνάρτηση (στην περίπτωση των φρυδιών, δύο) οι οποίες παρακολουθούν τις αλλαγές που έχουν γίνει. Αυτές οι συναρτήσεις περιγράφονται παρακάτω.

nose.m

Η συνάρτηση αυτή είναι υπεύθυνη για τις κινήσεις που πραγματοποιεί η μύτη. Η μύτη είναι το χαρακτηριστικό που κινείται λιγότερο απ' όλα στο πρόσωπο. Μπορούμε να υποθέσουμε ότι η μύτη στην ουσία δεν κινείται ανεξάρτητα, παρά μόνο αν κινηθεί ολόκληρο το κεφάλι. Για το λόγο αυτό τη χρησιμοποιούμε για να προσδιορίσουμε την κίνηση του κεφαλιού συνολικά. Η μύτη περιγράφεται από το μοντέλο με 11 σημεία τα οποία σχηματίζουν ένα ανοιχτό μονοπάτι, όπως φαίνεται στην παρακάτω εικόνα.



Εικόνα 4.1: Σημεία που περιγράφουν τη μύτη.

Για περισσότερη ακρίβεια χρησιμοποιούμε τα 4 πάνω σημεία (48, 49, 57, 58 στην Εικόνα 4.1) για να προσδιορίσουμε την κίνηση του κεφαλιού μιας και αυτά τα σημεία είναι πρακτικά ακίνητα στο πρόσωπο.

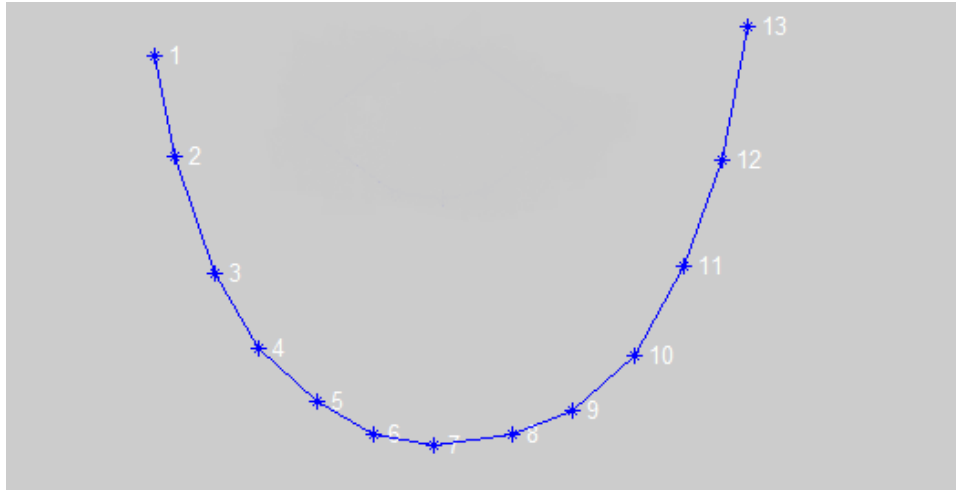
Η συνάρτηση δέχεται ως ορίσματα τα δύο αρχεία *asf*, από τα οποία υπολογίζουμε την κατά μέσο όρο συνολική μετατόπιση των τεσσάρων αυτών σημείων, η οποία και ισούται με τη μετατόπιση του κεφαλιού. Οι δύο μετατοπίσεις που υπολογίστηκαν, μία για τον οριζόντιο και μία για τον κατακόρυφο άξονα, επιστρέφονται στη συνάρτηση *features*.

head.m

Η συνάρτηση *head.m* καλείται από την *features.m* αμέσως μετά την *nose.m* και δέχεται σαν ορίσματα τις μετατοπίσεις που αυτή επέστρεψε. Η *head.m* εκτυπώνει στο command window την κίνηση που πραγματοποίησε το κεφάλι στον οριζόντιο και κάθετο άξονα. Μετά από παρατήρηση των αποτελεσμάτων επιλέχθηκαν με πειραματικό τρόπο κατώφλια για τις μετακινήσεις, ώστε να μην αναφέρεται μετατόπιση του κεφαλιού σε οποιαδήποτε κατεύθυνση εάν αυτή είναι τόσο μικρή που δεν είναι ορατή.

jaw.m

Στη συνέχεια καλείται από την *features.m* η *jaw.m* η οποία είναι υπεύθυνη για την μετακίνηση του σαγονιού. Το σαγόι δεν κάνει πολλές κινήσεις και κυρίως μας ενδιαφέρει η μετατόπιση του στον κάθετο άξονα, η οποία συνδέεται με το άνοιγμα και κλείσιμο του στόματος. Περιγράφεται από 13 σημεία τα οποία επίσης σχηματίζουν ένα ανοιχτό μονοπάτι όπως φαίνεται παρακάτω.



Εικόνα 4.2: Σημεία που περιγράφουν το σαγόι.

Η συνάρτηση αυτή δέχεται σαν ορίσματα τα δύο αρχεία *asf* και την κατακόρυφη μετακίνηση του κεφαλιού όπως αυτή προέκυψε από τη *pose.m*. Για να υπολογιστεί η πραγματική μετατόπιση του σαγονιού, πρέπει να λάβουμε υπ' όψη και την κίνηση που έκανε συνολικά το κεφάλι και να αφαιρέσουμε την επίδρασή της από την μετατόπιση του σαγονιού. Από τα *asf* των δύο frames που συγκρίνουμε μπορούμε και πάλι να υπολογίσουμε κατά μέσο όρο τη μετατόπιση του σαγονιού. Όλα τα σημεία του σαγονιού θεωρούμε ότι κινούνται ομοιόμορφα προς μια κατεύθυνση. Η συνάρτηση τέλος επιστρέφει στη *features.m* τη μετατόπιση του σαγονιού στον κάθετο άξονα.

eyes.m

Σειρά έχει η *eyes.m*. Η μόνη κίνηση που πραγματοποιούν τα μάτια είναι να ανοίγουν και να κλείνουν. Από το μοντέλο περιγράφονται με 8 σημεία το καθένα, σχηματίζοντας δύο κλειστά μονοπάτια όπως φαίνεται παρακάτω.



Εικόνα 4.3: Σημεία που περιγράφουν τα μάτια.

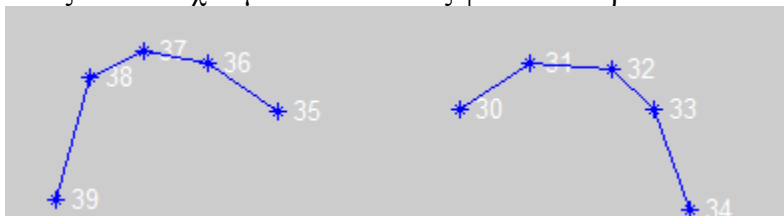
Γνωρίζοντας τη θέση των σημείων που περιγράφουν τα μάτια μπορούμε να αποφανθούμε εάν αυτά είναι πιο ανοιχτά ή πιο κλειστά στο test frame σε σχέση με το reference. Χρησιμοποιώντας την συνάρτηση *polyfit* του Matlab υπολογίζουμε την ευθεία που περιγράφει καλύτερα τα σημεία του κάθε ματιού (ελάχιστα τετράγωνα). Η ευθεία θα περνάει περίπου από τη μέση του ματιού αφού οι δύο γωνίες (σημεία 22,27 για το δεξί και 18,14 για το αριστερό) παραμένουν σχετικά σταθερές και τα σημεία στο πάνω και κάτω μέρος είναι ομοιόμορφα καταναμημένα. Στη συνέχεια με τη συνάρτηση *polyval* υπολογίζουμε την ακρίβεια με την οποία περιγράφει η ευθεία αυτή τα σημεία. Όσο μεγαλύτερο είναι το σφάλμα τόσο πιο ανακριβές είναι το ταίριασμα των σημείων στην ευθεία, άρα και πιο ανοιχτά τα μάτια. Συγκρίνοντας για κάθε μάτι το σφάλμα του reference και του test frame μπορούμε να καταλάβουμε αν το μάτι έχει ανοίξει ή έχει κλείσει.

Η συνάρτηση *eyes.m* παίρνει σαν όρισμα τα δύο αρχεία *asf*, κάνει τους παραπάνω υπολογισμούς και επιστρέφει δύο τιμές τις *lc* και *rc* οι οποίες είναι 1 εάν το αριστερό, και

αντίστοιχα, το δεξί μάτι είναι πιο κλειστό στο test frame απ' ότι στο reference και 0 αν είναι πιο ανοιχτό.

eyebrows.m

Για τα φρύδια ενδιαφέρον παρουσιάζει η μετατόπισή τους στον κάθετο άξονα καθώς και στον οριζόντιο, όταν αυτά πλησιάζουν μεταξύ τους. Τα φρύδια περιγράφονται από 5 σημεία το καθένα, σχηματίζοντας δύο ανοιχτά μονοπάτια όπως φαίνεται παρακάτω.

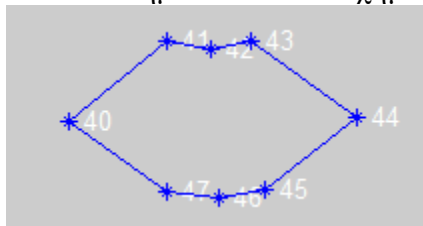


Εικόνα 4.4: Σημεία που περιγράφουν τα φρύδια.

Η συνάρτηση *eyebrows.m* παίρνει ως ορίσματα τα δύο αρχεία *asf* και την συνολική μετατόπιση του κεφαλιού στους δύο άξονες και υπολογίζει τη συνολική μετατόπιση των φρυδιών. Και πάλι θεωρούμε ότι όλα τα σημεία των φρυδιών κινούνται ομοίμορφα προς κάθε κατεύθυνση και, όπως και για το σαγόι, αφαιρούμε από τη συνολική μετατόπιση, τη μετατόπιση που οφείλεται στην κίνηση του κεφαλιού. Στην *features.m* επιστρέφει τη μετατόπιση και των δύο φρυδιών στους δύο άξονες (συνολικά τέσσερις τιμές).

mouth.m

Το στόμα μπορεί να ανοίξει, να κλείσει και να χαμογελάσει. Επίσης υπάρχουν διάφορες άλλες κινήσεις, όπως το βγάλσιμο της γλώσσας και διάφορες γκριμάτσες οι οποίες μπορούν να γίνουν με το στόμα. Οι γκριμάτσες αυτές έχουν σαν αποτέλεσμα την παραμόρφωση του στόματος σε τέτοιο βαθμό, ώστε αυτό να μην εντοπίζεται σωστά στην αναζήτηση των AAM. Αυτό το πρόβλημα υπάρχει ακόμα και στις περιπτώσεις που το στόμα ανοίγει πολύ. Επομένως λαμβάνουμε υπ' όψη μας μόνο τις κινήσεις εκείνες οι οποίες δεν οδηγούν σε αποτυχημένη αναζήτηση. Το στόμα αποτελείται από 8 σημεία τα οποία σχηματίζουν ένα κλειστό μονοπάτι.



Εικόνα 4.5: Σημεία που περιγράφουν το στόμα.

Λόγω της μορφής που έχει το στόμα, μπορούμε να το αντιμετωπίσουμε όπως και τα μάτια για να αποφανθούμε αν είναι ανοιχτό ή κλειστό. Η συνάρτηση *mouth.m* δέχεται για ορίσματα τα δύο αρχεία *asf*, τα *lc* και *rc* που επέστρεψε στο προηγούμενο βήμα η *eyes.m*, τις μετατοπίσεις του σαγονιού και του κεφαλιού στον κάθετο άξονα. Με τις *polyfit* και *polyval* υπολογίζουμε το σφάλμα όπως ακριβώς κάναμε και για τα μάτια. Αρχικά ελέγχουμε αν υπάρχει χαμόγελο. Όταν κάποιος χαμογελάει συμβαίνουν τα παρακάτω:

- Οι γωνίες του στόματος (σημεία 40 και 44) είναι ανασηκωμένες
- Τα μάτια είναι πιο κλειστά απ' ότι στην ουδέτερη έκφραση και
- Το στόμα είναι μετατοπισμένο προς τα πάνω.

Ελέγχοντας τις παραπάνω συνθήκες εντοπίζουμε την ύπαρξη ή μη χαμόγελου.

Η άλλη επιλογή είναι το στόμα να είναι πολύ ανοιχτό(πχ έκπληξη). Αυτό συμβαίνει όταν:

- Το σφάλμα που προκύπτει από τη polygon για το test frame είναι μεγαλύτερο από αυτό του reference frame και μεγαλύτερο από ένα κατώφλι το οποίο προέκυψε πειραματικά.
- Το σαγόνι είναι μετατοπισμένο προς τα κάτω.

Στη συνέχεια ελέγχουμε αν το σφάλμα για το reference frame είναι διαφορετικό από αυτό για το test frame και μικρότερο από ένα κατώφλι το οποίο επίσης προέκυψε πειραματικά. Σε αυτή την περίπτωση τα χείλη είναι πιεσμένα μεταξύ τους.

Τέλος συγκρίνοντας και πάλι τα σφάλματα των reference και test frames ελέγχουμε αν το στόμα είναι πιο ανοιχτό ή πιο κλειστό λαμβάνοντας υπ' όψη μας και το ενδεχόμενο να μην έχει γίνει καμία κίνηση.

Η συνάρτηση επιστρέφει μία τιμή sm η οποία είναι ίση με 1 εάν έχει εντοπιστεί χαμόγελο διαφορετικά είναι 0.

eyebrows_movement.m

Τέλος καλείται η *eyebrows_movement.m* η οποία ελέγχει την κίνηση των φρυδιών. Τα ορίσματα που παίρνει είναι: οι μετατοπίσεις, οριζόντιες και κάθετες, των φρυδιών, το sm που υπολογίστηκε στο προηγούμενο βήμα και τα lc, rc που αφορούν τα μάτια.

Ο πρώτος έλεγχος που γίνεται είναι εάν το άτομο του οποίου το πρόσωπο παρατηρούμε είναι συνοφρυωμένο. Για να ισχύει κάτι τέτοιο πρέπει να ισχύει μία από τις παρακάτω συνθήκες:

1. Τα φρύδια να έχουν μετατοπιστεί προς τα κάτω, τα μάτια να έχουν κλείσει και να μην υπάρχει χαμόγελο.
2. Τα φρύδια να έχουν μετατοπιστεί προς τα κάτω, να έχουν πλησιάσει μεταξύ τους στον οριζόντιο άξονα και να μην υπάρχει χαμόγελο.

Τέλος αν τα παραπάνω δεν ισχύουν ελέγχεται η μετατόπιση των φρυδιών μεμονωμένα και βγάζουμε ξεχωριστά συμπεράσματα για κάθε φρύδι.

search.m

Η τελευταία συνάρτηση που δημιουργήθηκε είναι η *search.m*. Αυτή είναι μία βοηθητική συνάρτηση που χρησιμοποιείται ανεξάρτητα όποτε χρειάζεται. Στην περίπτωση που αντί για βίντεο έχουμε ένα φάκελο με εικόνες στις οποίες θέλουμε να πραγματοποιηθεί η AAM αναζήτηση, καλούμε την *search.m* με όρισμα το όνομα του φακέλου. Εκείνη μετατρέπει τις εικόνες σε μορφή BMP και στη συνέχεια καλώντας τη συνάρτηση *aam.m*, που αναφέρθηκε παραπάνω, γίνεται η αναζήτηση. Στο τέλος διαγράφονται όλα τα περιττά αρχεία και ο φάκελος είναι στη μορφή που χρειάζεται ώστε να μπορέσουν τα αποτελέσματα αυτά της αναζήτησης να χρησιμοποιηθούν από το *tracking.m*.

Χρησιμοποιήθηκαν επίσης δυο από τις ήδη υπάρχουσες συναρτήσεις, η *readasf.m* και η *getasfpath.m*.

readasf.m

Η *readasf.m* δέχεται σαν όρισμα ένα αρχείο asf και από αυτό δημιουργεί τον point matrix, P, ο οποίος περιέχει όλα τα σημεία του μοντέλου. Η μορφή του πίνακα περιγράφεται στο Παράρτημα Β.

getasfpath.m

Η συνάρτηση αυτή δέχεται σαν όρισμα τον αριθμό ενός path και τον point matrix P και επιστρέφει έναν υποπίνακα του P ο οποίος περιέχει τα σημεία του path που δόθηκε.

Με τις δύο αυτές συναρτήσεις απομονώνουμε τα σημεία που περιγράφουν τα επιμέρους χαρακτηριστικά για να μελετηθεί η κίνησή τους.

4.2 Πειραματικά αποτελέσματα

Στη συνέχεια διεξήχθησαν πειράματα, προκειμένου να υπολογιστεί η ακρίβεια με την οποία το σύστημα που αναπτύχθηκε, αναγνωρίζει τις εκφράσεις του προσώπου. Για το σκοπό αυτό χρησιμοποιήθηκαν εικόνες από τη βάση MMIdb ([6]) και από τη βάση των Cohn-Kanade ([7],[8]). Η MMIdb περιέχει συνολικά 455 έγχρωμες εικόνες από 4 διαφορετικά άτομα. Στις εικόνες αυτές απεικονίζονται οι διαφορετικές Action Units. Στη βάση των Cohn – Kanade απεικονίζονται σε ασπρόμαυρες εικόνες οι βασικές εκφράσεις από 97 διαφορετικά άτομα.

Από τα πειράματα αυτά προέκυψε το συμπέρασμα, ότι μεγαλύτερη δυσκολία υπάρχει όταν το άτομο του οποίου το πρόσωπο μελετάται είναι συνοφρυωμένο ή πιέζει τα χείλη του. Σε αυτές τις δύο περιπτώσεις έγιναν τα περισσότερα λάθη στην αναγνώριση, καθώς το σύστημα απέτυχε κάποιες φορές να αναγνωρίσει τις εκφράσεις αυτές. Στις περισσότερες περιπτώσεις, το σύστημα τις μπερδεύει με το χαμόγελο. Αυτό οφείλεται στο γεγονός ότι και στις τρεις αυτές εκφράσεις υπάρχει κίνηση στο στόμα, η οποία πολλές φορές είναι παρόμοια, ανεξάρτητα από το αν το άτομο χαμογελάει, είναι συνοφρυωμένο ή πιέζει τα χείλη του.

Όσον αφορά το άνοιγμα και κλείσιμο του στόματος, τα αποτελέσματα ήταν αρκετά ακριβή μιας και βασίζονται στη ακριβή θέση των σημείων που περιγράφουν το στόμα. Αυτό σημαίνει ότι εάν η αναζήτηση AAM είναι ακριβής, τότε και η εκτίμηση για το αν το στόμα είναι ανοιχτό η κλειστό, θα είναι επίσης ακριβής. Τα λίγα λάθη που προέκυψαν, οφείλονταν στο ότι το σύστημα μπερδευε το χαμόγελο με το ανοιχτό στόμα.

Για τις κινήσεις που αφορούν τα φρύδια και τα μάτια, η ακρίβεια άγγιζε το 100%. Όσον αφορά τα φρύδια, αυτό οφείλεται στο γεγονός ότι η στην κίνηση που πραγματοποιείται συμμετέχουν όλα τα σημεία του φρυδιού με ομοιόμορφο τρόπο. Έτσι έχοντας τη μεταβολή στη θέση των σημείων, από την αναζήτηση AAM, εύκολα μπορούμε να υπολογίσουμε προς τα πού κινήθηκαν τα φρύδια. Για τα μάτια, οι κινήσεις που πραγματοποιούνται είναι σχετικά απλές. Έτσι εάν η αναζήτηση AAM μας δώσει ακριβή αποτελέσματα, ο υπολογισμός του σφάλματος από την polyval αποτελεί αξιόπιστο δείκτη για το αν τα μάτια είναι πιο ανοιχτά ή πιο κλειστά σε σχέση με το reference frame.

Τέλος χωρίς σφάλματα υπολογίζεται και η συνολική κίνηση του κεφαλιού από τη μετατόπιση των τεσσάρων σημείων στη μύτη.

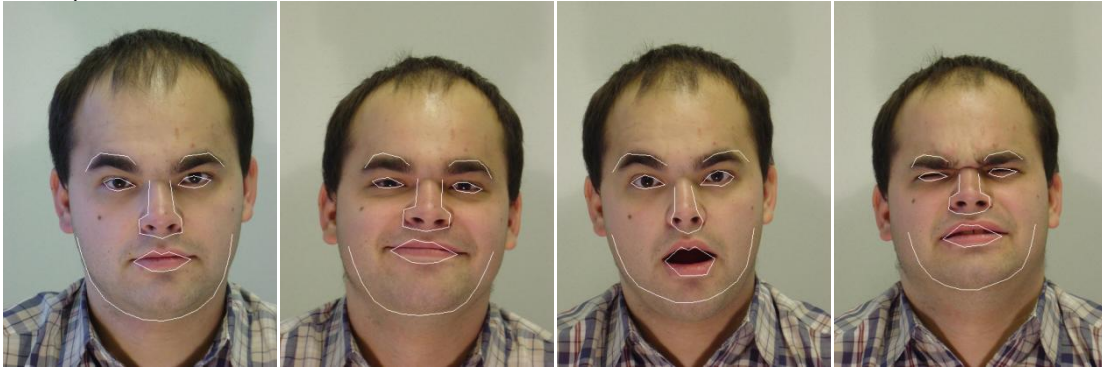
Στον παρακάτω πίνακα φαίνονται συγκεντρωτικά τα αποτελέσματα των πειραμάτων, όσον αφορά την ακρίβεια στην αναγνώριση των εκφράσεων.

Έκφραση	Ακρίβεια αναγνώρισης
Χαμόγελο	98.4%
Συνοφρύωμα	70.3%
Ανοιχτό/κλειστό στόμα	80.6%
Ανοιχτά/κλειστά μάτια	100%
Πιεσμένα χείλη	78.8%
Κίνηση στα φρύδια	96.7%
Κίνηση στο κεφάλι	100%

Πίνακας 4.1: Ακρίβεια αναγνώρισης εκφράσεων.

Στη συνέχεια φαίνονται κάποια ενδεικτικά παραδείγματα εκτέλεσης.

Αρχικά βλέπουμε κάποια από τα παραδείγματα για τα οποία το σύστημα έδωσε σωστά αποτελέσματα:



Εικόνα 4.6: Από αριστερά προς τα δεξιά: Ουδέτερη έκφραση, Χαμόγελο, Ανοιχτό στόμα και μάτια και ανασηκωμένα φρύδια, Συνοφρυωμένος.



Εικόνα 4.7: Από αριστερά προς τα δεξιά: Ουδέτερη έκφραση, Κλειστά μάτια, Συνοφρυωμένος, Χαμόγελο.



Εικόνα 4.8: Από αριστερά προς τα δεξιά: Ουδέτερη έκφραση, Ανοιχτό στόμα, Πιεσμένα χείλη.



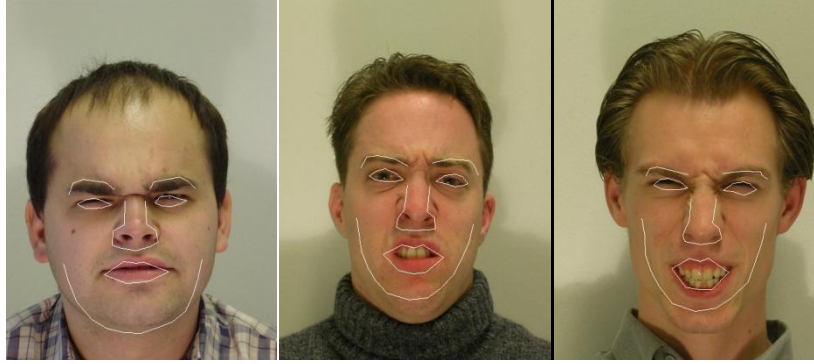
Εικόνα 4.9: Από αριστερά προς τα δεξιά: Ουδέτερη έκφραση, Ανοιχτό στόμα και μάτια και ανασηκωμένα φρύδια, Κλειστό στόμα, Συνοφρυωμένος.



Εικόνα 4.10: Από αριστερά προς τα δεξιά: Ουδέτερη έκφραση, Συνοφρυωμένος και ανοιχτό στόμα, Χαμόγελο, Χαμηλωμένα φρύδια.

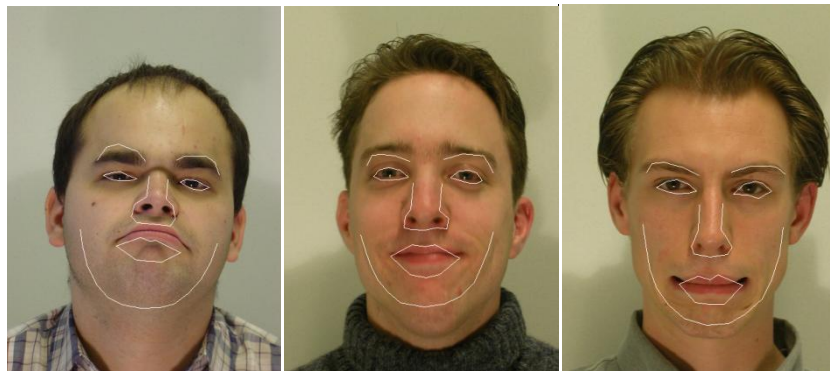
Τέλος ακολουθούν κάποια παραδείγματα, για τα οποία το σύστημα απέτυχε να αναγνωρίσει σωστά την έκφραση.

Στην Εικόνα 4.11 βλέπουμε κάποια παραδείγματα τα οποία αναγνωρίστηκαν ως χαμόγελα χωρίς να είναι. Αυτό οφείλεται στο ότι η κίνηση που έγινε στο στόμα, μοιάζει με αυτή του χαμόγελου και τα μάτια είναι κλειστά, κάτι που συμβαίνει και όταν κάποιος χαμογελάει.



Εικόνα 4.11: Λανθασμένη αναγνώριση ως χαμόγελο.

Στην Εικόνα 4.12 παρουσιάζονται τρία παραδείγματα για τα οποία το αποτέλεσμα της αναγνώρισης ήταν, λανθασμένα, ανοιχτό στόμα. Αυτό οφείλεται στην αποτυχία της αναζήτησης AAM στη περιοχή γύρω από το στόμα, όπως φαίνεται και από την εικόνα.



Εικόνα 4.12: Λανθασμένη αναγνώριση ως ανοιχτό στόμα.

5. Προκλήσεις και Μελλοντική Εργασία

Όσον αφορά την αναγνώριση εκφράσεων η προσοχή των ερευνητών στρέφεται στην αναγνώριση αυθόρμητων εκφράσεων. Η μεγαλύτερη πρόκληση είναι η έλλειψη επαρκών δεδομένων, εικόνων και βίντεο, στα οποία να καταγράφονται αυθόρμητες εκφράσεις. Αυτό οφείλεται στο ότι εάν κάποιος αντιληφθεί ότι οι αντιδράσεις του καταγράφονται, η έκφρασή του αυτόματα χάνει την αυθεντικότητά της. Αυτό μπορεί να αντιμετωπιστεί με χρήση κρυφής κάμερας η οποία να καταγράφει τις εκφράσεις κάποιων ατόμων και στη συνέχεια να ζητείται η συναίνεσή τους για τη χρήση του υλικού που καταγράφη. Σε αυτή την περίπτωση παραμένει το πρόβλημα ότι για τους σκοπούς της έρευνας απαιτούνται εικόνες με αλλαγές στο φωτισμό, και εικόνες που να μην φαίνεται ολόκληρο το πρόσωπο (πχ με τη χρήση γυαλιών, μαντηλιών κλπ).

Ο πιο συνηθισμένος τρόπος με τον οποίο προκαλούνται συναισθήματα στα άτομα των οποίων οι εκφράσεις καταγράφονται, είναι η προβολή βίντεο. Τα βίντεο αυτά είναι σχεδιασμένα με τέτοιο τρόπο ώστε να προκαλούν συγκεκριμένα συναισθήματα. Αυτή η τεχνική δουλεύει αρκετά καλά, όμως δεν είναι εξίσου εύκολο να προκληθούν όλα τα απαιτούμενα συναισθήματα. Η χαρά για παράδειγμα μπορεί να προκληθεί σχετικά εύκολα. Αντίθετα ο θυμός και ο φόβος απαιτούν την ενεργή συμμετοχή του ατόμου, κάτι που δεν επιτυγχάνεται απλά παρακολουθώντας ένα βίντεο. Πρέπει λοιπόν να βρεθεί κάποιος τρόπος ώστε να μπορούν να καταγραφούν όλα τα πιθανά συναισθήματα και αυτά να είναι αυθόρμητα.

Εκτός από τις βασικές εκφράσεις υπάρχουν πολλές ακόμη οι οποίες πρέπει να μπορούν να αναγνωριστούν. Είναι όμως πολύ δύσκολο, δυσκολότερο και από τις βασικές, να καταγραφούν ιδιαίτερα όταν πρόκειται για αυθόρμητες εκφράσεις.

Μία άλλη πρόκληση είναι οι διαφορές στις εκφράσεις και στα χαρακτηριστικά του προσώπου ατόμων διαφορετικής καταγωγής και ηλικίας. Τα συστήματα αναγνώρισης εκφράσεων πρέπει να δουλεύουν εξίσου καλά ανεξάρτητα από τα ιδιαίτερα χαρακτηριστικά του κάθε ατόμου.

Επιπλέον έρευνα απαιτείται επίσης όσον αφορά τις διαφορετικές γωνίες από τις οποίες μπορεί να παρατηρηθεί το πρόσωπο καθώς και στην αναγνώριση των μικροεκφράσεων.

Όσον αφορά τα Μοντέλα ενεργού περιγράμματος, πρέπει να ληφθούν υπ' όψη και μεταβολές στη θέση του κεφαλιού, στην έκφραση και στο φωτισμό όπου η αναζήτηση αντιμετωπίζει δυσκολίες.

Η αυτόματη τοποθέτηση των landmark points είναι επίσης απαραίτητη, λόγω του χρόνου και του κόπου που απαιτείται όταν αυτή γίνεται χειροκίνητα.

Για το σύστημα αναγνώρισης εκφράσεων που αναπτύχθηκε, μελλοντικά θα μπορούσε να γίνει μία ομαδοποίηση των εκφράσεων σε κλάσεις ανάλογα με τα συναισθήματα που αναπαριστούν και με ένα νευρωνικό δίκτυο ή με support vector machines να γίνεται και αναγνώριση του συναισθήματος.

Παράρτημα Α

Παρακάτω φαίνεται η μορφή ενός αρχείου acf.

```
#####  
#  
# Active Appearance Model Builder Configuration File  
#  
#####  
  
1 # Model reduction [1-n] (reduction factor = 1/x)  
0 # Model expansion [0-n] (pixels along the point normal)  
1 # Use convex hull [0|1] (off/on)  
0 # Verbose mode [0|1] (off/on)  
0 # Write registration movie [0|1] (off/on)  
0 # Write variance image [0|1] (off/on)  
0 # Produce model documentation [0|1] (off/on)  
1 # Use tangent space projection [0|1] (off/on)  
1 # Training method [ 0=PC Regression, 1=Jacobian (recommended) ]  
95 # Shape model truncation (percentage [0-100], -1=parallel analysis)  
95 # Texture model truncation (percentage [0-100], -1=parallel analysis)  
95 # Combined model truncation (percentage [0-100], -2=no combined model)  
1 # Subsampling of the training set (during training) [1-n]  
1 # Warping method [ 0=benchmark, 1=software, 2=hardware (requires OpenGL) ]
```

Παράρτημα Β

Παρακάτω φαίνεται ένα αρχείο asf για το μοντέλο του προσώπου. Το αρχείο αυτό περιέχει πληροφορίες σχετικά με τον αριθμό των σημείων του μοντέλου (number of model points) στην περίπτωση μας 58, για την host image, στην περίπτωση μας 01-1m.bmp, και ένα πίνακα με όλα τα σημεία του μοντέλου. Ο πίνακας αυτός έχει διαστάσεις (number of model points x 7). Η σημασία της κάθε στήλης περιγράφεται παρακάτω:

1. Αριθμός μονοπατιού στο οποίο ανήκει το σημείο.
2. Τύπος μονοπατιού (0 για ανοιχτό, 4 για κλειστό)
3. Συντεταγμένη του σημείου στον οριζόντιο άξονα.
4. Συντεταγμένη του σημείου στον κάθετο άξονα.
5. Αύξοντας αριθμός σημείου.
6. Το προηγούμενο σημείο σε αυτό το μονοπάτι.
7. Το επόμενο σημείο σε αυτό το μονοπάτι.

```
#####  
#  
#   AAM Shape File   -   written: Wednesday June 04 - 2014 [13:23]  
#  
#####  
  
#  
# number of model points  
#  
58  
  
#  
# model points  
#  
# format: <path#> <type> <x rel.> <y rel.> <point#> <connects from>  
<connects to>  
#  
0      4      0.39278128  0.59725987  0      0      1  
0      4      0.40088047  0.63613137  1      0      2  
0      4      0.41637510  0.68218203  2      1      3  
0      4      0.43335186  0.71194879  3      2      4  
0      4      0.45645116  0.73235508  4      3      5  
0      4      0.47904500  0.74536833  5      4      6  
0      4      0.50266843  0.74932161  6      5      7  
0      4      0.53310780  0.74532970  7      6      8  
0      4      0.55671066  0.73619078  8      7      9  
0      4      0.58141610  0.71444090  9      8      10  
0      4      0.60066027  0.67929818  10     9      11  
0      4      0.61540154  0.63794957  11     10     12  
0      4      0.62579174  0.58589818  12     11     12  
1      0      0.59990084  0.45232197  13     20     14  
1      0      0.58574028  0.44257642  14     13     15  
1      0      0.57317992  0.43539332  15     14     16  
1      0      0.55926627  0.43557402  16     15     17  
1      0      0.54310863  0.44781911  17     16     18  
1      0      0.55445613  0.45636333  18     17     19  
1      0      0.57099175  0.46331886  19     18     20
```

1	0	0.58747111	0.46295750	20	19	13
2	0	0.41622672	0.44806337	21	28	22
2	0	0.42971003	0.43745693	22	21	23
2	0	0.44154034	0.43441091	23	22	24
2	0	0.45960555	0.44404210	24	23	25
2	0	0.47416986	0.45310188	25	24	26
2	0	0.45953923	0.45579707	26	25	27
2	0	0.44214834	0.45703032	27	26	28
2	0	0.42933054	0.45679379	28	27	21
3	4	0.53798811	0.39057795	29	29	30
3	4	0.56335069	0.37432801	30	29	31
3	4	0.59274542	0.37618086	31	30	32
3	4	0.60806759	0.39099376	32	31	33
3	4	0.62067770	0.42663927	33	32	33
4	4	0.47331584	0.39177682	34	34	35
4	4	0.44773420	0.37392497	35	34	36
4	4	0.42530237	0.36958307	36	35	37
4	4	0.40607734	0.37924526	37	36	38
4	4	0.39355322	0.42300249	38	37	38
5	0	0.45255647	0.62558509	39	46	40
5	0	0.48742960	0.59685204	40	39	41
5	0	0.50331956	0.60028594	41	40	42
5	0	0.51758555	0.59697543	42	41	43
5	0	0.55549054	0.62433946	43	42	44
5	0	0.52239061	0.65029336	44	43	45
5	0	0.50594836	0.65263408	45	44	46
5	0	0.48720910	0.65073825	46	45	39
6	4	0.48536969	0.45009891	47	47	48
6	4	0.48306063	0.51016292	48	47	49
6	4	0.47278751	0.52375255	49	48	50
6	4	0.46890581	0.54663640	50	49	51
6	4	0.48114260	0.56007880	51	50	52
6	4	0.50057577	0.56076393	52	51	53
6	4	0.52172418	0.55684088	53	52	54
6	4	0.53856475	0.54590132	54	53	55
6	4	0.53699123	0.52249249	55	54	56
6	4	0.52733048	0.50957279	56	55	57
6	4	0.52559884	0.45130565	57	56	57

```

#
# host image
#
01-1m.bmp

```


Βιβλιογραφία

- [1] T.F. Cootes and C.J. Taylor. Statistical Models of Appearance for Computer Vision. March 2004, University of Manchester.
- [2] Mikkel B. Stegmann. Analysis and Segmentation of Face Images using Point Annotations and Linear Subspace Techniques. August 2002. Technical University of Denmark.
- [3] Vinay Bettadapura. Face Expression Recognition and Analysis: The State of the Art. Georgia Institute of Technology.
- [4] Dacher Keltner, Paul Ekman, Facial Expression of Emotion, In M.Lewis and J Haviland-Jones Handbook of emotions 2nd edition, 2000
- [5] David Matsumoto, Dacher Keltner, Michelle N. Shiota, Maureen O'Sullivan and Mark Frank, Facial Expressions of Emotion, Handbook of Emotions 3rd edition
- [6] M. Pantic, M. Valstar, R. Rademaker, and L. Maat, *Web-based Database for Facial Expression Analysis*, Proc. of IEEE International Conference on Multimedia and Expo (ICME05), July 2005.
- [7] Kanade, T., Cohn, J. F., & Tian, Y. (2000). Comprehensive database for facial expression analysis. Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition (FG'00), Grenoble, France, 46-53.
- [8] Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., & Matthews, I. (2010). The Extended Cohn-Kanade Dataset (CK+): A complete expression dataset for action unit and emotion-specified expression. Proceedings of the Third International Workshop on CVPR for Human Communicative Behavior Analysis (CVPR4HB 2010), San Francisco, USA, 94-101.