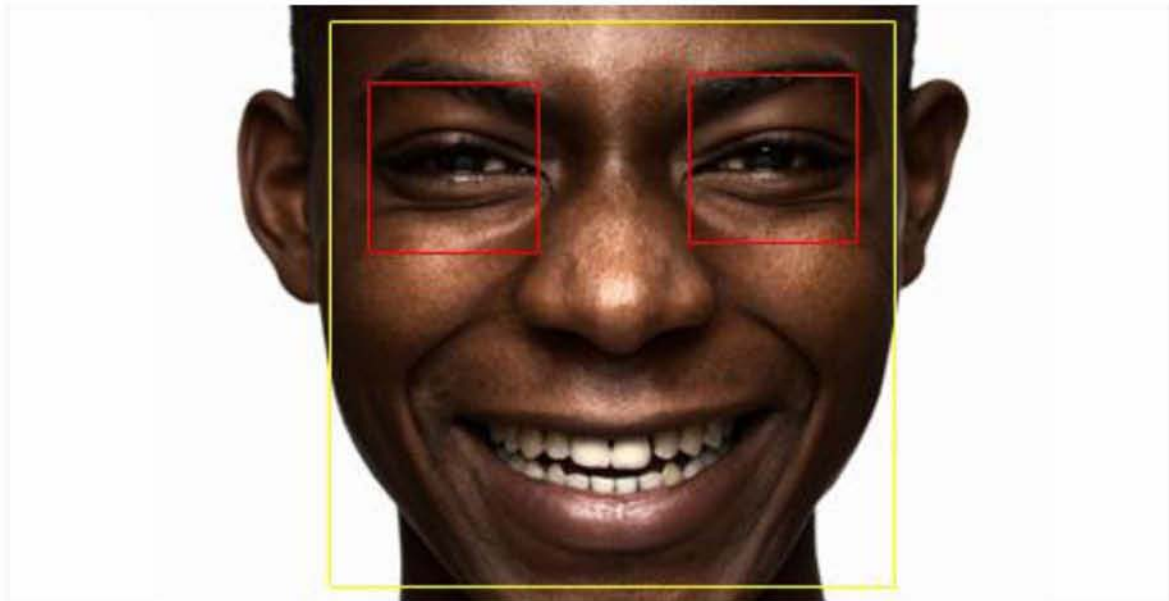


Πανεπιστήμιο Θεσσαλίας
Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών

Διπλωματική Εργασία
Αναγνώριση συναισθημάτων με χρήση βίντεο
Αντωνέλης Χάρης



Εξεταστική Επιτροπή
Επιβλέπων: Ιωάννης Κατσαβουνίδης
Μέλος: Γεράσιμος Ποταμιάνος

Βόλος 2014

Ευχαριστίες

Η παρούσα διπλωματική εργασία με τίτλο “Αναγνώριση συναισθημάτων με χρήση βίντεο” αποτελεί την ολοκλήρωση των σπουδών μου στο τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Πανεπιστημίου Θεσσαλίας. Η τελική διαμόρφωση της εργασίας είναι αποτέλεσμα εκτενούς αναζήτησης και έρευνας, πειραμάτων και δοκιμών καθώς και προϊόν υποστήριξης και καθοδήγησης συγκεκριμένων ανθρώπων τους οποίους θα ήθελα σε αυτό το σημείο να ευχαριστήσω θερμά.

Αρχικά θα ήθελα να ευχαριστήσω ιδιαίτερα τον επιβλέποντα καθηγητή της διπλωματικής μου εργασίας, κ. Κατσαβουνίδη Ιωάννη, για την εμπιστοσύνη που μου έδειξε κατά την επιλογή του θέματος, τις πολύτιμες συμβουλές και τις ουσιαστικές παρατηρήσεις για την ολοκλήρωση της εργασίας. Ακόμα, τον ευχαριστώ για τη γνώση που μου μετέφερε στα μαθήματα και το ειδικό θέμα που πραγματοποίησα με τη βοήθειά του.

Ιδιαίτερες ευχαριστίες θα ήθελα να δώσω και στο Στέργιο Πουλαράκη για τη βοήθεια, τη γνώση, τις συμβουλές και την υποστήριξη που μου έδωσε, καθώς και όλη την ομάδα φοιτητών, μεταπτυχιακών και διδακτορικών του κ. Κατσαβουνίδη, μέσα στην οποία όλοι μας παίρνουμε σπουδαία εφόδια για να δουλέψουμε αύριο ως μηχανικοί. Ευχαριστίες όμως, οφείλω σε όλους τους καθηγητές του ΤΗΜΜΥ οι γνώσεις, οι συμβουλές και οι παρατηρήσεις των οποίων συνέβαλαν στη λήψη του πτυχίου μου και αποτελούν σημαντικό εφόδιο ζωής.

Δεν μπορώ να παραλείψω να ευχαριστήσω τους φίλους και συμφοιτητές μου για την υποστήριξη τους και τη συνεχή ενθάρρυνση που μου επέδειξαν καθώς και για τα υπέροχα φοιτητικά χρόνια που μου χάρισαν. Τέλος θα ήθελα να ευχαριστήσω την οικογένεια μου, για την αμέριστη υποστήριξη που μου επέδειξε κατά τη διάρκεια των σπουδών μου η οποία αποτέλεσε ισχυρό κίνητρο για την επιτυχή ολοκλήρωσή τους.

Αντωνέλης Χάρης

Βόλος 2014

Περίληψη

Η αναγνώριση συναισθημάτων μέσω εικόνας και βίντεο αποτελεί πεδίο ανάπτυξης τα τελευταία χρόνια, ενώ η μελέτη του αντικειμένου της αναγνώρισης συναισθημάτων απασχολεί την ψυχολογική κοινότητα τις τελευταίες δεκαετίες. Βρίσκει εφαρμογή σε θέματα επικοινωνίας, ιατρικά θέματα, θέματα ασφαλείας. Βασίζεται στο γεγονός ότι υπάρχει καθολικότητα των εκφράσεων σε όλους τους ανθρώπους όλων των πολιτισμών. Τα βήματα που ακολουθούνται στην αναγνώριση είναι η ανίχνευση του προσώπου, η εξαγωγή χαρακτηριστικών και τελικά η κατηγοριοποίηση. Στην προσέγγιση που ακολουθεί η υπάρχουσα διπλωματική, χρησιμοποιείται η μέθοδος “Object Detection using Haar cascades”, μέσω της οποίας εντοπίζεται αρχικά το πρόσωπο και μετά το στόμα του αντικειμένου. Στη συνέχεια γίνεται επεξεργασία της εικόνας του στόματος, ώστε να εξαχθούν τα χαρακτηριστικά που διαχωρίζουν τρεις κατηγορίες συναισθημάτων: ουδέτερο, χαρά, έκπληξη. Η διαδικασία της κατηγοριοποίησης πραγματοποιείται με εκπαίδευση σε Support Vector Machines. Χρησιμοποιείται η βάση δεδομένων Cohn-Kanade και CK+ (Cohn-Kanade plus).

Abstract

During the past few years emotion recognition with the use of image and video processing is an active field of research, whereas emotion recognition has been embraced by the psychology community during the past few decades. Its most notable applications are communication, health and security issues. It is based on the fact that facial expressions of emotion are constant across cultures. Automatic emotion recognition systems are divided into three modules: face tracking and detection, feature extraction and expression classification. In our approach, we track the human face and the human mouth with the use of “Object Detection using Haar cascades” method. We continue by processing the image of the mouth, so we can extract the features that distinguish three categories of emotions: neutral, happiness, surprise. The classification process is performed by training our data using Support Vector Machines. We use the Cohn-Kanade and CK+ (Cohn-Kanade plus) database.

Περιεχόμενα

| | |
|--|----|
| 1.Εισαγωγή | 6 |
| Εκφράσεις ίδιες για όλους; | 7 |
| Έξι βασικά συναισθήματα | 8 |
| Σκόπιμες – αυθόρμητες εκφράσεις..... | 10 |
| Βάσεις δεδομένων αυθόρμητων εκφράσεων | 11 |
| Μικροεκφράσεις - micro expressions..... | 12 |
| Χαρακτηριστικά προσώπου | 12 |
| Χαρακτηριστικά ενός καλού συστήματος | 13 |
| Εντοπισμός προσώπου, παρακολούθηση και εξαγωγή χαρακτηριστικών | 13 |
| 2.Object detection using Haar cascades..... | 18 |
| Χαρακτηριστικά | 18 |
| Integral Image..... | 19 |
| Adaboost..... | 20 |
| Attentional cascade | 22 |
| 3.RGB to Grey | 25 |
| 4.Gaussian smoothing | 27 |
| 5.Canny edge detector..... | 30 |
| 6.Cohn-Kanade-dataset | 32 |
| 7.Εξαγωγή χαρακτηριστικών στη βάση δεδομένων Cohn-Kanade | 34 |
| 8.Χαρακτηριστικά | 36 |
| 9.Support Vector Machines | 45 |
| Αποτελέσματα | 47 |
| State of the art..... | 48 |
| 10.Χρόνος εκτέλεσης..... | 49 |
| 11.Συμπεράσματα και μελλοντική εργασία | 50 |
| Βιβλιογραφία..... | 51 |

1.Εισαγωγή

Οι μελέτες για τις εκφράσεις του προσώπου και τη φυσιognωμία ξεκίνησαν από την Αριστοτέλεια Εποχή (4ος αιώνας π.Χ.). Η φυσιognωμία ορίζεται ως η εκτίμηση του χαρακτήρα ή της προσωπικότητας ενός ατόμου από την εξωτερική τους εμφάνιση, ειδικά το πρόσωπο. Ενώ όμως όλα αυτά τα χρόνια το ενδιαφέρον για τη φυσιognωμία έχει χαθεί, η μελέτη των εκφράσεων του προσώπου υπήρξε σταθερά ενεργό θέμα. Οι θεμελιώδεις μελέτες για τις εκφράσεις του προσώπου, που αποτέλεσαν τη βάση της σημερινής έρευνας, ξεκινούν από τον 17ο αιώνα. Ένα λεπτομερές σημείωμα σχετικά με τις διάφορες εκφράσεις και την κίνηση των μυών της κεφαλής δόθηκε το 1649 από τον John Bulwer στο βιβλίο του “Pathomyotomia”. Μια άλλη ενδιαφέρουσα εργασία σχετικά με τις εκφράσεις του προσώπου (και τη φυσιognωμία) ήταν του Le Brun, Γάλλου ακαδημαϊκού και ζωγράφου. Το 1667, ο Le Brun έδωσε μια διάλεξη στο Royal Academy of Painting, που αργότερα γίνεται βιβλίο (1734). Είναι ενδιαφέρον να γνωρίζουμε ότι ηθοποιοί και καλλιτέχνες του 18ου αιώνα αναφέρονταν στο βιβλίο του προκειμένου να επιτευχθεί η “τέλεια απομίμηση της “γνήσιας” έκφρασης προσώπου”. Προχωρώντας στο 19ο αιώνα, ένα από τα σημαντικά έργα στην ανάλυση της έκφρασης του προσώπου που έχει άμεση σχέση με τη σύγχρονη επιστήμη της αυτόματης αναγνώρισης προσώπου, ήταν το έργο του Κάρολου Δαρβίνου. Το 1872, ο Δαρβίνος έγραψε μια πραγματεία που καθόρισε τις γενικές αρχές της έκφρασης και τα μέσα της έκφρασης τόσο στον άνθρωπο όσο και στα ζώα. Επίσης κατέταξε διάφορα είδη εκφράσεων σε κατηγορίες. Η κατηγοριοποίησή του ήταν η εξής [1]:

- κακή διάθεση, άγχος, θλίψη, κατήφεια, απόγνωση
- χαρά, κέφι, αγάπη, τρυφερά αισθήματα, αφοσίωση
- περισυλλογή, διαλογισμός, κακή διάθεση, κατσούφιασμα , αποφασιστικότητα
- μίσος, θυμός
- περιφρόνηση, αηδία, ενοχή , υπερηφάνεια
- έκπληξη, φόβος, τρόμος
- ντροπή, συστολή, σεμνότητα

Επιπλέον, ο Δαρβίνος καταλογογράφησε τις παραμορφώσεις του προσώπου που συμβαίνουν για καθεμία από τις παραπάνω κατηγορίες εκφράσεων. Για παράδειγμα, η σύσπαση των μυών γύρω από τα μάτια στη θλίψη, το σφιχτό κλείσιμο του στόματος στον προβληματισμό, η πίεση στις γωνίες του στόματος σε κακή διάθεση κλπ.

Ένα άλλο σημαντικό ορόσημο στη μελέτη των εκφράσεων του προσώπου και τα ανθρώπινα συναισθήματα είναι το έργο που επιτέλεσε ο ψυχολόγος Paul Ekman και οι συνεργάτες του το 1970. Το έργο τους είναι εξαιρετικά σημαντικό και έχει μεγάλη επιρροή στην ανάπτυξη των σύγχρονων αυτόματων αναγνωριστών των εκφράσεων του προσώπου.

Παραδοσιακά, όπως είδαμε, οι εκφράσεις του προσώπου έχουν μελετηθεί από τους κλινικούς και κοινωνικούς ψυχολόγους, ιατρούς, ηθοποιούς και καλλιτέχνες. Ωστόσο, το τελευταίο τέταρτο του 20ου αιώνα, με τις εξελίξεις στους τομείς της ρομποτικής, των γραφικών και του computer vision, animators και επιστήμονες της επιστήμης των υπολογιστών ξεκίνησαν να δείχνουν ενδιαφέρον για τη μελέτη των εκφράσεων του προσώπου. Το πρώτο βήμα προς την αυτόματη αναγνώριση των εκφράσεων του προσώπου πραγματοποιήθηκε το 1978. Ο Suwa και οι συνάδελφοί του παρουσίασαν ένα σύστημα για την ανάλυση των εκφράσεων του προσώπου από μια αλληλουχία εικόνων, χρησιμοποιώντας είκοσι σημεία παρακολούθησης. Παρά το γεγονός ότι το σύστημα αυτό προτάθηκε το 1978, οι ερευνητές δεν ακολούθησαν αυτή τη γραμμή μελέτης μέχρι τις αρχές της δεκαετίας του 1990. Ο λόγος γι' αυτό είναι ο εξής: η αυτόματη αναγνώριση των εκφράσεων του προσώπου απαιτεί ισχυρό αλγόριθμο ανίχνευσης προσώπου. Αυτά ήταν θέματα έρευνας προς ανάπτυξη εκείνο τον καιρό. Μέχρι τα τέλη της δεκαετίας του 1990, φθινή υπολογιστική δύναμη άρχισε να γίνεται διαθέσιμη. Αυτό οδήγησε στην ανάπτυξη ισχυρών αλγορίθμων ανίχνευσης και εντοπισμού προσώπου στις αρχές του 1990. Την ίδια στιγμή, η αλληλεπίδραση ανθρώπου-υπολογιστή άρχισε να γίνεται δημοφιλής. Οι ερευνητές διαπίστωσαν ότι χωρίς συστήματα αυτόματου εντοπισμού εκφράσεων και συναισθημάτων, οι υπολογιστές θα παραμένουν “ψυχροί”, σχετικά με τη συναισθηματική κατάσταση του χρήστη. Όλοι αυτοί οι παράγοντες οδήγησαν σε αύξηση του ενδιαφέροντος για την ανάπτυξη τέτοιων συστημάτων.

Εκφράσεις ίδιες για όλους;

Ένας τρόπος να δείχνουμε συναισθήματα είναι μέσω των αλλαγών στις εκφράσεις προσώπου. Αλλά είναι αυτές οι εκφράσεις καθολικές σε όλους τους πολιτισμούς; Για μεγάλο χρονικό διάστημα, ανθρωπολόγοι και ψυχολόγοι μελετούσαν την απάντηση στην ερώτηση αυτή. Με βάση τη θεωρία του για την εξέλιξη, ο Δαρβίνος υποστήριξε ότι ήταν καθολική. Ωστόσο, οι απόψεις ήταν ποικίλες και δεν υπήρχε γενική συναίνεση. Το 1971, οι Ekman και Friesen διεξήγαγαν μελέτες σε δυτικούς και ανατολικούς πολιτισμούς και ανέφεραν ότι οι εκφράσεις του προσώπου ήταν πράγματι σταθερές στους πολιτισμούς. Τα αποτελέσματα αυτής της έρευνας έγιναν αποδεκτά από την επιστημονική κοινότητα. Το 1994, ο Russell έγραψε μια κριτική που αμφισβητούσε τους ισχυρισμούς τους. Το ίδιο έτος, οι Ekman και Izard απάντησαν σε αυτή την κριτική με ισχυρά

αποδεικτικά στοιχεία και διέψευσαν σημείο προς σημείο τους ισχυρισμούς του Russell. Από τότε, έχει θεωρηθεί ως ένα αδιαμφισβήτητο γεγονός ότι η αναγνώριση των συναισθημάτων από τις εκφράσεις του προσώπου είναι καθολική και σταθερή σε όλους τους πολιτισμούς. Διαπολιτισμικές έρευνες έδειξαν ότι, αν και η ερμηνεία των εκφράσεων είναι καθολική και σταθερή σε όλους τους πολιτισμούς, η έκφραση των συναισθημάτων μέσω των αλλαγών στο πρόσωπο εξαρτώνται απ' το κοινωνικό πλαίσιο. Για παράδειγμα, όταν Αμερικάνοι και Ιάπωνες παρακολούθησαν βίντεο που προκαλούσαν το ίδιο συναισθήματα, παρουσίασαν παρόμοιες εκφράσεις του προσώπου. Ωστόσο, παρουσία μιας αρχής, οι Ιάπωνες ήταν πολύ περισσότερο διστακτικοί να δείξουν τα πραγματικά συναισθήματά τους μέσω αλλαγών στις εκφράσεις του προσώπου.

Έξι βασικά συναισθήματα

Στις διαπολιτισμικές τους μελέτες, οι Ekman και Friesen χρησιμοποίησαν μόνο έξι συναισθήματα [1]:

- χαρά
- λύπη
- θυμό
- έκπληξη
- αηδία
- φόβο

Σύμφωνα με τους ίδιους τα έξι συναισθήματα που μελετήθηκαν ήταν αυτά που βρέθηκαν από παραπάνω από ένα ερευνητή να είναι διαχωρίσιμα σε κάθε ένα εγγράμματο πολιτισμό. Αυτές οι έξι εκφράσεις είναι γνωστές ως “βασικές”, “πρωτότυπες” ή “αρχέτυπες” εκφράσεις. Από τις αρχές της δεκαετίας του 1990, οι ερευνητές έχουν επικεντρωθεί στην ανάπτυξη συστημάτων που αναγνωρίζουν τις έξι αυτές βασικές εκφράσεις.

Εκτός από τις έξι βασικές εκφράσεις, το ανθρώπινο πρόσωπο δημιουργεί εκφράσεις για μια ποικιλία από άλλα συναισθήματα. Το 2000, ο Parrott διέκρινε 136 τέτοιες συναισθηματικές καταστάσεις, τις οποίες κατέταξε σε ξεχωριστές κατηγορίες και υποκατηγορίες. Αυτή η κατηγοριοποίηση φαίνεται στον παρακάτω πίνακα. Τα τελευταία χρόνια υπήρξαν προσπάθειες να διακρίνουμε συναισθήματα πέρα από τα έξι βασικά, κάτι το οποίο όμως ξεφεύγει από τους σκοπούς της εργασίας αυτής.

| Βασικό συναίσθημα | Δευτερεύον συναίσθημα | Τριτεύοντα συναισθήματα |
|--------------------------|------------------------------|---|
| Αγάπη | Στοργή | Στοργή, αγάπη, συμπάθεια, έλξη, φροντίδα, τρυφερότητα, συμπόνια, συναισθηματισμός |
| | Λαγνεία | Διέγερση, επιθυμία, πάθος |
| | Λαχτάρα | Λαχτάρα |
| Χαρά | Χαρά | Ψυχαγωγία, ευδαιμονία, ευθυμία, χαρά, απόλαυση, αγαλλίαση |
| | Όρεξη | Ενθουσιασμός, ζήλος, όρεξη, συγκίνηση, χαρά |
| | Ικανοποίηση | Ικανοποίηση, χαρά |
| | Υπερηφάνεια | Υπερηφάνεια, θρίαμβος |
| | Αισιοδοξία | Προθυμία, ελπίδα, αισιοδοξία |
| | Ανακούφιση | Ανακούφιση |
| Έκπληξη | Έκπληξη | |
| Θυμός | Ερεθισμός | Ανησυχία, ενόχληση, γκρίνια |
| | Εξόργιση | Αγανάκτηση, απογοήτευση |
| | Οργή | Θυμός, οργή, αγανάκτηση, εχθρότητα, αγριότητα, πίκρα, μίσος, απέχθεια, περιφρόνηση, αντιπάθεια, δυσαρέσκεια |
| | Αηδία | Αηδία, αποστροφή, περιφρόνηση |
| | Ζήλια | Φθόνος, ζήλια |
| Θλίψη | Ταλαιπωρία | Αγωνία, πόνος |

| | | |
|-------|-------------|---|
| | Θλίψη | Κατάθλιψη, απελπισία, απόγνωση, κατήφεια, θλίψη, δυστυχία, μελαγχολία |
| | Απογοήτευση | Απογοήτευση, δυσαρέσκεια |
| | Ντροπή | Ενοχή, ντροπή, λύπη, τύψεις |
| | Παραμέληση | Παραμέληση, αποξένωση, απομόνωση, απόρριψη, ήττα, ανασφάλεια, αμηχανία, ταπείνωση |
| | Λύπηση | Λύπηση, συμπάθεια |
| Φόβος | Φρίκη | Σοκ, φόβος, τρόμος, φρίκη, πανικός, υστερία |
| | Νευρικότητα | Άγχος, νευρικότητα, υπερένταση, ανησυχία, φόβος, αγωνία |

Σκόπιμες – αυθόρμητες εκφράσεις

Μέχρι τώρα είδαμε τις βασικές και μη-βασικές εκφράσεις. Ας δούμε τώρα μια δεύτερη κατηγοριοποίηση, τις σκόπιμες και τις αυθόρμητες εκφράσεις. Σκόπιμες είναι οι εκφράσεις που κάποιος θα πραγματοποιήσει όταν θα του ζητηθεί. Αυτή είναι συνήθως η κατάσταση όταν κάποιος βρίσκεται στο εργαστήριο υπό συνθήκες δοκιμής. Σε αντίθεση, οι αυθόρμητες εκφράσεις είναι οι γνήσιες ανθρώπινες εκφράσεις. Αυτές είναι οι εκφράσεις που βλέπουμε σε καθημερινές συνομιλίες, βλέποντας ταινίες κλπ. Από τις αρχές του 1990, μέχρι τα τελευταία χρόνια, οι περισσότερες έρευνες για αυτόματη αναγνώριση συναισθημάτων βασίζονται σε σκόπιμες-στημένες εκφράσεις. Αυτό συμβαίνει επειδή οι σκόπιμες εκφράσεις είναι εύκολο να ληφθούν σε βάση δεδομένων. Πολλοί ψυχολόγοι έχουν αποδείξει ότι οι στημένες και οι αυθόρμητες εκφράσεις διαφέρουν στην εμφάνιση, στο χρονικό διάστημα και στα χαρακτηριστικά. Επιπλέον, αποδεικνύεται ότι πολλές από τις εκφράσεις που έχουν χρησιμοποιηθεί σε σημερινά συστήματα είναι “υπερβολικές”, όπως φαίνεται στην παρακάτω εικόνα. Αντίθετα, οι φυσικές εκφράσεις είναι πιο ήπιες.



Εικόνα 1.1: Σκόπιμες, φανερά υπερβολικές εκφράσεις έκπληξης, αποστροφής και θυμού.

Βάσεις δεδομένων αυθόρμητων εκφράσεων

Ένας από τους σημαντικότερους παράγοντες στην ανάπτυξη ενός νέου συστήματος αναγνώρισης συναισθημάτων είναι η επιλογή της βάσης δεδομένων. Αν χρησιμοποιείται μία βάση δεδομένων από όλους τους ερευνητές, η απόδοση του νέου συστήματος, συγκρινόμενη με τα προηγούμενα, γίνεται μια απλή και γραμμική υπόθεση. Ωστόσο, η κατασκευή μιας αξιόπιστης βάσης δεδομένων για την αναγνώριση συναισθημάτων είναι μια δύσκολη υπόθεση. Όπως είπαμε παραπάνω, υπάρχουν διαφορές ανάμεσα στις σκόπιμες και τις αυθόρμητες εκφράσεις. Τα βασικά προβλήματα που διαπιστώθηκαν από τον Sebe και τους συναδέλφους του είναι:

- Διαφορετικοί άνθρωποι εκφράζουν τα ίδια συναισθήματα με διαφορετική ένταση
- Αν ο συμμετέχων στο πείραμα αντιληφθεί ότι φωτογραφίζεται, η έκφραση χάνει την αυθεντικότητά της
- Ακόμα και αν δεν αντιληφθεί ότι φωτογραφίζεται, το περιβάλλον του εργαστηρίου δεν ενθαρρύνει αυθόρμητες εκφράσεις.

Η λύση που δόθηκε από την ομάδα του Sebe ήταν η εξής: δημιούργησαν ένα χώρο που άνθρωποι μπορούσαν να παρακολουθήσουν βίντεο που δημιουργούσαν συναισθήματα. Οι εκφράσεις των ανθρώπων αυτών καταγράφονταν σε βίντεο από μια κρυμμένη κάμερα. Αφού η καταγραφή ολοκληρωνόταν, τους γνωστοποιούνταν και η ομάδα τους ζητούσε την άδειά τους να χρησιμοποιήσει το υλικό στην έρευνα. Το 50% των συμμετεχόντων έδωσε την άδεια. Κατόπιν, περιέγραφαν τα πραγματικά συναισθήματα που είχαν νιώσει τη στιγμή της καταγραφής. Από την έρευνα, η ομάδα συμπέρανε ότι ήταν πολύ δύσκολο να προκαλέσεις μια μεγάλη γκάμα συναισθημάτων, όπως ο φόβος και η λύπη. Επίσης διαπίστωσαν ότι οι αυθόρμητες εκφράσεις

μπορούν να δημιουργήσουν λάθος συμπεράσματα. Για παράδειγμα, κάποιοι άνθρωποι είχαν έκφραση που έδειχνε λυπημένη, ενώ αισθάνονταν καλά στην πραγματικότητα.

Μικροεκφράσεις - micro expressions

Μικροεκφράσεις είναι οι εκφράσεις που προβάλλουν οι άνθρωποι όταν προσπαθούν να κρύψουν τα αληθινά τους συναισθήματα. Αντίθετα από τις κανονικές εκφράσεις, οι μικρο εκφράσεις διαρκούν μόνο 1/25 του δευτερολέπτου. Δεν υπήρξε καμία έρευνα που να γίνεται σχετικά με τις μικροεκφράσεις, κυρίως λόγω του γεγονότος ότι αυτές οι εκφράσεις εμφανίζονται πολύ γρήγορα και είναι πολύ πιο δύσκολο να συλληφθούν και να αποτυπωθούν. Ο Ekman έχει γράψει λεπτομερώς σχετικά με τη χρήση των μικροεκφράσεων στον εντοπισμό της απάτης και του ψέματος.

Χαρακτηριστικά προσώπου

Έχοντας συζητήσει τα συναισθήματα και τις συναφείς εκφράσεις του προσώπου, ας δούμε τώρα τα χαρακτηριστικά (features) του προσώπου. Τα χαρακτηριστικά του προσώπου χωρίζονται σε μόνιμα και παροδικά. Μόνιμα είναι χαρακτηριστικά όπως τα μάτια, τα χείλη, τα φρύδια και τα μάγουλα. Παροδικά είναι χαρακτηριστικά όπως γραμμές προσώπου, ρυτίδες, αυλάκια που εμφανίζονται σε αλλαγές στο πρόσωπο και εξαφανίζονται σε ένα ουδέτερο ύφος. Ας δούμε τώρα τα βασικά χαρακτηριστικά του προσώπου και το ρόλο που παίζουν στην αναγνώριση των συναισθημάτων. Η περισσότερη πληροφορία για την αναγνώριση των εκφράσεων βρίσκεται στο στόμα και στα φρύδια. Αυτό αποδεικνύεται από τους Pardas και Bonafonte. Το έργο τους δείχνει ότι η έκπληξη, η ευχαρίστηση και η απέχθεια έχουν πολύ υψηλή αναγνωρισιμότητα (100%, 93,4%, 97.3% αντιστοίχως) επειδή περιλαμβάνουν σαφή κίνηση του στόματος και των φρυδιών. Ένα ακόμα ενδιαφέρον αποτέλεσμα από την εργασία τους δείχνει ότι το στόμα περιέχει περισσότερη πληροφορία από τα φρύδια. Οι δοκιμές που διεξάχθηκαν μόνο με το στόμα να είναι ορατό έδωσαν 78% ακρίβεια αναγνώρισης, ενώ αντίστοιχα τα φρύδια έδωσαν 50%. Ένα ακόμα αποτέλεσμα σχετιζόμενο με τη μελέτη των Bourel et al. έδειξε ότι η θλίψη αναγνωρίζεται κυρίως από το στόμα. Μπορούμε να συμπεράνουμε ότι, με σειρά προτεραιότητας, το στόμα και τα φρύδια είναι τα πιο σημαντικά χαρακτηριστικά του προσώπου. Οι Kotsia et al. έφτασαν στο συμπέρασμα ότι η εξέταση του δεξιού ή του αριστερού μισού του προσώπου δεν αλλάζει τα αποτελέσματα. Αυτό είναι προφανές από τη συμμετρικότητα του ανθρώπινου προσώπου. Ένα ακόμα σημείο που εξετάζεται είναι οι διαφορές στα χαρακτηριστικά του προσώπου μεταξύ διαφορετικών ανθρώπων, όπως η υφή του δέρματος και η ύπαρξη τριχών, η διαφορά ανοίγματος ματιού ανάμεσα στους Ασιάτες και τους

Ευρωπαίους κλπ.

Χαρακτηριστικά ενός καλού συστήματος

Ένα σύστημα αναγνώρισης συναισθημάτων πρέπει να έχει τα παρακάτω χαρακτηριστικά [1]:

- Θα πρέπει να είναι πλήρως αυτόματο
- Θα πρέπει να έχει τη δυνατότητα να επεξεργαστεί βίντεο και εικόνες
- Θα πρέπει να λειτουργεί σε πραγματικό χρόνο
- Θα πρέπει να μπορεί να αναγνωρίσει αυθόρμητες εκφράσεις
- Θα πρέπει να μπορεί να λειτουργεί σε διαφορετικές συνθήκες φωτισμού
- Θα πρέπει να δουλεύει σχετικά καλά ακόμα και με κομμάτια του προσώπου κρυμμένα
- Θα πρέπει να δουλεύει σε ανθρώπους με διαφορετικό χρώμα δέρματος και ηλικία
- Θα πρέπει να δουλεύει με εικόνες διαφορετικών αναλύσεων

Μέχρι τώρα, διαφορετικές ερευνητικές ομάδες έχουν εστιάσει σε διαφορετικά από τα παραπάνω κριτήρια. Παραμένει στοιχείο μελέτης η ενοποίηση των ερευνών αυτών προκειμένου να δημιουργηθεί ένα σχεδόν ιδανικό σύστημα.

Εντοπισμός προσώπου, παρακολούθηση και εξαγωγή χαρακτηριστικών

Τα συστήματα αυτόματης αναγνώρισης συναισθημάτων χωρίζονται σε τρεις ενότητες:

- Ανίχνευση προσώπου
- Εξαγωγή χαρακτηριστικών
- Κατηγοριοποίηση συναισθήματος

Το πρώτο βήμα στην ανάλυση προσώπου είναι να εντοπιστεί το πρόσωπο στην εικόνα ή το βίντεο που έχουμε. Η έρευνα στους τομείς του εντοπισμού προσώπου υπήρξε πολύ δραστήρια και υπάρχει πολύ μεγάλη βιβλιογραφία. Παρακάτω θα περιγραφούν οι βασικοί αλγόριθμοι εντοπισμού προσώπου που χρησιμοποιήθηκαν τα τελευταία χρόνια (από τις αρχές της δεκαετίας του 2000).

Μία από τις μεθόδους που αναπτύχθηκε στις αρχές της δεκαετίας του 1990 για ανίχνευση προσώπων ήταν ο Kanade-Lucas-Tomasi. Το 1998, οι Kanade και Schneiderman ανέπτυξαν ένα

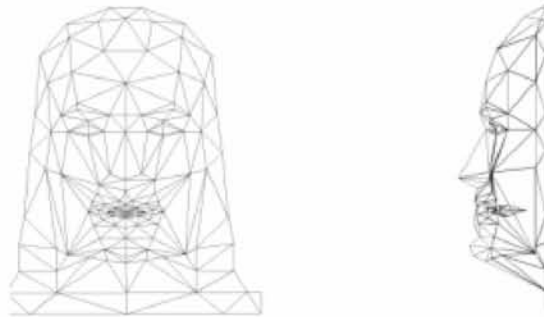
σταθερό αλγόριθμο εντοπισμού προσώπων, χρησιμοποιώντας στατιστικές μεθόδους, ο οποίος χρησιμοποιήθηκε εκτενώς από τότε που προτάθηκε. Το 2004, οι Viola και Jones ανέπτυξαν μία μέθοδο χρησιμοποιώντας τον αλγόριθμο Adaboost, η οποία είναι πολύ γρήγορη και μπορεί γρήγορα να εντοπίσει πρόσωπα “ανφάς”. Αυτός είναι και ο αλγόριθμος που χρησιμοποιείται και στην παρούσα εργασία. Άλλοι γνωστοί αλγόριθμοι προτάθηκαν από τους Sung και Poggio το 1998, Rowel et al. το 1998 και Roth et al. το 2000.

Μία μέθοδος εντοπισμού προσώπου που χρησιμοποιήθηκε ευρέως είναι η Piecewise Bezier Volume Deformation (PBVD) tracker, από τους Tao και Huang. Η μέθοδος χρησιμοποιεί ένα generic 3D wire frame μοντέλο του προσώπου που συνδέεται με 16 Bezier volumes. Η ίδια μέθοδος χρησιμοποιείται στην ανάλυση κίνησης προσώπου και στο computer animation.

Μία άλλη μέθοδος είναι αυτή που χρησιμοποιεί το μοντέλο προσώπου 3D Candidate. Το μοντέλο αυτό κατασκευάζεται από ένα σύνολο από τρίγωνα. Αρχικά αναπτύχθηκε απ' τον Rydfalk το 1987 για model based computer animation. Το μοντέλο προσώπου αργότερα τροποποιήθηκε από τους Stromberg, Welsh και Ahlberg, με τα μοντέλα τους γνωστά ως Candidate-1, Candidate-2 και Candidate-3 αντίστοιχα. Σήμερα χρησιμοποιείται περισσότερο το μοντέλο Candidate-3. Η παραπάνω μέθοδος είναι μία από τις δημοφιλέστερες που χρησιμοποιούνται σήμερα από τους ερευνητές του τομέα.



Μοντέλο προσώπου Candidate - 1



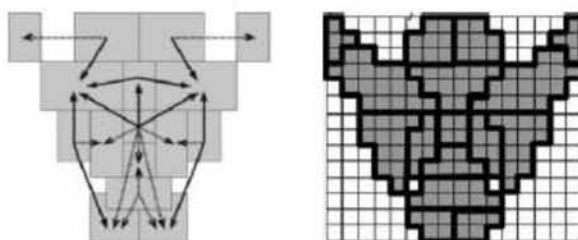
Μοντέλο προσώπου Candidate - 2



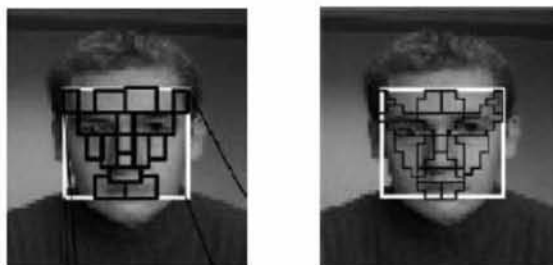
Μοντέλο προσώπου Candidate - 3

Εικόνα 1.2: Μοντέλα προσώπου Candidate

Ακόμα μία μέθοδος είναι η Ratio Template Tracker. Ο αλγόριθμος αυτός προτάθηκε απ' τον Sinha το 1996. Ο Scassellati το χρησιμοποίησε προκειμένου να φτιάξει ένα σύστημα που εντοπίζει τα μάτια εντοπίζοντας πρώτα το πρόσωπο σε πραγματικό χρόνο. Το 2004, οι Anderson και McOwen τροποποίησαν τον αλγόριθμο ενσωματώνοντας το “Golden Ratio”. Με αυτή την προσθήκη, ο αλγόριθμος δούλεψε καλύτερα υπό διαφορετικές συνθήκες φωτισμού. Ο αλγόριθμος επιτυχημένα χρησιμοποιήθηκε στην αναγνώριση συναισθημάτων.



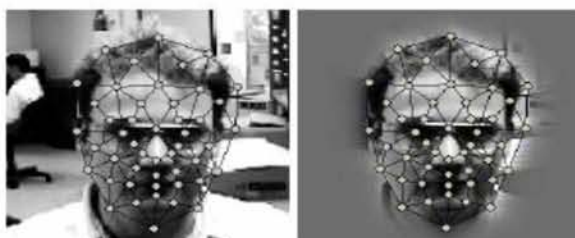
Η εικόνα αριστερά είναι το μοντέλο για τη μέθοδο 'Radio Template'. Αποτελείται από 16 περιοχές (γκρι κουτάκια) και 23 σχέσεις (βέλη). Στα δεξιά το μοντέλο της μεθόδου 'Spatial Radio Template'.



Η εικόνα στα αριστερά δείχνει την εφαρμογή του μοντέλου 'Radio Template' σε ένα πρόσωπο. Αντίστοιχα στη δεξιά εικόνα εφαρμόζεται το μοντέλο 'Spatial Radio Template'.

Εικόνα 1.3: Μέθοδος Radio Template

Μία άλλη μέθοδος που πραγματοποιεί γρήγορο και αξιόπιστο εντοπισμό είναι το σύστημα PersonSpotter. Αναπτύχθηκε το 1998 από τους Steffens et al. για εντοπισμό πραγματικού χρόνου. Ο αλγόριθμος εντοπίζει περιοχές ενδιαφέροντος που περιέχουν κινούμενα αντικείμενα. Αυτό το κάνει συγκρίνοντας τις εικόνες και βρίσκοντας αλλαγές. Εντοπιστές δέρματος και εντοπιστές συνοχής εφαρμόζονται μετά προκειμένου να βρουν το πρόσωπο. Το σύστημα αυτό, εξαιτίας της μεθόδου αυτής, είναι χρήσιμο για εντοπισμό του προσώπου σε γρήγορη κίνηση του background.



Εικόνα 1.4: Το μοντέλο της μεθόδου PersonSpotter και background suppression.

Τέλος, έχουν προταθεί και άλλες ενδιαφέρουσες προσεγγίσεις. Αντί για εντοπισμό του προσώπου, οι Tian et al. χρησιμοποίησαν Multistate Face Models για να εντοπίσουν μόνιμα και μη-μόνιμα χαρακτηριστικά του προσώπου ξεχωριστά. Τα μόνιμα χαρακτηριστικά (χειίλη, μάτια φρύδια και

ζυγωματικά) εντοπίζονται χρησιμοποιώντας διαφορετικούς εντοπιστές. Χρησιμοποιείται η μέθοδος Canny edge detection για να εντοπιστούν μη-μόνιμα χαρακτηριστικά, όπως αυλάκια.

2.Object detection using Haar cascades

Η μέθοδος αυτή προτάθηκε το 2001 από τους Paul Viola και Michael Jones στο άρθρο “Rapid Object Detection using a Boosted Cascade of Simple Features” [3]. Είναι η μέθοδος που χρησιμοποιείται στα πρώτα στάδια της εργασίας προκειμένου να βρεθεί αρχικά το πρόσωπο και μετά το στόμα. Η μέθοδος αυτή βασίζεται σε τρεις βασικούς παράγοντες που συμβάλλουν στην επιτυχία της:

- 1) ένας νέος τρόπος αναπαράστασης μιας εικόνας, η “Integral Image”, μας επιτρέπει να υπολογίσουμε τα χαρακτηριστικά που χρησιμοποιούνται στον αλγόριθμό μας με μεγάλη ταχύτητα
- 2) ένας αλγόριθμος εκμάθησης, βασισμένος στον αλγόριθμο AdaBoost, ο οποίος διαλέγει ένα μικρό αριθμό από χαρακτηριστικά από ένα πολύ μεγαλύτερο σύνολο και τα χρησιμοποιεί για να εκπαιδεύσει πολύ ικανούς κατηγοριοποιητές
- 3) μία μέθοδος που συνδυάζει πολύπλοκους κατηγοριοποιητές σε ένα “cascade”, που επιτρέπει μεγάλη ποσότητα άχρηστης πληροφορίας να μην εξεταστεί και να επικεντρωθεί στα σημεία που περιέχουν πληροφορία με μεγαλύτερη πιθανότητα

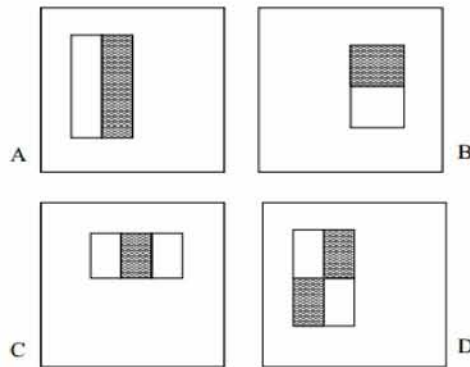
Χαρακτηριστικά

Ο αλγόριθμος κατηγοριοποιεί ανάλογα αν σε μία εικόνα υπάρχει ένα αντικείμενο ή όχι. Κριτήριο αποτελούν οι τιμές απλών χαρακτηριστικών (features) και όχι οι τιμές των pixels της εικόνας. Οι λόγοι είναι δύο:

- 1) τα απλά χαρακτηριστικά μπορούν να κωδικοποιήσουν τις γνώσεις που έχουμε βλέποντας διαισθητικά το πρόβλημα, με βάση την ανθρώπινη αντίληψη και εμπειρία
- 2) στο συγκεκριμένο αλγόριθμο η χρήση απλών χαρακτηριστικών τον επιταχύνει σε μεγάλο βαθμό

Τα απλά χαρακτηριστικά που χρησιμοποιούνται στον αλγόριθμο βασίζονται στις συναρτήσεις Haar, οι οποίες χρησιμοποιήθηκαν από τους Parageorgiou et al. Συγκεκριμένα, είναι τρεις κατηγορίες χαρακτηριστικών. Η τιμή ενός χαρακτηριστικού-δύο-τετραγώνων είναι η διαφορά στο άθροισμα των pixels που περιέχονται στα τετράγωνα. Η τιμή ενός χαρακτηριστικού-τριών-τετραγώνων

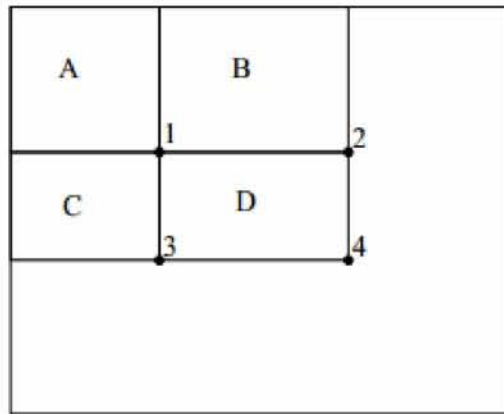
υπολογίζει το άθροισμα των εξωτερικών, αφαιρώντας το άθροισμα από το εσωτερικό. Η τιμή ενός χαρακτηριστικού-τεσσάρων-τετραγώνων είναι η διαφορά των διαγωνίων ζευγαριών τετραγώνων. Αν θεωρήσουμε μια περιοχή 24x24 pixels, τα χαρακτηριστικά Haar που μπορούμε να βρούμε είναι πάνω από 180.000! Το πρόβλημα αυτό λύνεται με την Integral Image.



Εικόνα 2.1: Παραδείγματα χαρακτηριστικών. Το άθροισμα των pixels στα άσπρα τετράγωνα αφαιρείται από το άθροισμα των pixels στα γκρι τετράγωνα. Οι εικόνες (A) και (B) έχουν χαρακτηριστικά-δύο-τετραγώνων, η εικόνα (C) χαρακτηριστικά-τριών-τετραγώνων και η εικόνα (D) χαρακτηριστικά-τεσσάρων-τετραγώνων.

Integral Image

Τα παραπάνω χαρακτηριστικά θα ήθελαν πολύ χρόνο για να υπολογιστούν ξεχωριστά το καθένα. Εκμεταλλευόμενοι τον επαναλαμβανόμενο υπολογισμό χαρακτηριστικών σε υπο-παράθυρα, προεπεξεργαζόμαστε τον πίνακα Integral Image. Στον πίνακα αυτό, το σημείο x,y περιλαμβάνει το άθροισμα των pixels πάνω και αριστερά του σημείου. Ένα παράδειγμα υπολογισμού δίνεται στην εικόνα 2.2. Χρησιμοποιώντας τον πίνακα αυτό, το άθροισμα τιμών pixels για κάθε χαρακτηριστικό-δύο-τετραγώνων μπορεί να υπολογιστεί σε τέσσερις πράξεις και η διαφορά σε οκτώ. Επειδή όμως τα τετράγωνα βρίσκονται δίπλα το ένα στο άλλο, χρειάζονται τελικά 6 πράξεις. Τα χαρακτηριστικά-τριών-τετραγώνων χρειάζονται οκτώ πράξεις και τα χαρακτηριστικά-τεσσάρων-τετραγώνων εννιά πράξεις. Από τα 180.000 χαρακτηριστικά όμως, ποια θα χρησιμοποιήσουμε εμείς;



Εικόνα 2.2: Το άθροισμα των pixels που περιέχονται στο τετράγωνο D μπορεί να υπολογιστεί με τέσσερις πράξεις. Η τιμή της Integral Image στο σημείο 1 είναι το άθροισμα των pixels στο τετράγωνο A. Η τιμή στο σημείο 2 είναι $A+B$, στο σημείο 3 είναι $A+C$, και στο σημείο 4 $A+B+C+D$. Το άθροισμα στο D μπορεί να υπολογιστεί ως $4+1-(2+3)$.

Adaboost

Στην απλή έκδοσή του, ο αλγόριθμος Adaboost χρησιμοποιείται για να βελτιώσει την απόδοση πολλών αδύναμων κατηγοριοποιητών. Οι Freund και Schapire απέδειξαν ότι το λάθος εκπαίδευσης του τελικού κατηγοριοποιητή, αποτέλεσμα των πολλών αδύναμων κατηγοριοποιητών, τείνει στο μηδέν εκθετικά με βάση τον αριθμό των γύρων. Στην υπόθεση που έκαναν οι Viola και Jones, η οποία προέκυψε απ' τα πειράματα, ένας πολύ μικρός αριθμός από χαρακτηριστικά μπορούν να συνδυαστούν με αποτέλεσμα έναν ισχυρό κατηγοριοποιητή. Η δυσκολία είναι στο να βρει κάποιος τα χαρακτηριστικά αυτά.

Στον αλγόριθμο Adaboost, κάθε αδύναμος κατηγοριοποιητής είναι ένας κατηγοριοποιητής που, χρησιμοποιώντας μόνο ένα χαρακτηριστικό, κατηγοριοποιεί τις εικόνες. Σε κάθε βήμα επιλέγουμε τον αδύναμο κατηγοριοποιητή που κατηγοριοποίησε καλύτερα, δηλαδή με τα λιγότερα λάθος αποτελέσματα, το σύνολο των εικόνων. Στην πράξη, κανένα χαρακτηριστικό από μόνο του δεν μπορεί να κατηγοριοποιήσει με καλά αποτελέσματα. Χαρακτηριστικά που επιλέχθηκαν στα αρχικά στάδια είχαν ποσοστά λάθους ανάμεσα σε 0.1 και 0.3. Χαρακτηριστικά που ο αλγόριθμος επέλεξε σε μετέπειτα στάδια είχαν ποσοστό λάθους μεταξύ 0.4 και 0.5. Αφού στον πρώτο γύρο επιλέξει το χαρακτηριστικό που είναι το πιο αποτελεσματικό, προετοιμάζει τις εικόνες που κατηγοριοποιήθηκαν λάθος από το χαρακτηριστικό αυτό ώστε να έχουν περισσότερο βάρος στον

επόμενο γύρο. Έτσι, το επόμενο χαρακτηριστικό θα διαλεχτεί έτσι ώστε να μπορέσει να κατηγοριοποιήσει σωστά κυρίως αυτά που το προηγούμενο έκανε λάθος. Με την απλή αυτή συλλογιστική, οι αδύναμοι κατηγοριοποιητές μπαίνουν σε τέτοια σειρά, ώστε να έχουν το καλύτερο δυνατό αποτέλεσμα. Η τελική κατηγοριοποίηση γίνεται με βάση το συνδυασμό από τα αποτελέσματα των αδύναμων κατηγοριοποιητών. Ο πλήρης αλγόριθμος ακολουθεί:

- Έχοντας εικόνες $(x_1, y_1), \dots, (x_n, y_n)$ όπου $y_i = 0, 1$ για αρνητικά και θετικά παραδείγματα αντίστοιχα.
- Αρχικοποιώ βάρη $w_{1,i} = \frac{1}{2m}, \frac{1}{2l}$ για $y_i = 0, 1$ αντίστοιχα, όπου m και l οι αριθμοί των αρνητικών και θετικών αντίστοιχα.
- Για $t = 1, \dots, T$:

1. Κανονικοποιώ τα βάρη,

$$w_{t,i} \leftarrow \frac{w_{t,i}}{\sum_{j=1}^n w_{t,j}}$$

έτσι ώστε w_t να είναι μια κατανομή πιθανότητας.

2. Για κάθε χαρακτηριστικό, j , εκπαιδεύω ένα κατηγοριοποιητή h_j που χρησιμοποιεί μόνο ένα χαρακτηριστικό. Το λάθος υπολογίζεται ως $\epsilon_j = \sum_i w_i |h_j(x_i) - y_i|$.
3. Διαλέγω τον κατηγοριοποιητή, h_t με το μικρότερο λάθος ϵ_t .
4. Ενημερώνω τα βάρη:

$$w_{t+1,i} = w_{t,i} \beta_t^{1-e_i}$$

όπου $e_i = 0$ αν το παράδειγμα x_i έχει κατηγοριοποιηθεί σωστά, $e_i = 1$ αλλιώς, και $\beta_t = \frac{\epsilon_t}{1-\epsilon_t}$.

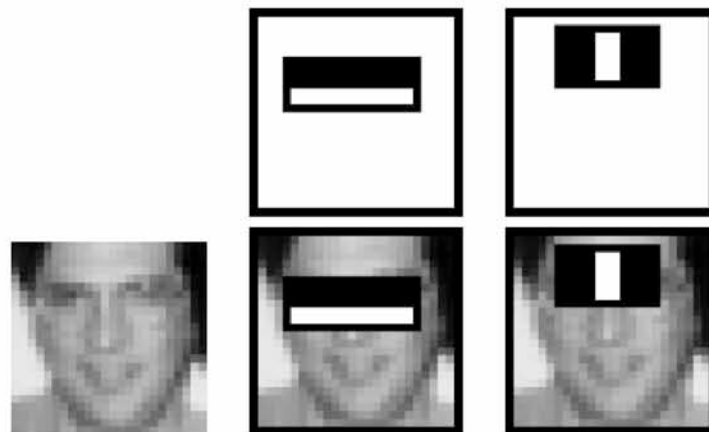
- Ο τελικός δυνατός κατηγοριοποιητής είναι:

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

$$\text{όπου } \alpha_t = \log \frac{1}{\beta_t}$$

Στην πράξη ο αλγόριθμος Adaboost, όταν χρησιμοποιήθηκε με 200 χαρακτηριστικά για εντοπισμό προσώπου, πέτυχε κατά 95%, με ποσοστό ψευδώς θετικό 1/14084. Αυτά τα αποτελέσματα θεωρούνται ικανοποιητικά, όχι όμως αρκετά καλά για μια εφαρμογή πραγματικού χρόνου. Όσο αφορά το κόστος υπολογισμού, ο κατηγοριοποιητής αυτός ήταν μάλλον ταχύτερος από αυτούς που μέχρι τώρα έχουν εκδοθεί, απαιτώντας 0.7 δευτερόλεπτα να εντοπίσει πρόσωπα σε μια εικόνα 384x288 pixels. Δυστυχώς ο καλύτερος τρόπος για να βελτιώσεις τα αποτελέσματα είναι να προσθέσεις χαρακτηριστικά, κάτι που αυξάνει άμεσα το υπολογιστικό κόστος.

Τα αρχικά χαρακτηριστικά που επιλέχθηκαν για τον εντοπισμό του προσώπου εξηγούνται άμεσα διαισθητικά. Το πρώτο χρησιμοποιεί το γεγονός ότι η περιοχή των ματιών είναι συχνά πιο σκούρη από την περιοχή της μύτης και των ζυγωματικών. Το δεύτερο χρησιμοποιεί το γεγονός ότι η περιοχή ανάμεσα στα μάτια είναι λιγότερο σκούρη από τα μάτια.



Εικόνα 2.3: Το πρώτο και δεύτερο χαρακτηριστικό που διαλέγει η Adaboost. Τα δύο αυτά χαρακτηριστικά παρουσιάζονται πάνω και εφαρμόζονται σε ένα πρόσωπο κάτω. Το πρώτο χαρακτηριστικό μετρά τη διαφορά στην ένταση της εικόνας μεταξύ της περιοχής των ματιών και της περιοχής πάνω απ' τα ζυγωματικά. Βασίζεται στο γεγονός ότι η περιοχή των ματιών είναι συνήθως πιο σκούρη απ' τα ζυγωματικά. Το δεύτερο χαρακτηριστικό συγκρίνει την διαφορά έντασης της εικόνας ανάμεσα στην περιοχή των ματιών και στην περιοχή πάνω από τη μύτη.

Attentional cascade

Ως cascade μπορούμε να ορίσουμε ένα δέντρο αποφάσεων, στο οποίο κάτι μπορεί να είναι Σωστό ή Λάθος. Το σωστό περνάει στο επόμενο στάδιο, ενώ το Λάθος κόβεται. Τελικά, ότι γίνει αποδεκτό

από το cascade θα πρέπει να έχει περάσει όλα τα στάδια ως Σωστό.

Προκειμένου να μειωθεί το υπολογιστικό κόστος και να υπάρχει καλύτερος εντοπισμός, οι Viola και Jones χρησιμοποιούν ένα cascade από κατηγοριοποιητές που έχουν δημιουργήσει με τον αλγόριθμο Adaboost. Το κλειδί είναι το γεγονός ότι μικρότεροι και αποτελεσματικότεροι κατηγοριοποιητές μπορούν να κατασκευαστούν, με σκοπό να αποκλείουν τις περισσότερες εικόνες που δεν περιέχουν πρόσωπο, ενώ δέχονται σχεδόν όλες τις εικόνες που το περιέχουν. Απλοί κατηγοριοποιητές διώχνουν την πλειοψηφία των υπο-παραθύρων που δεν έχουν πρόσωπο, ώστε αργότερα πιο σύνθετοι κατηγοριοποιητές με χαμηλότερα ψευδώς θετικά ποσοστά να ακολουθήσουν. Με βάση τον ορισμό που δώσαμε παραπάνω, για να γίνει αποδεκτό ότι σε ένα υπο-παραθύρο υπάρχει πρόσωπο, θα πρέπει να έχει γίνει αποδεκτό από όλους τους κατηγοριοποιητές του cascade. Στην απλή έκδοση του αλγορίθμου Adaboost, το όριο τοποθετείται με κριτήριο να υπάρχει χαμηλό ποσοστό λάθους. Εδώ το κριτήριο είναι να ελαχιστοποιηθούν τα ψευδώς αρνητικά. Έτσι χαμηλώνουμε το όριο. Αυτό επιτρέπει και κάποια ψευδώς θετικά να περάσουν, κάτι που δεν μας ενοχλεί ιδιαίτερα, γιατί πιθανότερα θα αποκλειστούν στις επόμενες φάσεις του cascade.

Για παράδειγμα, ένας κατηγοριοποιητής πρώτου σταδίου που λειτουργεί πολύ αποδοτικά μπορεί να κατασκευαστεί από ένα ισχυρό κατηγοριοποιητή δύο-χαρακτηριστικών, μειώνοντας το όριο ώστε να ελαχιστοποιηθούν τα ψευδώς αρνητικά. Το όριο μπορεί να τοποθετηθεί έτσι, ώστε να εντοπίζει 100% των προσώπων με ποσοστό ψευδώς θετικό 40%. Τα δύο αυτά χαρακτηριστικά βρίσκονται στην εικόνα 2.3. Ο υπολογισμός αυτού του κατηγοριοποιητή πραγματοποιείται με 60 εντολές μηχανής. Δύσκολα κάποιο άλλο φίλτρο θα απέρριπτε επιτυχώς τόσα πολλά παραδείγματα.

Η επιλογή του cascade γίνεται κατανοητή από το γεγονός ότι μέσα σε κάθε εικόνα η μεγάλη πλειοψηφία των υπο-παραθύρων δεν περιέχουν πρόσωπο. Έτσι είναι αποδοτικό να απορρίψουμε όσα περισσότερα τέτοια υπο-παραθύρα με το ελάχιστο υπολογιστικό κόστος.

Το cascade του αλγορίθμου εντοπισμού προσώπου έχει 38 στάδια με παραπάνω από 6000 χαρακτηριστικά. Παρ' όλα αυτά, επιτυγχάνει κατά μέσο όρο γρήγορα αποτελέσματα. Σε ένα δύσκολο σύνολο δεδομένων που περιέχει 507 πρόσωπα και 75 εκατομμύρια υπο-παραθύρα, τα πρόσωπα εντοπίζονται με κατά μέσο όρο 10 υπολογισμούς χαρακτηριστικών ανά υπο-παραθύρο. Αυτό είναι εφικτό, επειδή η μεγάλη πλειοψηφία των υπο-παραθύρων απορρίπτονται από το πρώτο ή το δεύτερο στάδιο. Σε έναν επεξεργαστή 700MHz Pentium 3, ο αλγόριθμος επεξεργάζεται μια περιοχή 384x288 pixels σε περίπου 0.067 δευτερόλεπτα. Αυτό είναι περίπου 15 φορές γρηγορότερο από τον αλγόριθμο Rowley-Baluja-Kanade και 600 φορές γρηγορότερο από τον αλγόριθμο Schneiderman-Kanade.

Χρησιμοποιώντας τον παραπάνω αλγόριθμο, απομονώνουμε από την εικόνα το πρόσωπο και από το πρόσωπο το στόμα. Στη συνέχεια, θα θέλαμε να κρατήσουμε τη “γραμμή” ανάμεσα στα δύο χείλη, ώστε να δούμε το σχήμα που έχει. Έτσι, μπορούμε να κάνουμε διάκριση ανάμεσα στο ουδέτερο ύφος ενός ανθρώπου, το χαμογελαστό, ή την έκπληξη. Προκειμένου να βρούμε τη γραμμή αυτή, θα χρησιμοποιήσουμε έναν edge detector, ώστε να εκμεταλλευτούμε τη σκιά που υπάρχει στην ένωση των δύο χειλιών. Πρώτα, πρέπει η εικόνα να μετατραπεί σε ασπρόμαυρη.


3.RGB to Grey

Υπάρχουν διαφορετικές μέθοδοι για να απεικονιστεί και να αποθηκευτεί ένα χρώμα στον υπολογιστή. Συγκεκριμένα, κάθε χρώμα μπορεί να δημιουργηθεί από συνδυασμούς των χρωμάτων κόκκινο, πράσινο, μπλε. Έτσι, κάθε συνδυασμός εντάσεων των τριών αυτών χρωμάτων μας δίνει ένα χρώμα άρα και κάθε pixel αποθηκεύεται ως ένας τέτοιος συνδυασμός. Στο γκρι χρώμα, δηλαδή στην απεικόνιση στην οποία υπάρχουν μόνο αποχρώσεις του γκρι, από το λευκό ως το μαύρο, υπάρχει μόνο μία τιμή αντί για τρεις και αυτή είναι η ένταση του συγκεκριμένου pixel. Η μετατροπή από RGB σε Grey γίνεται με διάφορες μεθόδους [5]:

- Η μέθοδος της φωτεινότητας υιοθετεί τον αλγόριθμο $(\max(R,G,B) + \min(R,G,B))/2$.
- Η μέθοδος του μέσου όρου απλά βρίσκει τον μέσο όρο των R,G,B, δηλαδή $(R+G+B)/3$.
- Η μέθοδος luminosity, αντί για το μέσο όρο, βάζει διαφορετικά βάρη σε κάθε χρώμα. Αυτό βασίζεται στο γεγονός ότι το ανθρώπινο μάτι είναι πιο ευαίσθητο στο πράσινο χρώμα, έτσι ο αλγόριθμος του δίνει μεγαλύτερο βάρος. Ο αλγόριθμος μετατροπής RGB to Grey της opencv χρησιμοποιεί τα παρακάτω βάρη:

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

Στις παρακάτω εικόνες φαίνεται η διαφορά των τριών αλγορίθμων:

| | |
|----------------------|---|
| Αρχική εικόνα |  |
| Μέθοδος φωτεινότητας |  |
| Μέθοδος μέσου όρου |  |
| Μέθοδος luminosity |  |

Εικόνα 3.1: Αλγόριθμοι RGB to Gray

Εφ' όσον έχουμε μετατρέψει την εικόνα μας σε ασπρόμαυρη, δηλαδή σε αποχρώσεις του γκρι, μπορούμε να προχωρήσουμε στον εντοπισμό της γραμμής του χείλους. Προκειμένου να εφαρμοστεί σωστά ο Canny edge detector που θα χρησιμοποιήσουμε, πρέπει πρώτα να περάσουμε την εικόνα από ένα low pass filter. Ο edge detector βρίσκει “απότομες” διαφορές ανάμεσα στα pixels. Έτσι, θα θέλαμε να εξομαλύνουμε τα απομονωμένα pixels που έχουν μεγάλη διαφορά με τους γείτονές τους, ώστε ο edge detector να μας δώσει ως αποτέλεσμα μόνο τη γραμμή στα χείλη και όχι θόρυβο. Για την εξομάλυνση αυτή χρησιμοποιούμε το Gaussian smoothing.

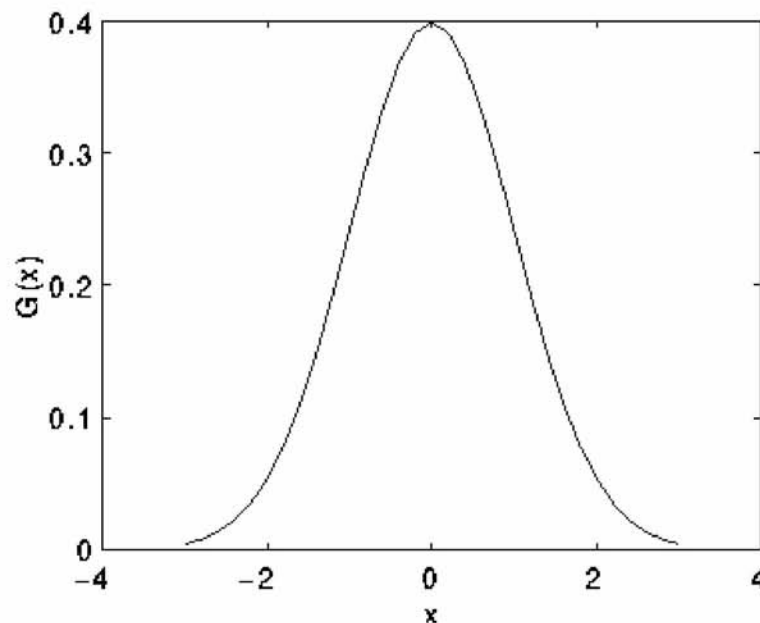
4. Gaussian smoothing

Ο Gaussian smoothing operator είναι ένας δύο διαστάσεων convolution operator, που χρησιμοποιείται για να “θολώσει” (blur) εικόνες και να απομακρύνει τις λεπτομέρειες και το θόρυβο. Μοιάζει με το mean filter, αλλά χρησιμοποιεί διαφορετικό kernel, που δημιουργεί το καμπανοειδές σχήμα της Gaussian κατανομής [6].

Η Gaussian κατανομή σε μία διάσταση έχει της εξής μορφή:

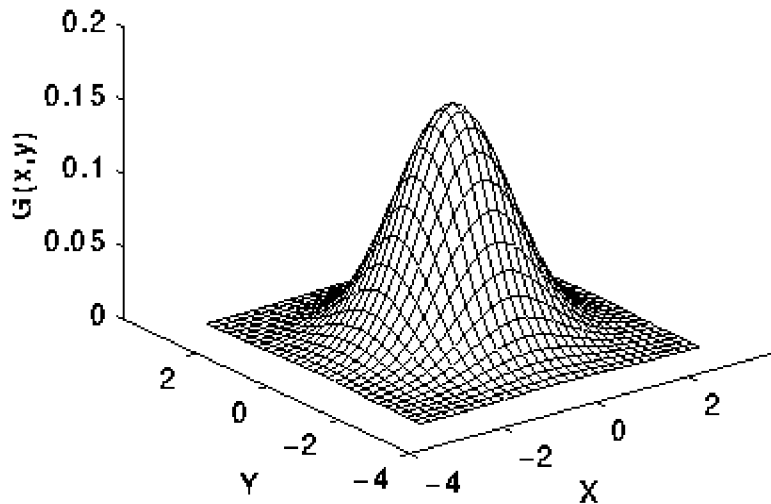
$$G(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}$$

όπου σ είναι η τυπική απόκλιση της κατανομής. Επίσης υποθέσαμε ότι η κατανομή έχει μέσο όρο μηδέν. Η κατανομή φαίνεται στην παρακάτω εικόνα.



Στις δύο διαστάσεις, η κατανομή είναι η εξής:

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$$



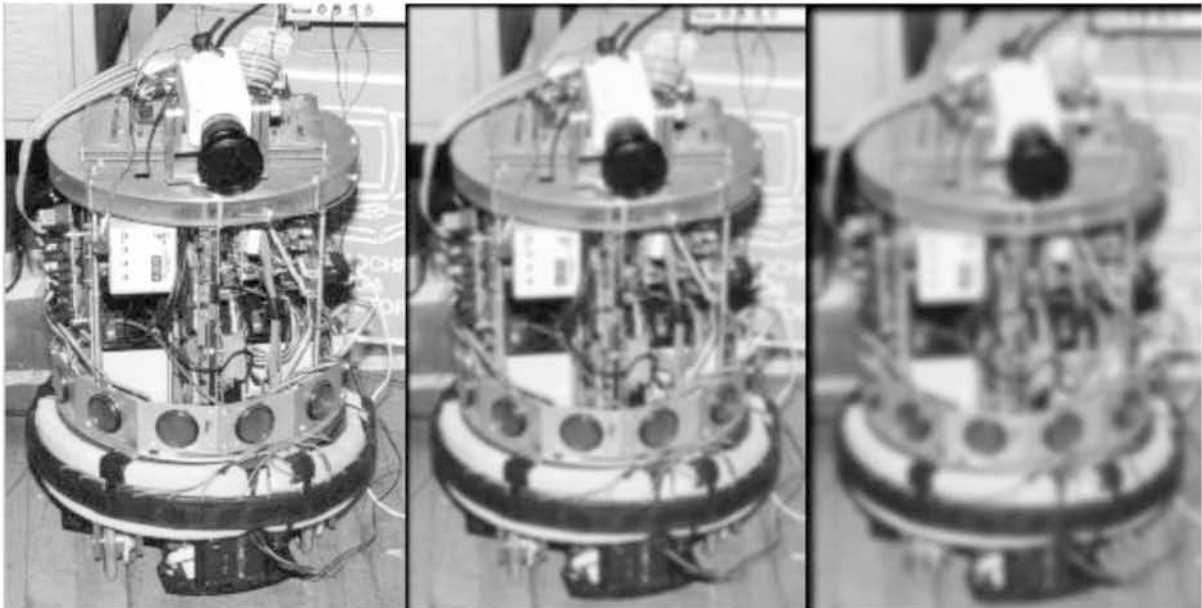
Ο σκοπός του Gaussian smoothing είναι να χρησιμοποιηθεί η μορφή της 2-D κατανομής πάνω στην εικόνα για να εξαλειφθεί ο θόρυβος και αυτό γίνεται μέσω της συνέλιξης. Εφ' όσον η εικόνα είναι ένα σύνολο από διακριτά pixels χρειάζεται να δημιουργήσουμε μια διακριτή προσέγγιση της συνάρτησης Gauss πριν τη συνέλιξη. Θεωρητικά, η συνάρτηση αυτή είναι μη μηδενική παντού, κάτι το οποίο θα απαιτούσε ένα άπειρο kernel συνέλιξης. Στην πράξη, δεν μας χρειάζεται τόση λεπτομέρεια. Στην παρακάτω εικόνα βλέπουμε ένα kernel που θα μπορούσαμε να χρησιμοποιήσουμε για τη συνέλιξη.

$$\frac{1}{273}$$

| | | | | |
|---|----|----|----|---|
| 1 | 4 | 7 | 4 | 1 |
| 4 | 16 | 26 | 16 | 4 |
| 7 | 26 | 41 | 26 | 7 |
| 4 | 16 | 26 | 16 | 4 |
| 1 | 4 | 7 | 4 | 1 |

Αφού έχουμε δημιουργήσει τον kernel, το Gaussian smooth μπορεί να πραγματοποιηθεί με τις γνωστές μεθόδους συνέλιξης. Στην πραγματικότητα, η συνέλιξη μπορεί να πραγματοποιηθεί αρκετά γρήγορα, αφού ο kernel μπορεί να χωριστεί σε δύο παράγοντες, x και y. Έτσι η δισδιάστατη συνέλιξη μπορεί να γίνει κάνοντας συνέλιξη πρώτα κατά x και μετά κατά y.

Ακολουθούν παραδείγματα χρήσης του αλγορίθμου με διαφορετική τυπική απόκλιση και μέγεθος kernel.



Εικόνα 4.1: Από αριστερά: 1. Αυθεντική εικόνα. 2. Επεξεργασμένη εικόνα με $\sigma = 1.0$ και μέγεθος *kernel* 5x5. 3. Επεξεργασμένη εικόνα με $\sigma = 2$ και μέγεθος *kernel* 9x9.

5.Canny edge detector

Ο Canny edge detector αναπτύχθηκε απ' τον John F. Canny το 1986. Χρησιμοποιεί ένα πολυεπίπεδο αλγόριθμο προκειμένου να εντοπίσει μια μεγάλη γκάμα ακμών. Σκοπός του Canny ήταν να δημιουργήσει τον ιδανικό αλγόριθμο εντοπισμού ακμών. Τα κριτήρια που χρησιμοποίησε ήταν [8]:

- μικρό ποσοστό λάθους: να βρίσκει δηλαδή μόνο πραγματικές ακμές
- καλός εντοπισμός: η απόσταση ανάμεσα στα pixels των ακμών που εντοπίστηκαν και στα πραγματικά pixels ακμών της εικόνας να είναι η ελάχιστη
- μια ακμή στην εικόνα να εντοπίζεται μόνο μία φορά

Τα βήματα που ακολουθεί ο αλγόριθμος Canny είναι τα εξής:

1. μείωση του θορύβου Επειδή ο αλγόριθμος είναι ευαίσθητος στην ύπαρξη θορύβου, φιλτράρουμε την εικόνα με ένα φίλτρο Gaussian smooth, όπως εξηγήσαμε παραπάνω.
2. προσδιορισμός της κλίσης της φιλτραρισμένης εικόνας. Αυτό πραγματοποιείται κάνοντας συνέλιξη με μάσκες παραγωγίσης Στη συγκεκριμένη εργασία, ο αλγόριθμος χρησιμοποιεί

$$G_x = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix}$$
$$G_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ +1 & +2 & +1 \end{bmatrix}$$

αυτές του Sobel:

- Η κατεύθυνση θ στρογγυλοποιείται σε μια από τις τέσσερις γωνίες (0,45, 90, 135).
3. καταστολή των μη μέγιστων τιμών [7]. Η μέθοδος αυτή αφαιρεί pixels που δεν θεωρούνται μέρη της ακμής. Έτσι μένουν μόνο οι λεπτές ακμές. Αυτό γίνεται ως εξής: έχοντας υπολογίσει την κλίση της εικόνας, βρίσκουμε το τοπικό μέγιστο του πλάτους για τα pixels. Για κάθε pixel, ελέγχουμε αν το πλάτος των pixels κάθετα στην κατεύθυνσή του είναι μικρότερο από το πλάτος του ίδιου. Αν είναι, τότε αποτελεί μέρος της ακμής.
 4. καταφλίωση υστέρησης. Το τελικό βήμα. Ο αλγόριθμος χρησιμοποιεί δύο τιμές καταφλίου (άνω και κάτω):
 1. αν η κλίση ενός pixel είναι μεγαλύτερη από την τιμή του άνω καταφλίου, το pixel

θεωρείται ακμή.

2. αν η κλίση είναι κάτω από το την τιμή του κάτω κατωφλίου, το pixel απορρίπτεται.
3. αν η κλίση είναι ανάμεσα στις τιμές των δύο κατωφλίων, τότε το pixel γίνεται αποδεκτό μόνο αν συνδέεται με pixel που έχει τιμή μεγαλύτερη του άνω κατωφλίου.



Εικόνα 5.1: Εικόνες πριν και μετά την εφαρμογή του αλγορίθμου Canny edge detector.

6.Cohn-Kanade-dataset

Ο αυτόματος εντοπισμός εκφράσεων και η αναγνώριση συναισθημάτων γίνεται όλο και σημαντικότερο πεδίο έρευνας. Χρησιμοποιείται σε πολύ σημαντικές εφαρμογές, όπως ασφάλεια, επικοινωνία ανθρώπου-υπολογιστή, ασφάλεια στην οδήγηση, ιατρικές εφαρμογές κ.α. Το γεγονός αυτό κάνει αναγκαία μια βάση δεδομένων με τέτοια χαρακτηριστικά, ώστε να μπορούν διαφορετικά ερευνητικά αποτελέσματα να γενικευθούν και να συγκριθούν. Η βάση δεδομένων Cohn-Kanade [9] δημιουργήθηκε το 2000 με αυτό το σκοπό. Από τότε, έχει γίνει μια από τις περισσότερο χρησιμοποιούμενες βάσεις δεδομένων για ανάπτυξη και αξιολόγηση αλγορίθμων. Το 2010 επεκτάθηκε με 27% περισσότερα αντικείμενα και 22% περισσότερες εικόνες, καταλήγοντας στη βάση δεδομένων CK+ [10].

Η βάση δεδομένων CK+ περιέχει αντικείμενα διαφορετικών ηλικιών, με διαφορετικό σχήμα προσώπου, χρώμα, υφή δέρματος, ηλικία, διαφορετική κουλτούρα. Στην αναγνώριση συναισθημάτων, οι παράγοντες αυτοί παίζουν μεγάλο ρόλο στη διαφορά της ίδιας έκφρασης από αντικείμενο σε αντικείμενο. Το άνοιγμα των ματιών διαφέρει για παράδειγμα ανάμεσα στους Ασιάτες και τους λαούς της βόρειας Ευρώπης. Ο τρόπος που κάποιος χαμογελάει διαφέρει ανάλογα με την κουλτούρα, την ηλικία. Ακόμα και η διάρκεια μιας έκφρασης διαφέρει από αντικείμενο σε αντικείμενο. Μια αξιόπιστη βάση δεδομένων πρέπει να έχει επαρκή αριθμό εκφράσεων από ανθρώπους όλων των παραπάνω συνδυασμών.

Η αρχική βάση δεδομένων CK περιείχε 486 σειρές εκφράσεων από 97 αντικείμενα. Η CK+ περιέχει 486 σειρές εκφράσεων από 123 αντικείμενα. Η λήψη έγινε με δύο κάμερες Panasonic AG-7500. Οι συμμετέχοντες είχαν ηλικία 18 – 50 χρονών, 69% γυναίκες, 81% Ευρωπαίοι και Αμερικανοί, 13% Αφροαμερικανοί και 6% άλλης καταγωγής. Κάθε σειρά εκφράσεων ξεκινάει και τελειώνει σε ουδέτερη έκφραση. Τα συναισθήματα που περιέχονται είναι: χαρά, έκπληξη, λύπη, αηδία, θυμός και φόβος. Οι εικόνες έχουν μέγεθος 640x480 ή 640x490 pixels. Η λήψη γίνεται ανά δευτερόλεπτο.

Η βάση δεδομένων CK+ θεωρείται ότι πληρεί τα κριτήρια μιας αξιόπιστης βάσης δεδομένων. Στόχος είναι να προστεθούν ακόμα περισσότερα αντικείμενα με μεγάλη γκάμα συναισθημάτων καθώς και να γίνει περαιτέρω προσπάθεια για να προστεθούν αυθόρμητες εκφράσεις, ή εκφράσεις που να τις προσεγγίζουν ακόμα καλύτερα.



Εικόνα 6.1: Παραδείγματα εικόνων απο τη βάση δεδομένων CK+.

7.Εξαγωγή χαρακτηριστικών στη βάση δεδομένων Cohn-Kanade

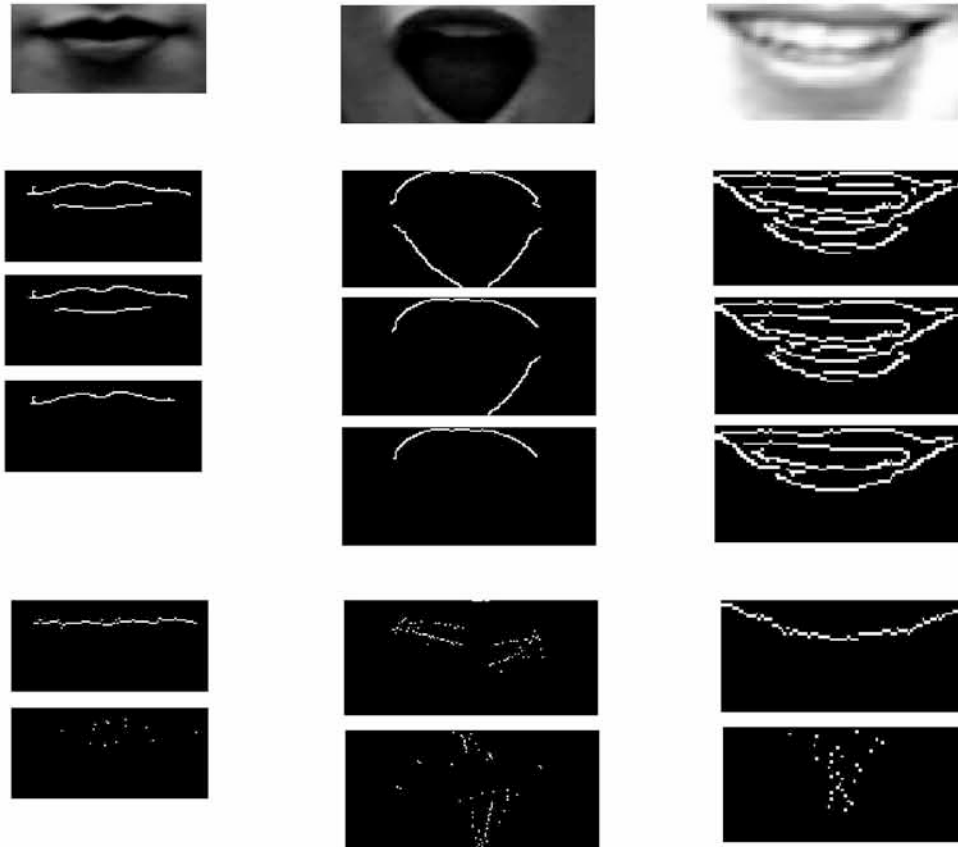
Από τη βάση δεδομένων έχουν χρησιμοποιηθεί μόνο τα αντικείμενα που είχαν εικόνες και στα τρία συναισθήματα που μας ενδιαφέρουν. Για κάθε μια από τις εικόνες, αρχικά εντοπίζουμε το πρόσωπο. Στη συνέχεια, περιοριζόμαστε στο κάτω 1/3 του προσώπου, όπου και ψάχνουμε το στόμα. Αφού το εντοπίσουμε, περνάμε την εικόνα από φίλτρο Gaussian smoothing, καταλήγοντας στην εικόνα του στόματος έτοιμη να περάσει από Canny edge detector. Τα βήματα φαίνονται παρακάτω για μία εικόνα από κάθε συναίσθημα.



Εικόνα 7.1: Οι εικόνες της πάνω σειράς είναι οι αυθεντικές εικόνες του πρώτου αντικειμένου, μία από κάθε συναίσθημα. Στην επόμενη σειρά περιορίζουμε την εικόνα στο 1/3 του προσώπου, ώστε να ψάξουμε γρηγορότερα για το στόμα. Στην τρίτη σειρά έχουμε εντοπίσει το στόμα, το οποίο περνάει από Gaussian smoothing. Το αποτέλεσμα, στο οποίο εφαρμόζεται ο Canny edge detector, βρίσκεται στην τελευταία σειρά.

Ο Canny edge detector εφαρμόζεται με τρία διαφορετικά κατώφλια, με βάση και την περιγραφή του αλγορίθμου που έγινε παραπάνω. Τα κατώφλια είναι: 70-210, 80-240, 100-300. Στη συνέχεια, ενώνουμε τα τρία αποτελέσματα σε ένα με τον εξής τρόπο: Για τις αντίστοιχες στήλες των τριών εικόνων, υπολογίζουμε τον μέσο όρο των υψών των pixels στις στήλες αυτές και δημιουργούμε μια καινούρια εικόνα με μόνο ένα στοιχείο σε κάθε στήλη, με ύψος τον μέσο όρο. Αντίστοιχα και κατά

γραμμές. Έτσι καταλήγουμε σε δύο εικόνες για κάθε εικόνα εισόδου, μία με τον μέσο όρο των pixels των τριών εικόνων στον άξονα x και μία με τον μέσο όρο των pixels των τριών εικόνων στον άξονα y . Παρακάτω φαίνεται το αποτέλεσμα για τις τρεις εικόνες που δείξαμε.

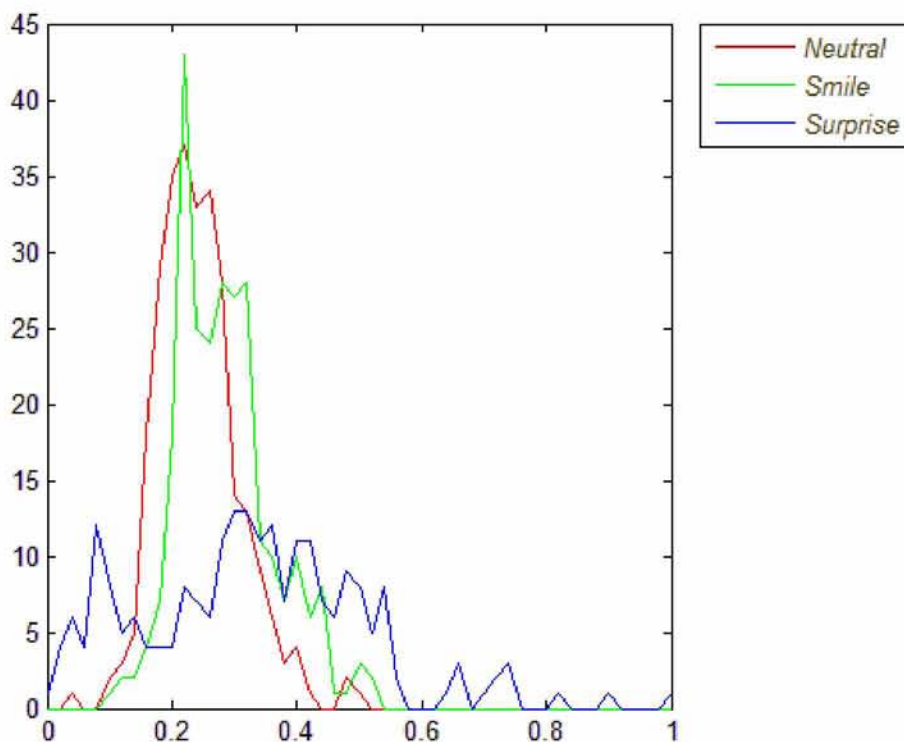


Εικόνα 7.2: Οι εικόνες της πρώτης σειράς είναι το αποτέλεσμα του Gaussian blur. Στη δεύτερη σειρά είναι τα αποτελέσματα μετά τον αλγόριθμο Canny, με διαφορετικά κατώφλια. Στην τρίτη σειρά είναι ο μέσος όρος των παραπάνω τριών εικόνων στους άξονες x και y .

8.Χαρακτηριστικά

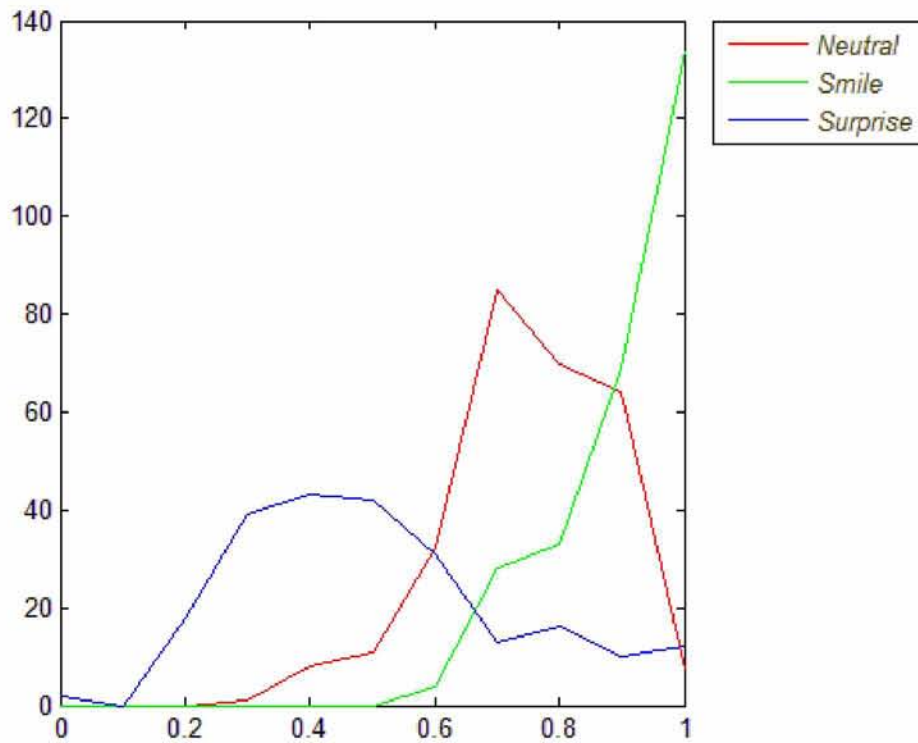
Κατά τη διάρκεια των πειραμάτων δοκιμάστηκαν διάφορα χαρακτηριστικά, ξεκινώντας με τη λογική ποια από αυτά διαχωρίζουν καλύτερα τις τρεις κατηγορίες συναισθημάτων. Αφού εξάγουμε τα χαρακτηριστικά, ένας τρόπος για να δούμε την αποδοτικότητα του κάθε χαρακτηριστικού είναι το ιστόγραμμα. Στο ιστόγραμμα κατανέμονται οι τιμές του χαρακτηριστικού για κάθε εικόνα σε ένα γράφημα, από την ελάχιστη ως τη μέγιστη, ανάλογα με τον αριθμό που κάθε τιμή επαναλήφθηκε. Έτσι μπορούμε να δούμε αν οι περιοχές συγκέντρωσης των τιμών του χαρακτηριστικού για κάθε συναίσθημα είναι διαφορετικές ή κοινές. Στη δεύτερη περίπτωση το χαρακτηριστικό δεν κάνει καλό διαχωρισμό. Στην πρώτη περίπτωση, το χαρακτηριστικό διαχωρίζει ικανοποιητικά. Ο συνδυασμός των σωστών χαρακτηριστικών θα μας δώσει το καλύτερο αποτέλεσμα. Τα χαρακτηριστικά στα οποία καταλήξαμε είναι εννιά.

1. Μέσος όρος ύψους στον άξονα y. Η έκπληξη σε αυτό το χαρακτηριστικό διαχωρίζεται από τα άλλα δύο συναισθήματα, λόγω του στρογγυλού σχήματος που έχει. Αυτό μειώνει τις τιμές στο μέσο όρο του ύψους δίνοντας μια πιο ομοιόμορφη κατανομή στο ιστόγραμμα. Αντίθετα τα άλλα δύο συναισθήματα έχουν χαμηλές τιμές.



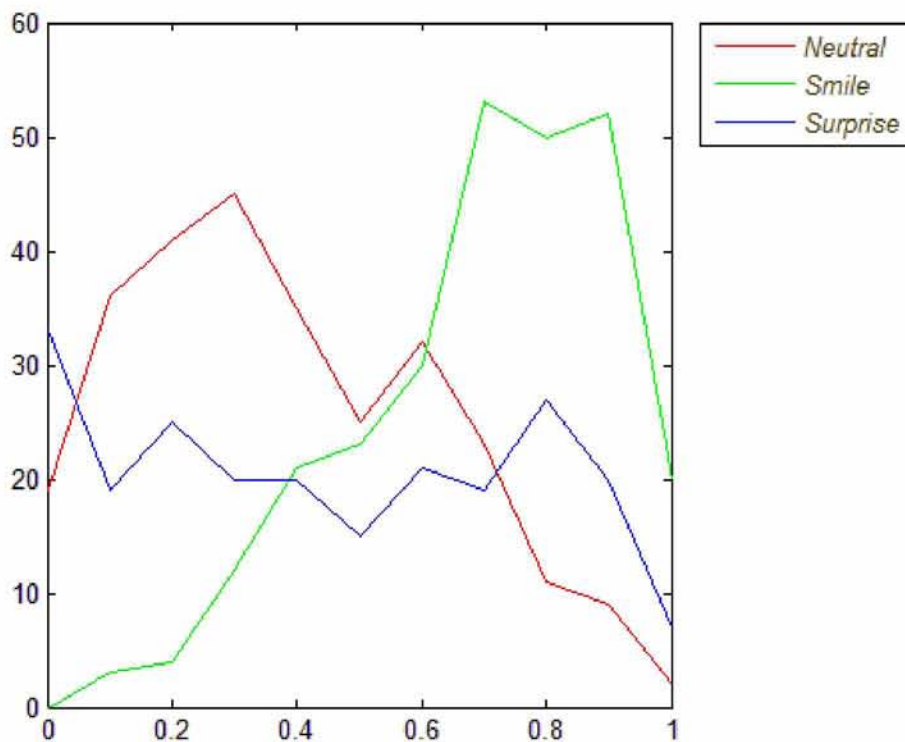
Εικόνα 8.1: Χαρακτηριστικό 1: Μέσος όρος ύψους στον άξονα y.

2. Αριθμός στηλών με τουλάχιστον ένα στοιχείο στον άξονα x. Παρατηρούμε ότι το χαμόγελο συνήθως πιάνει όλο στόμα στο πλάτος της εικόνας. Το ουδέτερο συναίσθημα συνήθως το 80%, ενώ η έκπληξη, και πάλι επειδή έχει στρόγγυλο σχήμα έχει το σύνολο των τιμών του χαρακτηριστικού στη μέση του άξονα.



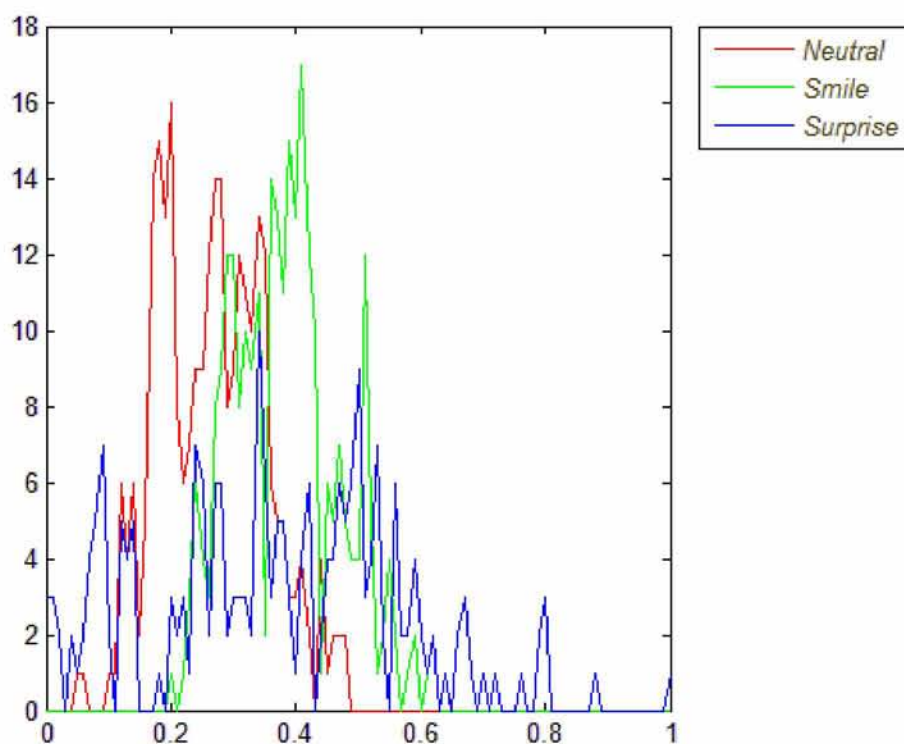
Εικόνα 8.2: Χαρακτηριστικό 2: Αριθμός στηλών με τουλάχιστον ένα στοιχείο στον άξονα x.

3. Ποσοστό των στηλών που το ύψος τους είναι μεγαλύτερο από το ύψος του μεσαίου στοιχείου της εικόνας. Το χαρακτηριστικό αυτό χρησιμοποιείται για να διαχωρίσουμε το χαμόγελο από τα άλλα δύο συναισθήματα. Το χαμόγελο συνήθως έχει σχήμα καμπύλης με το μεσαίο σημείο να είναι χαμηλότερα από τα υπόλοιπα, όπως φαίνεται και στο παρακάτω σχήμα.



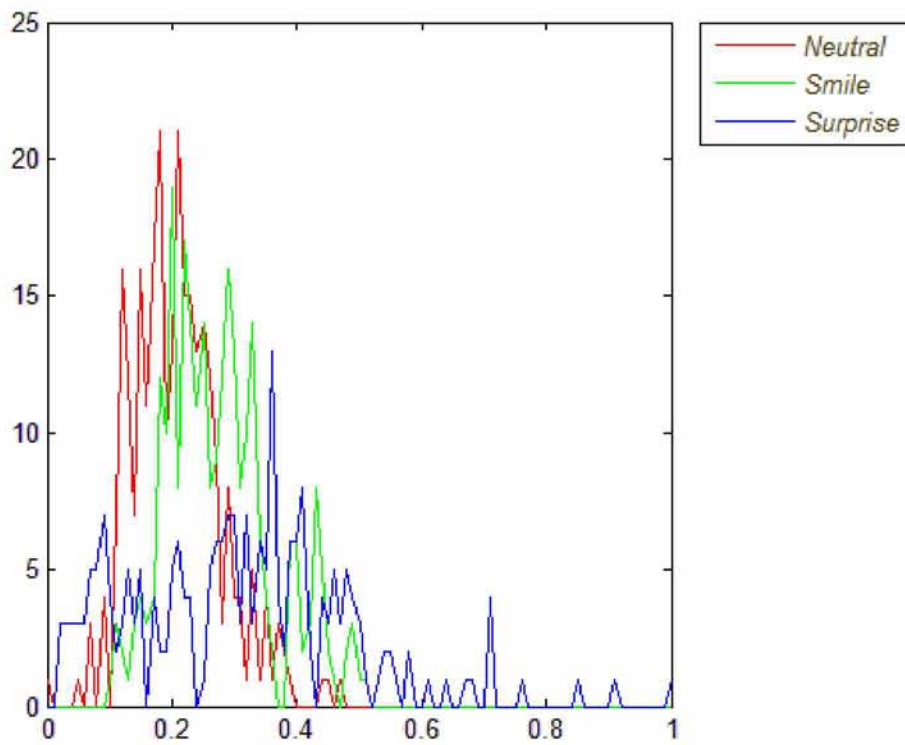
Εικόνα 8.3: Χαρακτηριστικό 3: Ποσοστό των στηλών που το ύψος τους είναι μεγαλύτερο από το ύψος του μεσαίου στοιχείου της εικόνας.

4. Μέσος όρος ύψους στο δεύτερο 1/3 του άξονα x. Από τα πειράματα παρατηρήθηκε ότι το χαμόγελο και το ουδέτερο συναίσθημα έχουν τις ίδιες τιμές στο δεύτερο 1/3 του άξονα x, ενώ η έκπληξη είναι περισσότερο ομοιόμορφη.



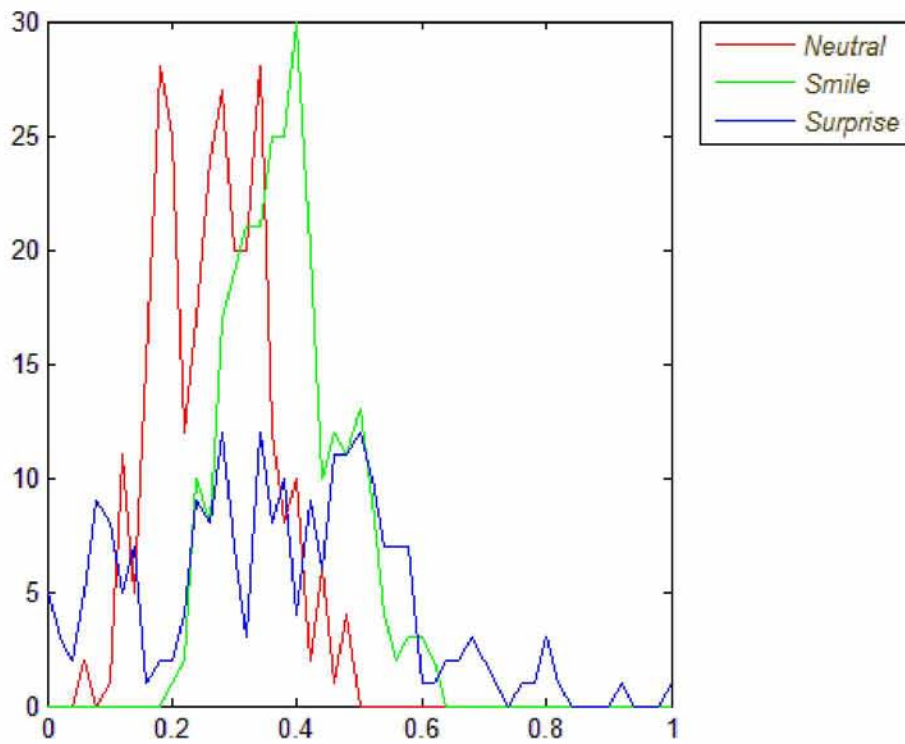
Εικόνα 8.4: Χαρακτηριστικό 4: Μέσος όρος ύψους στο δεύτερο 1/3 του άξονα x.

5. Μέσος όρος ύψους στον άξονα y με χρήση βαρών. Σε αυτή την περίπτωση κάθε στήλη έχει διαφορετικό βάρος στον υπολογισμό του μέσου όρου, ανάλογα με τα pixels που έχουν προστεθεί από τις τρεις εικόνες του Canny edge detector.



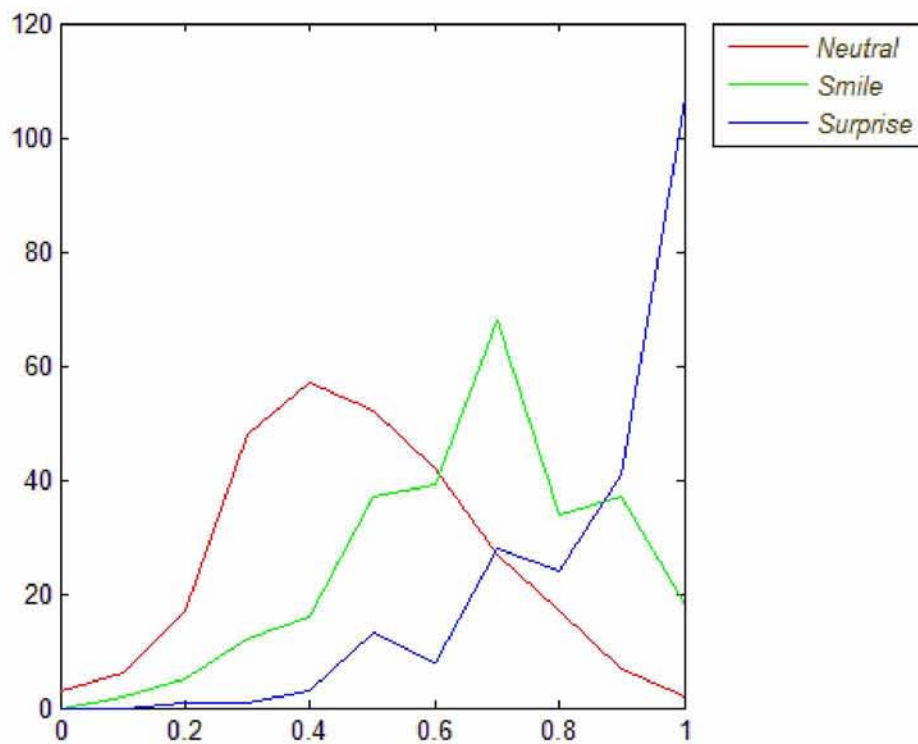
Εικόνα 8.5: Χαρακτηριστικό 5: Μέσος όρος ύψους στον άξονα y με βάρη.

6. Μέσος όρος ύψους στο δεύτερο 1/3 του άξονα x με βάρη. Όμοια με παραπάνω.



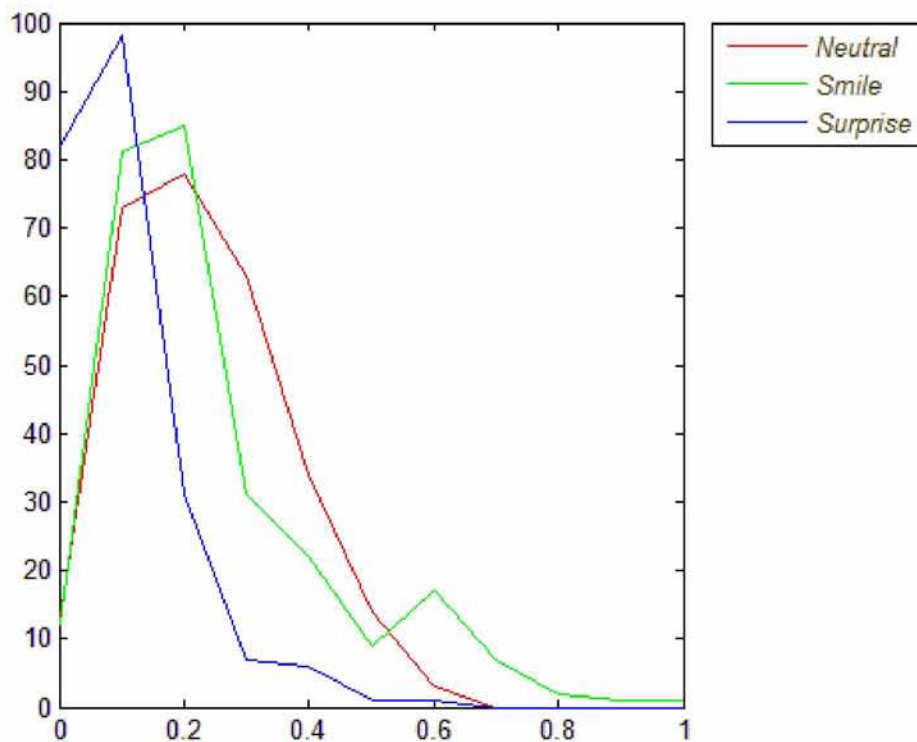
Εικόνα 8.6: Χαρακτηριστικό 6: Μέσος όρος ύψους στο δεύτερο 1/3 του άξονα x με βάρη.

7. Αριθμός γραμμών με τουλάχιστον ένα στοιχείο στον άξονα y . Το ουδέτερο συναίσθημα μοιάζει συνήθως με ευθεία γραμμή παράλληλη με τον άξονα x . Αυτό έχει ως συνέπεια να έχει πολύ λίγες γραμμές που να έχουν τουλάχιστον ένα στοιχείο. Η έκπληξη αντίθετα ως κύκλος προβάλλεται στον άξονα y με στοιχεία στην πλειοψηφία των γραμμών. Το χαμόγελο βρίσκεται ενδιάμεσα.



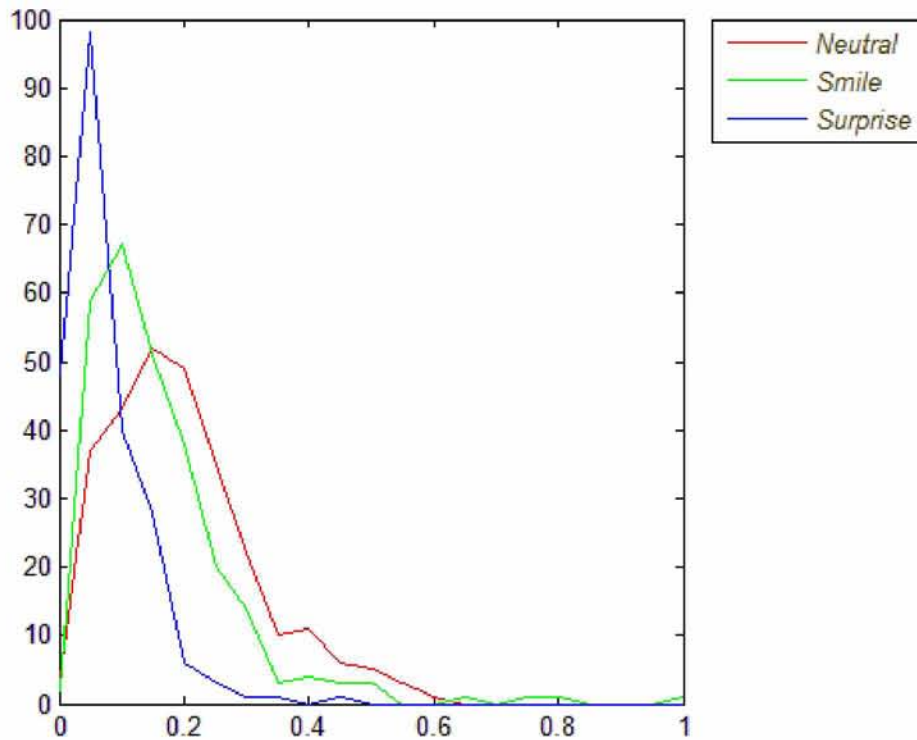
Εικόνα 8.7: Χαρακτηριστικό 7: Αριθμός γραμμών με τουλάχιστον ένα στοιχείο στον άξονα y .

8. Διασπορά στο μέσο όρο κατά τον άξονα y . Παρατηρείται από τα πειράματα ότι η έκπληξη έχει τη μικρότερη διασπορά. Αυτό και πάλι εξηγείται από το γεγονός ότι στο μέσο όρο των γραμμών, ο κύκλος μετατρέπεται σε μια “ευθεία” παράλληλη στον άξονα y στο κέντρο της εικόνας. Μεγαλύτερη διασπορά έχει το ουδέτερο συναίσθημα. Το χαμόγελο βρίσκεται ενδιάμεσα.



Εικόνα 8.8: Χαρακτηριστικό 8: Διασπορά στο μέσο όρο στον άξονα y .

9. Διασπορά στο μέσο όρο κατά τον άξονα y με βάρη.



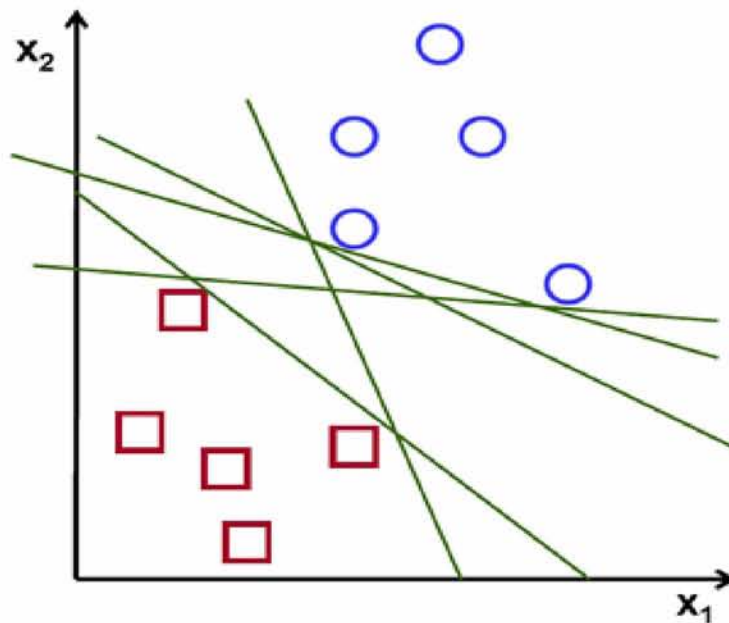
Εικόνα 8.9: Χαρακτηριστικό 9: Διασπορά στο μέσο όρο κατά τον άξονα y με βάρη.

Τα χαρακτηριστικά αυτά χρησιμοποιούνται αρχικά στη φάση εκπαίδευσης, όπου ο κατάλληλος συνδυασμός τους μας δίνει τα όρια που διαχωρίζουν τις τρεις κλάσεις – συναισθήματα. Στη συνέχεια, για κάθε εικόνα εισόδου βρίσκουμε τα χαρακτηριστικά αυτά και με βάση τα όρια σε ποια κλάση ανήκει. Η δουλειά αυτή πραγματοποιείται με τους ταξινομητές.

9.Support Vector Machines

Ο SVM είναι ένας ταξινομητής που καθορίζεται από ένα υπερεπίπεδο διαχωρισμού [11]. Με άλλα λόγια, έχοντας δεδομένα εκπαίδευσης (εποπτευόμενη μάθηση), ο αλγόριθμος παράγει ένα βέλτιστο υπερεπίπεδο που ταξινομεί νέα δεδομένα. Το πώς παράγεται το βέλτιστο υπερεπίπεδο γίνεται κατανοητό με το παρακάτω παράδειγμα:

Για ένα γραμμικά διαχωρίσιμο σύνολο δεδομένων σημείων δύο διαστάσεων, πρέπει να βρούμε μια διαχωριστική ευθεία γραμμή:



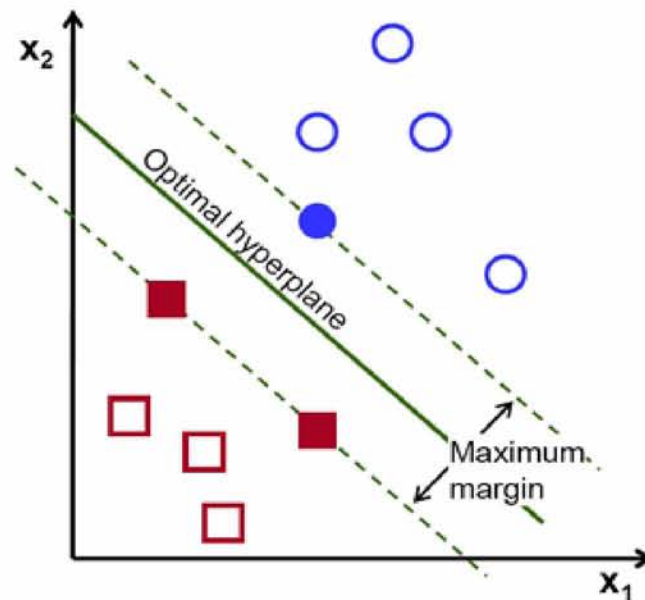
Εικόνα 9.1: Το παράδειγμα αυτό αφορά ευθείες και σημεία στο Καρτεσιανό σύστημα αντί για υπερεπίπεδα και διανύσματα σε χώρο μεγαλύτερων διαστάσεων. Αυτό γίνεται για λόγους κατανόησης, με το ίδιο σκεπτικό λειτουργεί ο αλγόριθμος και για περισσότερες διαστάσεις.

Στην παραπάνω εικόνα υπάρχουν παραπάνω από μία ευθείες που ικανοποιούν την αρχική ερώτηση. Ο SVM λύνει το παραπάνω πρόβλημα με το εξής κριτήριο:

Μια ευθεία δεν είναι καλή αν περνάει πολύ κοντά στα σημεία, επειδή θα είναι ευαίσθητη στο θόρυβο και η ταξινόμηση δεν θα γενικεύεται σωστά. Έτσι ο στόχος είναι να βρεθεί η ευθεία που περνάει όσο μακρύτερα γίνεται απ' όλα τα σημεία.

Ο αλγόριθμος SVM μεγιστοποιεί την απόσταση μεταξύ των διανυσμάτων υποστήριξης και ενός υπερεπιπέδου διαχωρισμού. Τα διανύσματα υποστήριξης είναι τα δεδομένα εκπαίδευσης που

βρίσκονται πιο κοντά στο υπερεπίπεδο και καθορίζουν το περιθώριό του (margin). Ο αλγόριθμος μεγιστοποιεί το περιθώριο αυτό.



Εικόνα 9.2: Το βέλτιστο υπερεπίπεδο.

Το υπερεπίπεδο ορίζεται ως:

$$f(x) = \beta_0 + \beta^T x,$$

το βέλτιστο υπερεπίπεδο μπορεί να αναπαρασταθεί με άπειρους διαφορετικούς τρόπους κλιμακώνοντας το διάνυσμα βαρών και το bias. Η αναπαράσταση που επιλέγουμε είναι:

$$|\beta_0 + \beta^T x| = 1$$

Όπου το x αναπαριστά τα διανύσματα υποστήριξης. Στη συνέχεια βρίσκουμε την απόσταση μεταξύ ενός σημείου x και του υπερεπιπέδου:

$$\text{distance} = \frac{|\beta_0 + \beta^T x|}{\|\beta\|}.$$

Το περιθώριο, M , ισούται με τη διπλάσια απόσταση από τα κοντινότερα δεδομένα:

$$M = \frac{2}{\|\beta\|}$$

Το πρόβλημα μεγιστοποίησης του M είναι ισοδύναμο με το πρόβλημα της ελαχιστοποίησης μιας

συνάρτησης $L(\beta)$ που υπόκειται σε κάποιους περιορισμούς:

$$\min_{\beta, \beta_0} L(\beta) = \frac{1}{2} \|\beta\|^2 \text{ υπόκειται σε } y_i(\beta^T x_i + \beta_0) \geq 1 \quad \forall i,$$

όπου y το διάνυσμα με τις κατηγορίες των δεδομένων εκπαίδευσης.

Αυτό είναι ένα πρόβλημα βελτιστοποίησης που λύνεται χρησιμοποιώντας πολλαπλασιαστές Lagrange, για να βρούμε το διάνυσμα βαρών και το bias του βέλτιστου υπερεπιπέδου.

Αποτελέσματα

Χρησιμοποιώντας 9 χαρακτηριστικά, τα οποία κανονικοποιούμε βρίσκοντας το ελάχιστο και το μέγιστο στοιχείο, χωρίσαμε τη βάση δεδομένων σε training και test set, με ποσοστό 75% και 25% αντίστοιχα και κάναμε τα εξής πειράματα:

- Leave-*one*-user-out cross-validation (LOOCV), όπου training set θεωρείται όλο το training set εκτός ενός αντικειμένου, και test set το αντικείμενο αυτό (user-independent). Επαναλαμβάνεται για όλα τα αντικείμενα. Έγινε ξεχωριστό πείραμα για κάθε εικόνα χωριστά και για κάθε συναίσθημα χωριστά (βάση πλειοψηφίας). Στην πρώτη περίπτωση, κατηγοριοποιούμε το συναίσθημα μόνο με βάση την εικόνα εισόδου. Στη δεύτερη περίπτωση, αποφασίζουμε για κάθε frame του συγκεκριμένου συναισθήματος και κατηγοριοποιούμε συνολικά το σύνολο των frames με την απόφαση που πάρθηκε για την πλειοψηφία.
- Έλεγχος στο test set για κάθε εικόνα και για κάθε συναίσθημα.

Τα ίδια πειράματα εφαρμόσαμε και με αλγόριθμο ταξινόμησης τον Nearest Neighbor. Λόγω απλότητας του αλγορίθμου τα αποτελέσματα (όπως περιμέναμε) ήταν χειρότερα του SVM.

| Πειράματα | Support Vector Machine | Nearest Neighbor |
|---|------------------------|------------------|
| Leave- <i>one</i> -out cross-validation (LOOCV) για κάθε εικόνα | 86.20% | 80.98% |
| Leave- <i>one</i> -out cross-validation (LOOCV) για κάθε συναίσθημα | 87.50% | 83.33% |
| Test set για κάθε εικόνα | 86.80% | 75.35% |
| Test set για κάθε συναίσθημα | 92.15% | 76.47% |

State of the art

Το σύστημα με το οποίο συγκρίνουμε τα αποτελέσματα εντοπίζει επτά συναισθήματα αντί για τρία. Βασίζεται στα Active Appearance Models (AAM), ενώ χρησιμοποιεί δύο ομάδες χαρακτηριστικών (similarity-normalized shape features & canonical appearance features). Τα αποτελέσματα προέρχονται από το συνδυασμό των δύο ομάδων, ενώ η κατηγοριοποίηση γίνεται με SVM. Από τους ίδιους προτείνεται η μέθοδος Leave-one-out cross-validation. Το αποτέλεσμα τους είναι 82.85% [1] για επτά συναισθήματα. Συγκεκριμένα:

- Θυμός: 75.00%
- Αηδία: 94.74%
- Φόβος: 65.22%
- Χαρά: 100%
- Λύπη: 68%
- Έκπληξη: 77.09%
- Ουδέτερο: 100%

10.Χρόνος εκτέλεσης

Ο συνολικός χρόνος εκτέλεσης υπολογίζεται από τη στιγμή που δίνεται η εικόνα εισόδου στο πρόγραμμα, μέχρι τη στιγμή που μας δίνεται ως έξοδος η κατηγορία συναισθήματος του αντικειμένου. Έχουμε μετρήσει και το χρόνο για κάθε συνάρτηση ξεχωριστά. Ο συνολικός χρόνος κατά μέσο όρο υπολογίζεται 0.095 seconds, δηλαδή 95ms. Συνεπώς υπολογίζονται περίπου 10.5 frames/second. Αξιοσημείωτο είναι ότι το 97% περίπου του χρόνου αυτού ξοδεύεται στους αλγόριθμους εντοπισμού του προσώπου και στη συνέχεια του στόματος. Παρακάτω ακολουθεί πίνακας με το μέσο όρο χρόνου ανά συνάρτηση:

| Συνάρτηση | Χρόνος σε ms |
|-------------------------------|--------------|
| -detectFacialFeatures | 95.2 |
| --Faces HaarDetect algorithm | 79.8 |
| --detectMouth | 15.4 |
| ---mouth Haardetect algorithm | 12.1 |
| ---cvtColor | 0.21 |
| ---cvSmooth | 0.43 |
| ---cvCanny x3 | 3.12 |
| ---featureExct | 2.9 |

11. Συμπεράσματα και μελλοντική εργασία

Η υπάρχουσα εργασία εντοπίζει αποτελεσματικά το πρόσωπο και το στόμα ενός ανθρώπου. Το ποσοστό επιτυχίας στην κατηγοριοποίηση είναι ικανοποιητικό, ενώ είναι σημαντικό ότι γενικεύεται, εξαιτίας της χρήσης της βάσης δεδομένων CK+ που έχει αντικείμενα διαφορετικών φύλων, πολιτισμών, ηλικίας. Ο χρόνος επίσης θεωρείται ικανοποιητικός, αφού μπορεί να εντοπίσει το συναίσθημα ενός ανθρώπου παραπάνω από δέκα φορές το δευτερόλεπτο, δεδομένου ότι η ανθρώπινη έκφραση δύσκολα αλλάζει παραπάνω από τρεις φορές το δευτερόλεπτο.

Το βασικό σημείο που θα μπορούσε να ενισχύσει την ικανότητα για σωστή κατηγοριοποίηση του αλγορίθμου είναι να συλλέγεται ταυτόχρονα πληροφορία και από άλλα μέρη του προσώπου, όπως τα μάτια, τα φρύδια, τα ζυγωματικά. Ακόμα θα μπορούσε να συνδυαστεί με μεθόδους όπως η Active Appearance Model (AAM) [12] και να δίνουν συνδυασμένα αποτελέσματα. Φυσικά θα μπορούσε να δοκιμαστεί με άλλες βάσεις δεδομένων ώστε να γενικεύεται ακόμα περισσότερο. Τέλος, θα μπορούσαν στις κατηγορίες συναισθημάτων να μουν και άλλα συναισθήματα, όπως θυμός, απογοήτευση, ενόχληση, φόβος.

Βιβλιογραφία

[1] Bettadapura, Vinay. "Face expression recognition and analysis: the state of the art." *arXiv preprint arXiv:1203.6722* (2012).

[2]

http://docs.opencv.org/trunk/doc/py_tutorials/py_objdetect/py_face_detection/py_face_detection.html

[3] Viola, P.; Jones, M., "Rapid object detection using a boosted cascade of simple features," Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on , vol.1, no., pp.I-511,I-518 vol.1, 2001.

[4] http://www.robots.ox.ac.uk/~az/lectures/cv/adaboost_matas.pdf.

[5] <http://www.johndcook.com/blog/2009/08/24/algorithms-convert-color-grayscale/>.

[6] <http://homepages.inf.ed.ac.uk/rbf/HIPR2/gsmooth.htm>.

[7] Σκολαρίκης Μιχαήλ, "Οπτική αναγνώριση και παρακολούθηση αντικειμένων με εξαγωγή χαρακτηριστικών σχήματος σε περιβάλλον κίνησης ρομπότ", Διπλωματική εργασία, Τμήμα Ηλεκτρολόγων και Μηχανικών Ηλεκτρονικών Υπολογιστών, Αριστοτέλειο Πανεπιστήμιο, 2011.

[8] http://docs.opencv.org/doc/tutorials/imgproc/imgtrans/canny_detector/canny_detector.html.

[9] Kanade, T.; Cohn, J.F.; YingLi Tian, "Comprehensive database for facial expression analysis," Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on , vol., no., pp.46,53, 2000.

[10] Lucey, P.; Cohn, J.F.; Kanade, T.; Saragih, J.; Ambadar, Z.; Matthews, I., "The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression," Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on , vol., no., pp.94,101, 13-18 June 2010.

[11] A Tutorial on Support Vector Machines for Pattern Recognition Christopher J.C. Burges, "A

Tutorial on Support Vector Machines for Pattern Recognition,” *Data Mining and Knowledge Discovery*, vol. 2, pp. 121-167, 1998.

[12] Cootes, Timothy F., Gareth J. Edwards, and Christopher J. Taylor, "Active appearance models," *Computer Vision—ECCV'98*. Springer Berlin Heidelberg, pp. 484-498, 1998.