



Τμήμα Μηχανικών Πληροφορικής, Υπολογιστών και
Τηλεπικοινωνιών
Πρόγραμμα Μεταπτυχιακών Σπουδών στη Ρομποτική

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Αναγνώριση εκφράσεων προσώπου με χρήση deep neural networks και του ενεργού μοντέλου εμφάνισης

Βαΐα Ζαχαράκη Α.Μ.:10

Σέρρες, Σεπτέμβριος 2020
Διεθνές Πανεπιστήμιο της Ελλάδος

Εργασία που υποβλήθηκε στο
Πρόγραμμα Μεταπτυχιακών Σπουδών στη Ρομποτική,
του Διεθνούς Πανεπιστημίου της Ελλάδος,
για τη μερική εκπλήρωση υποχρεώσεων για το Δίπλωμα Ειδίκευσης στη Ρομποτική

Επιβλέπων καθηγητής: Αθανάσιος Νικολαΐδης

Ευχαριστίες

Με την ολοκλήρωση αυτής της εργασίας θα ήθελα να εκφράσω τις ευχαριστίες μου στους ανθρώπους που βοήθησαν στην πραγματοποίησή της. Αρχικά στον καθηγητή μου τον κύριο Αθανάσιο Νικολαΐδη για την υπομονή και την καλή συνεργασία. Έπειτα την αδερφή μου και την οικογένειά μου για την συμβολή τους και στήριξή τους καθ' όλη τη διάρκεια συγγραφής της μεταπτυχιακής εργασίας. Τέλος, θα ήθελα να ευχαριστήσω τους φίλους μου που με στήριξαν όλο αυτό το διάστημα.

*Βαΐα Ζαχαράκη
Σέρρες, Σεπτέμβριος 2020*

Περίληψη

Η εργασία αυτή αποσκοπεί στην ανάλυση των εκφράσεων ενός συνόλου δεδομένων προσώπου από τον πραγματικό κόσμο. Για να λάβουμε τις αλλαγές στο σχήμα και την εμφάνιση του προσώπου σε μια εικονοσειρά, έχει αναπτυχθεί μια έκδοση του AAM (Ενεργό μοντέλο εμφάνισης). Αυτό το μοντέλο τροφοδοτείται στη συνέχεια στα προσφάτως δημοφιλή βαθιά νευρωνικά δίκτυα για να επιτευχθεί καλή κατηγοριοποίηση διαφορετικών εκφράσεων προσώπου.

Λέξεις κλειδιά – AAM, ενεργό μοντέλο εμφάνισης, αναγνώριση εκφράσεων προσώπου, βαθιά νευρωνικά δίκτυα

Abstract

This project aims to analyze the expressions of a set of face data from the real world. To get the changes in the shape and appearance of the face in an image sequence, a version of AAM (Active Appearance Model) has been developed. This model is then fed to the recently popular deep neural networks to achieve good classification of different facial expressions.

Keywords – AAM, Active appearance model, facial expression recognition, deep neural networks

Περιεχόμενα

| | |
|--|----|
| 1. Εισαγωγή..... | 8 |
| 1.1 Διάρθρωση εργασίας..... | 8 |
| 2. Εκφράσεις προσώπου..... | 9 |
| 2.1 Ιστορική αναδρομή..... | 9 |
| 2.2 Γενικά για τις εκφράσεις προσώπου..... | 10 |
| 2.3 Διάφορες εφαρμογές | 13 |
| 2.4 Παραμετροποίηση προσώπου..... | 15 |
| 2.4.1 Σύστημα κωδικοποίησης κινήσεων προσώπου..... | 15 |
| 2.4.2 Παράμετροι Facial Animation..... | 19 |
| 2.5 Παγκόσμιες εκφράσεις προσώπου..... | 21 |
| 2.6 Εκφράσεις προσώπου αποτέλεσμα των συναισθηματικών αντιδράσεων..... | 22 |
| 3. Στατιστικά μοντέλα..... | 24 |
| 3.1 Εισαγωγή..... | 24 |
| 3.2 Στατιστικά μοντέλα σχήματος..... | 26 |
| 3.3 Ενεργά μοντέλα εμφάνισης..... | 29 |
| 3.3.1 Γενικά..... | 29 |
| 3.3.2 Δημιουργία μοντέλου..... | 30 |
| 3.3.3 Εκπαίδευση μοντέλου..... | 33 |
| 3.3.4. Εφαρμογή μοντέλου σε εικόνα..... | 34 |
| 3.3.5 Πειραματικά αποτελέσματα εφαρμογής μοντέλου..... | 36 |
| 3.3.6 Ακολουθίες παρατηρήσεων και ακολουθίες σημείων..... | 39 |
| 3.3.7 Εντοπισμός προσώπου..... | 40 |
| 3.3.8 Viola-Jones ανιχνευτής προσώπου..... | 40 |
| 4. Τεχνητή νοημοσύνη | 42 |
| 4.1 Μηχανική μάθηση | 43 |

| | |
|---|----|
| 4.2 Νευρωνικά δίκτυα | 43 |
| 4.3 Βαθιά νευρωνικά δίκτυα | 45 |
| 4.4 Επιβλεπόμενη μάθηση νευρωνικών δικτύων | 46 |
| 4.5 Συνελκτικά νευρωνικά δίκτυα | 48 |
| 4.6 Προ-εκπαιδευμένα Νευρωνικά δίκτυα – Pre-trained NN..... | 50 |
| 4.7 Μοντέλο GoogleNet | 51 |
| | |
| 5. Εκπαίδευση μοντέλων και αποτελέσματα..... | 53 |
| 5.1 Αρχικοποίηση μοντέλου | 53 |
| 5.2 Προετοιμασία δεδομένων | 54 |
| 5.3 Εκπαίδευση μοντέλου..... | 55 |
| | |
| 6. Σύνοψη και συμπεράσματα | 59 |
| | |
| 7. Μελλοντικές κατευθύνσεις | 60 |

1. Εισαγωγή

Η σημασία της έκφρασης του ανθρώπινου προσώπου είναι μεγάλη όσο και το ενδιαφέρον ως προς το πώς αυτή αναλύεται. Οι εκφράσεις αυτές παρουσιάζονται σε πολλές και διαφορετικές εφαρμογές, όμως εκτός από αυτό, οι εικόνες του ανθρώπινου προσώπου διαφέρουν από τις άλλες κατηγορίες αντικειμένων γιατί έχουν το χαρακτηριστικό ότι οι άνθρωποι μπορούν να τις αναγνωρίζουν. Είναι εύκολο να εντοπιστούν από τους ανθρώπους ακόμα και οι πολύ μικρές αλλαγές σε ένα πρόσωπο. Οι άνθρωποι αναγνωρίζουν τα πρόσωπα και ταυτοποιούν τα αισθήματα που παρουσιάζονται. Έχουν δηλαδή ανεπτυγμένη την ικανότητα κωδικοποίησης και αξιολόγησης των εκφράσεων του προσώπου. Αυτός είναι ο λόγος που μας απασχολεί ιδιαίτερα η ανάλυση έκφρασης του προσώπου.

Στο ανθρώπινο πρόσωπο απεικονίζεται η πληροφορία σχετικά με τα συναισθήματα του ατόμου. Η ερμηνεία, επομένως, των εκφράσεων μπορεί να αποδειχθεί αρκετά χρήσιμη σε πολλές περιπτώσεις, κατά τη διάρκεια κοινωνικών επαφών, για την επίλυση προβλημάτων, για την επικοινωνία μεταξύ ατόμων διαφορετικών εθνικοτήτων, ή ακόμα και για τον εντοπισμό αλήθειας ή ψεύδους. Εξαιτίας της μεγάλης σημασίας των εκφράσεων και του ρόλου που παίζουν στην καθημερινή ζωή, το θέμα αυτό έχει απασχολήσει τους επιστήμονες με εκτεταμένες έρευνες. Ένα σύστημα που μπορεί και αναγνωρίζει τις εκφράσεις, φαίνεται να έχει πιο ανθρώπινη ταυτότητα και γίνεται λιγότερο απρόσωπο εφόσον αλληλεπιδρά με τον χρήστη άμεσα.

1.1 Διάρθρωση εργασίας

Η εργασία αυτή αποτελείται από 7 κεφάλαια. Στο 1ο γίνεται περιγραφή του ζητήματος της αναγνώρισης των εκφράσεων του προσώπου. Στο 2ο κεφάλαιο περιγράφονται οι κατηγορίες των εκφράσεων και το πόσο σημαντικές είναι για τον άνθρωπο. Επίσης, γίνεται ανασκόπηση της σχετικής βιβλιογραφίας από όταν πρωτοξεκίνησε η σχετική έρευνα μέχρι και σήμερα. Το 3ο κεφάλαιο αναφέρεται στα στατιστικά μοντέλα σχήματος και ειδικότερα τα ενεργά μοντέλα εμφάνισης. Παρουσιάζεται η χρήση τους για τον εντοπισμό του ανθρώπινου προσώπου και των χαρακτηριστικών του σε εικόνες. Στο 4ο κεφάλαιο περιγράφεται το σύστημα αυτόματης αναγνώρισης εκφράσεων που αναπτύχθηκε με νευρωνικά δίκτυα για τους σκοπούς αυτής της εργασίας. Το 5ο κεφάλαιο αναφέρεται στην εκπαίδευση των μοντέλων και τα πειραματικά

αποτελέσματα. Το 6ο κεφάλαιο αναφέρεται στη σύνοψη και τα συμπεράσματα, ενώ το 7ο στις μελλοντικές κατευθύνσεις.

2. Εκφράσεις Προσώπου

2.1 Ιστορική αναδρομή

Μελέτες για τις εκφράσεις του προσώπου και της φυσιογνωμίας, γενικότερα, έχουν ξεκινήσει από πολύ νωρίς τον 4ο αιώνα π.Χ. την εποχή του Αριστοτέλη. Ως φυσιογνωμία ορίζεται η ερμηνεία του χαρακτήρα και της προσωπικότητας ενός ατόμου από την εξωτερική του εμφάνιση και, ειδικά, από το πρόσωπό του. Ορισμένες από τις σημαντικότερες μελέτες σχετικά με τις εκφράσεις του προσώπου, που αποτέλεσαν τη βάση της σύγχρονης έρευνας, τοποθετούνται χρονικά στον 17ο αιώνα. Σε βιβλίο του που γράφτηκε το 1649, την «Παθομυοτομία», ο John Bulwer περιγράφει λεπτομερώς τις διαφορετικές εκφράσεις που είναι κινήσεις των μυών του κεφαλιού. Ενδιαφέρον, επίσης, παρουσιάζει το έργο του Γάλλου ζωγράφου και ακαδημαϊκού Le Brun στο οποίο βασίστηκαν οι ηθοποιοί και καλλιτέχνες του 18ου αιώνα προκειμένου να ερμηνεύσουν και να αποδώσουν την «τέλεια μίμηση των αυθεντικών εκφράσεων του προσώπου».

Ένα σημαντικό έργο που επιδρά στην επιστήμη του σήμερα χρονολογείται από τον 19^ο αιώνα. Το έργο αυτό είναι του Καρόλου Δαρβίνου και αποτελεί θεμέλιο λίθο στην αυτόματη αναγνώριση εκφράσεων προσώπου. Η έρευνά του το 1872 αφορούσε στις γενικές αρχές που αναφέρονται στις εκφράσεις του προσώπου και, συγκεκριμένα, στους τρόπους που εκφράζονται οι άνθρωποι και τα ζώα.

- στεναχώρια
- χαρά
- συλλογισμός
- θυμός
- αηδία
- έκπληξη
- ντροπή

Εκεί κατέγραψε τις παραμετροποιήσεις του προσώπου που σημειώνονται σε κάθε μία από τις προηγούμενες κατηγορίες εκφράσεων. Δηλαδή, κατέγραψε τη σύσπαση των μυών που

περιβάλλουν τα μάτια όταν κάποιος είναι θλιμμένος και το χαμήλωμα του πιγουνιού όταν κάποιος είναι άκεφος.

Σημαντικό ορόσημο για τη 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 ιδιαίτερα σημαντικό και επηρεάζει ακόμα και σήμερα τα συστήματα αυτόματης αναγνώρισης εκφράσεων.

Οι εκφράσεις του προσώπου μελετώνται από ψυχολόγους, γιατρούς, καλλιτέχνες και ηθοποιούς. Τα τελευταία χρόνια, με τις τελευταίες εξελίξεις στους τομείς της ρομποτικής, στα γραφικά υπολογιστών και στην όραση των υπολογιστών, παρουσιάστηκε ιδιαίτερο ενδιαφέρον από ανθρώπους που εργάζονται στη δημιουργία κινουμένων σχεδίων (animators) αλλά και μηχανικούς ερευνητές των εκφράσεων του ανθρώπινου προσώπου.

Το πρώτο βήμα για την αυτόματη αναγνώριση εκφράσεων έγινε από τον Suwa και τους συνεργάτες του όταν το 1978 παρουσιάστηκε ένα σύστημα για την ανάλυση των εκφράσεων μίας ακολουθίας εικόνων, η οποία χρησιμοποιούσε 20 σημεία εντοπισμού. Παρόλο που το σύστημα προτάθηκε το 1978, οι ερευνητές δεσυνέχισαν προς αυτήν την κατεύθυνση μέχρι την αρχή της δεκαετίας του '90. Αυτό συνέβαινε διότι έως το 1980 τα συστήματα εντοπισμού του προσώπου που απαιτούνται για την αυτόματη αναγνώριση των εκφράσεων βρίσκονταν ακόμη σε στάδιο εξέλιξης.

Στο τέλος της δεκαετίας του '80 και στις αρχές του '90 έγινε ευρέως διαθέσιμη η φθηνή υπολογιστική ισχύς και είχε ως αποτέλεσμα την ανάπτυξη συστημάτων εντοπισμού του προσώπου. Επιπλέον, άρχισαν να γίνονται πιο δημοφιλείς οι τομείς της Αλληλεπίδρασης Ανθρώπου-Υπολογιστή και της Συναισθηματικής Υπολογιστικής. Έτσι, δόθηκε το έναυσμα για να ανανεωθεί το ενδιαφέρον για την ανάπτυξη συστημάτων αυτόματης αναγνώρισης εκφράσεων.

2.2 Γενικά για τις εκφράσεις προσώπου

Η αλλαγή της έκφρασης του προσώπου είναι ένας από τους τρόπους επίδειξης των συναισθημάτων του ανθρώπου. Μέχρι το τέλος της δεκαετίας του '60, υπήρχε η άποψη ότι οι εκφράσεις του προσώπου ήταν λιγότερο αξιόπιστες και με μικρή επικοινωνιακή αξία. Οι σχετικοί συγγραφείς της εποχής εκείνης ανέφεραν πολλά παραδείγματα που κάποια άτομα χαμογελούσαν ενώ παρακολουθούσαν κάτι που ήταν άσχημο, στα οποία δεν αντιστοιχούσαν οι εκφράσεις του

προσώπου με το συναίσθημα που ένιωθε εκείνη τη δεδομένη στιγμή ο άνθρωπος. Σήμερα, όμως, υπάρχουν αρκετές έρευνες που επιβεβαιώνουν αυτή τη σχέση.

Σύμφωνα με έρευνες, το χαμόγελο που συνοδεύεται από το σήκωμα των μάγουλων, συνδέεται με θετικά συναισθήματα σε μικρής και μεγαλύτερης ηλικίας ενήλικες. Οι εκφράσεις της διασκέδασης και της αμηχανίας, το αυθόρμητο χαμόγελο και το γέλιο, εκφράζουν διακριτά συναισθήματα. Έχει αποδειχθεί ότι το πόσο δυνατό είναι το γέλιο και το χαμόγελο φαίνεται από το πόσο αστείο τελικά ήταν αυτό που συνέβη.

Οι εκφράσεις του προσώπου κατά το Δαρβίνο είναι μέρος πιο περίπλοκων αντιδράσεων και παρουσιάζονται σε συνδυασμό με κινήσεις του σώματος και αντιδράσεις, όπως η φωνή, η στάση του σώματος, και οι διακυμάνσεις του νευρικού συστήματος.

Οι ερευνητές της θεωρίας της εξέλιξης υποστηρίζουν, επίσης, ότι οι εκφράσεις του προσώπου είναι πολύ σημαντικές και όχι μόνο κάποια ένδειξη της εσωτερικής κατάστασης του ατόμου. Επηρεάζουν τις κοινωνικές επαφές, περιέχουν πληροφορίες για τη συναισθηματική κατάσταση, τη διάθεση, τη σχέση με τον παραλήπτη της έκφρασης. Τα συμπεράσματα αυτά προέκυψαν έπειτα από έρευνες στις οποίες συμμετείχαν γονείς που επικοινωνούσαν με τα παιδιά τους καθώς και από παρακολούθηση συμπεριφορών όπως η φιλία, το φλερτ, το παιχνίδι και η άσκηση βίας.

Τα χαρακτηριστικά του προσώπου παίζουν σημαντικό ρόλο στην αναγνώριση εκφράσεων. Μπορεί να είναι είτε παροδικά, είτε μόνιμα. Μόνιμα θεωρούνται τα χείλη, τα μάγουλα, η μύτη και τα μάτια ενώ παροδικά θεωρούνται οι ρυτίδες έκφρασης που χάνονται όταν βρίσκεται το άτομο σε μια ουδέτερη συναισθηματική κατάσταση, όπως επίσης παροδικές είναι και οι γραμμές που εμφανίζονται ενδιάμεσα από τα φρύδια κατά την διάρκεια του συναισθήματος θυμού.

Στο στόμα και στα φρύδια εμπεριέχεται το μεγαλύτερο μέρος πληροφορίας σχετικά με την έκφραση του προσώπου που εξετάζεται. Σύμφωνα με ερευνητές, η έκπληξη, η χαρά, και η απέχθεια 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. Έχει σημειωθεί ότι το στόμα μπορεί να έχει περισσότερη πληροφορία από τα φρύδια. Πειράματα που έγιναν με διαθέσιμο μόνο το στόμα, είχαν ποσοστό επιτυχίας 78% αλλά το ποσοστό έπεσε στο 50% όταν υπήρχαν μόνο τα φρύδια. Σύμφωνα με διαφορετική μελέτη, στο στόμα περιέχεται η πληροφορία για την έκφραση της λύπης. Οι έρευνες που έγιναν με μόνο ορισμένα χαρακτηριστικά του προσώπου ευδιάκριτα είναι ιδιαίτερα σημαντικές, αφού σε πραγματικές συνθήκες μπορεί να υπάρξει διαθέσιμο μόνο μέρος του προσώπου. Καθημερινά αντικείμενα όπως γυαλιά, κάποιο

μαντήλι, μάσκες, τροποποιούν την εικόνα του προσώπου. Αξιοσημείωτο είναι το γεγονός ότι αν έχουμε στη διάθεσή μας μόνο το μισό του προσώπου τα αποτελέσματα δεν αλλάζουν αφού οι εκφράσεις είναι συμμετρικές.

Οι Ekman και Friesen χρησιμοποίησαν σε έρευνές τους μόνο έξι συναισθήματα: χαρά, έκπληξη, στεναχώρια, θυμός, απέχθεια και φόβος. Αυτά είναι τα βασικά συναισθήματα. Πέρα από τα 6 βασικά συναισθήματα, το πρόσωπο του ανθρώπου έχει τη δυνατότητα να εκφράσει περαιτέρω κατηγορίες συναισθημάτων. Οι ερευνητές μελετούν περισσότερες από τις έξι αυτές εκφράσεις οι οποίες περιγράφουν κάποιο συναίσθημα, μελετώντας πώς οι εκφράσεις εξελίχθηκαν χρονικά, το πού κατευθύνεται το βλέμμα, το κεφάλι και τι στάση έχει το σώμα.

Καταγράφηκαν εκφράσεις που αφορούν την ντροπή ή την αμηχανία ενός ατόμου (Keltner 1995, Keltner&Buswell 1997, Keltner&Harker 1998) την συμπόνια (Eisenberg&al. 1989), διαφορές που υπάρχουν ανάμεσα στο γέλιο και το χαμόγελο (Keltner&Bonanno 1997). Έχουν δημοσιευθεί έρευνες σχετικές με το κόκκινο χρώμα του προσώπου (Leary, Britt, Cutlip, &Templeton 1992, Shearn et al. 1990) και την εξαγωγή της γλώσσας (Haidt&Keltner 1999) τα οποία μπορούν επίσης να εκφράζουν συναίσθημα. Το 2000, ο Parrot σημείωσε 136 συναισθηματικές καταστάσεις που εκφράζουν οι άνθρωποι, τις οποίες διαχώρισε σε ξεχωριστές κλάσεις και υποκλάσεις.

Τα συστήματα που δημιουργούνται κατά κύριο λόγο ασχολούνται με την αναγνώριση εκφράσεων που είναι βασικές και σε λιγότερες περιπτώσεις στις υπόλοιπες μη βασικές εκφράσεις. Οι προσπάθειες που σχετίζονται με την έρευνα την εξέλιξης μιας έκφρασης στη διάρκεια του χρόνου είναι αρκετά λίγες. Πιο συγκεκριμένα, οι έρευνες εστιάζουν στην παρατήρηση της έναρξης, της κορύφωσης καθώς και της λήξης κάθε μιας έκφρασης.

Επιπλέον, μια άλλη κατηγοριοποίηση είναι σε προμελετημένες ή αυθόρμητες εκφράσεις. Τεχνητές και προμελετημένες εκφράσεις θεωρούνται αυτές που συμβαίνουν κατ' απαίτηση. Συνήθως, αυτό παρατηρείται σε άτομα που συμμετέχουν σε έρευνες και γνωρίζουν ότι βρίσκονται υπό παρατήρηση σε κάποιον χώρο. Αντιθέτως, οι εκφράσεις που συμβαίνουν αυθόρμητα είναι αυτές που συμβαίνουν χωρίς σκέψη. Παρατηρούνται στην καθημερινότητα όταν γίνονται συζητήσεις, παρακολούθηση ταινίας ή κάποιας σκηνής. Τα συστήματα αυτόματης αναγνώρισης εκφράσεων, λόγω του ότι αυτές είναι εύκολο να αναγνωριστούν και να καταγραφούν, εστιάζουν περισσότερο στην αναγνώριση των προμελετημένων εκφράσεων. Είναι αρκετά δύσκολο να δημιουργηθεί μία βάση δεδομένων που περιέχει εικόνες και βίντεο με αυθόρμητες εκφράσεις. Σύμφωνα με έρευνες, οι εκφράσεις που είναι προμελετημένες μπορούν να περιγράψουν σε ικανοποιητικό βαθμό τα συναισθήματα της αμηχανίας, της ντροπής, του γέλιου και της συμπόνιας και σε μικρότερα

βασικά συναισθήματα. Οι αυθόρμητες εκφράσεις ξεχωρίζουν από τις προμελετημένες στην εμφάνισή τους και στα χρονικά χαρακτηριστικά τους. Επίσης, οι προμελετημένες εκφράσεις φαίνεται να είναι κάπως υπερβολικές και μερικές φορές ψεύτικες, εν αντιθέσει με τις αυθόρμητες που συνήθως είναι πιο διακριτικές.

Μία ακόμη κατηγορία είναι οι μικροεκφράσεις. Αυτές συμβαίνουν στις περιπτώσεις που το άτομο δεν εκφράζει το πραγματικό του συναίσθημα και κάνει έναν μικρό μορφαρμό. Οι μικροεκφράσεις έχουν διάρκεια μερικά κλάσματα του δευτερολέπτου, εν αντιθέσει με τις αυθόρμητες εκφράσεις. Εξαιτίας της δυσκολίας που προκύπτει κατά την καταγραφή αυτών των μικροεκφράσεων δεν έχουν γίνει αρκετές μελέτες προς αυτή την κατεύθυνση. Μελετήθηκαν από τον Ekman οι μικροεκφράσεις που αφορούν στον εντοπισμό ψέματος. Κομμάτι της μελέτης του ήταν και όσες εκφράσεις ξεκινούν να φαίνονται στο ανθρώπινο πρόσωπο πριν την αλλαγή του συναισθήματός του. Διαρκούν αρκετά περισσότερο από όσο διαρκούν οι μικροεκφράσεις αλλά δεν είναι ολοκληρωμένες στα χρονικά χαρακτηριστικά τους.

Όσον αφορά στο κατά πόσο κάποιο άτομο λέει ψέματα ή αλήθεια, αυτό βρίσκεται εύκολα από τις εκφράσεις που παίρνει το πρόσωπο του. Αυτό είναι πρόβλημα για τους ανθρώπους γιατί δεν έχουν όλοι την ικανότητα αυτή να το εντοπίσουν. Οι Ekman, Friesen, O'Sullivan και Scherer το 1980 σύγκριναν απαντήσεις διαφόρων παρατηρητών που έβλεπαν ειλικρινείς και μη ειλικρινείς εκφράσεις. Για τις ειλικρινείς εκφράσεις, τα συμπεράσματα των ανθρώπων αυτών ήταν σωστά, σε αντίθεση με τις μη-ειλικρινείς. Κάποιοι από αυτούς μπόρεσαν να συμπεράνουν με ακρίβεια ακόμα και τις μη-ειλικρινείς εκφράσεις και από αυτό προκύπτει το συμπέρασμα πως υπάρχει δυνατότητα σωστού συμπεράσματος σχετικά με το αν κάποιος λέει ψέματα με απλή παρατήρηση στο πρόσωπό του.

2.3 Διάφορες εφαρμογές

Πολλές και σημαντικές εφαρμογές χρησιμοποιούν σήμερα συστήματα αυτόματης αναγνώρισης εκφράσεων. Ιδιαίτερα, μετά την πρόσφατη εξέλιξη της ρομποτικής και των ανθρωποειδών ρομπότ, η χρήση αυτών των συστημάτων είναι απαραίτητη. Τα τελευταία χρόνια, τα ρομπότ έχουν όλο και περισσότερη σχέση με τον άνθρωπο αφού συναντώνται συχνά σε χώρους που κατοικούν και εργάζονται. Είναι επιτακτική η ανάγκη να γίνονται «εξυπνότερα» και να «κατανοούν» την διάθεση

και τα συναισθήματα των ανθρώπων. Η οπτική επικοινωνία μεταξύ ανθρώπου και υπολογιστή εξαρτάται άμεσα από τα συστήματα αναγνώρισης των εκφράσεων.

Η επικοινωνία μεταξύ των ανθρώπων καθώς και οι αντιδράσεις σε οποιαδήποτε συναισθηματική κατάσταση στην οποία βρίσκονται τη δεδομένη στιγμή πρέπει να προσομοιωθεί και από τους υπολογιστές. Στη συναισθηματική υπολογιστική βασίζεται η κοινότητα των ερευνητών που ασχολείται με την αλληλεπίδραση του ανθρώπου με τον υπολογιστή. Σημαντικό ρόλο παίζει η αναγνώριση των εκφράσεων στην κατανόηση των συναισθημάτων καθώς επίσης και στη δημιουργία μιας αξιόπιστης επικοινωνίας και αλληλεπίδρασης ανθρώπου – υπολογιστή. Φυσικά, εκτός από τη συναισθηματική υπολογιστική και τη ρομποτική που αναφέρθηκε υπάρχουν και άλλες εφαρμογές στις οποίες βρίσκουν εφαρμογή τα συστήματα αναγνώρισης εκφράσεων, σε τομείς όπως: στις τηλεπικοινωνίες, στα κινούμενα σχέδια, στα βιντεοπαιχνίδια, στην ψυχιατρική, στα συστήματα ασφάλειας των αυτοκινήτων, σε εκπαιδευτικά λογισμικά, στην τηλεόραση και στο ραδιόφωνο για επιλογή του προγράμματος ανάλογα με το πως νιώθει ο χρήστης.



Εικόνα 1. Εφαρμογές αυτόματης αναγνώρισης εκφράσεων

Επιπλέον, υπάρχουν αρκετές σημαντικές και αποτελεσματικές εφαρμογές σε πραγματικό χρόνο. Παράδειγμα αποτελεί σύστημα το οποίο χρησιμοποιεί ψηφιακούς χαρακτήρες για να μπορεί να

αναπαραστήσει τις εκφράσεις ενός χρήστη, αλλά και συστήματα που, ενώ διαρκεί κάποια συνομιλία, εισάγεται στη συνομιλία αυτή εμοτίση που σχετίζεται με την έκφραση του προσώπου κάποιου χρήστη.

Μελλοντικά, αναμένεται να υπάρξουν παρουσιάσεις πολλών νέων και καινοτόμων εφαρμογών διότι τα συστήματα αναγνώρισης των εκφράσεων φαίνεται να είναι αποτελεσματικά, ιδιαίτερα σε εφαρμογές πραγματικού χρόνου.

2.4 Παραμετροποίηση προσώπου

Η έκφραση του προσώπου που εξετάζεται μπορεί να παραμετροποιηθεί με βάση τις κινήσεις των μυών. Εν συνεχεία, οι παράμετροι χρησιμοποιούνται για να αναπαραστήσουν τις εκφράσεις του προσώπου. Οι δύο κύριες τεχνικές για την παραμετροποίηση είναι:

- α) Το FAC που παρουσιάστηκε πρώτη φορά από τους Ekman και Friesen το 1977 και
- β) Οι παράμετροι κινήσεων προσώπου που αποτελούν μέρος του προτύπου MPEG-4 (συνθετική/φυσική υβριδική κωδικοποίηση) το 1998.

2.4.1 Σύστημα κωδικοποίησης κινήσεων προσώπου

Οι ερευνητικές ομάδες που ασχολούνται με τις εκφράσεις του προσώπου, πριν θεμελιωθούν τα FACS, βασίζονταν στην ανθρώπινη κρίση και παρατήρηση του υπό εξέταση προσώπου και έπειτα παρουσίαζαν τα αποτελέσματά τους. Αυτό, φυσικά, δεν θεωρείται επιστημονική εξήγηση αφού οι άνθρωποι δεν μπορεί να είναι πάντα σωστοί και αξιόπιστοι ή ακριβείς. Αρκετά εύκολα, κάποιος παρατηρητής μπορεί και επηρεάζεται από εξωτερικές πληροφορίες, όπως η φωνή του ανθρώπου, και τα αποτελέσματα που εξάγονται συνδέονται με το πολιτιστικό υπόβαθρό τους.

Υπάρχουν περιορισμοί που προκύπτουν κατά τη χρήση ανθρώπου – παρατηρητή, οι οποίοι μπορεί να μη ληφθούν υπόψη κατά την αναπαράσταση εκφράσεων με κάποιο σύνολο σταθερών παραμέτρων. Όλα αυτά βοηθούν στο να παρατηρηθούν ανεξάρτητα οι παράμετροι με παράκαμψη της συμπεριφοράς του προσώπου συνολικά.

Γενικά, τα συστήματα αυτά προσεγγίζουν την αναγνώριση εκφράσεων μέσω των μυών του προσώπου. Αφού εντοπιστούν οι διάφοροι μύες που είναι αναξάρτητοι μεταξύ τους ή συνδυάζονται σε ομάδες, τότε προκαλούνται αλλαγές στο πρόσωπο και στην εμφάνισή του. Αυτές οι αλλαγές και οι μύες που είναι ένας ή σε ομάδες προκαλούν αλλαγές που λέγονται Action Units. Αυτά τα Action Units είναι που αποτελούν τα FACS. Παρακάτω γίνεται αναφορά σε αυτές:
















- AU1: σήκωμα του έσω φρυδιού. Συνδέεται με τους μύες μετωπικός & Pars μέσος
- AU 2: σήκωμα του έξω φρυδιού. Συνδέεται με τους μύες μετωπικός & Pars μέσος
- AU 26: προς τα κάτω σαγόι. Συνδέεται με τους μύες Masetter, προσωρινός και εσωτερικός περυγοειδής

Υπάρχουν και AUs που δεν προκαλούνται από μύες του προσώπου. Για παράδειγμα:

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

Οι AUs μπορεί να είναι προσθετικές ή μη-προσθετικές. Οι προσθετικές είναι αυτές που η εμφάνιση της κάθε AU είναι ανεξάρτητη, ενώ μη-προσθετικές αυτές που επηρεάζονται η κάθε μία από την εμφάνιση της άλλης. Η κάθε έκφραση αναπαρίσταται σαν συνδυασμός μίας ή και παραπάνω προσθετικών ή όχι AUs. Ένα παραδειγμα είναι αυτό του συναισθήματος φόβου που μπορεί να γίνει με συνδυασμό των εξής AUs: 1, 2 και 26.

Σχόλιο [u1]: Μήπως το αντίθετο;

| NEUTRAL | AU 1 | AU 2 | AU 4 | AU 5 |
|---|---|---|---|--|
|  |  |  |  |  |
| 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: AUstovn ματιών και συνδυασμοί.

Παραδείγματα των AUs που αφορούν στο άνω και κάτω μέρος του προσώπου καθώς και τις κινήσεις τους απεικονίζονται στις εικόνες 2 και 3.

| | | | | |
|---|--|--|---|--|
| <i>NEUTRAL</i> | AU 9 | AU 10 | AU 12 | AU 20 |
|  |  |  |  |  |
| 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. |
| AU 15 | 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 | | | |
|  |  | | | |

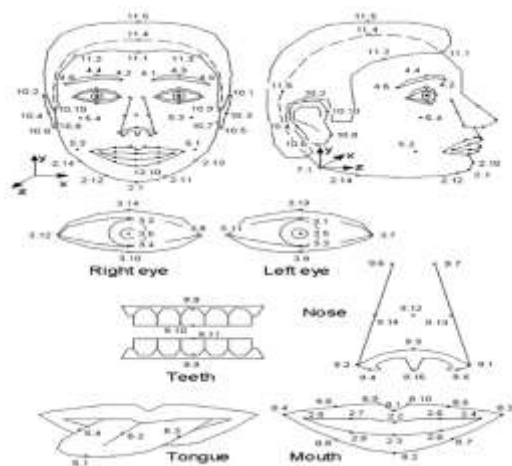
Εικόνα 3: AUs κάτω προσώπου.

2.4.2 Παράμετροι Facial Animation

Κατά τη δεκαετία του '90, οι ερευνητές των κινουμένων σχεδίων σε υπολογιστή συνάντησαν παρόμοια θέματα με αυτά που αντιμετώπισαν οι ερευνητές της αυτόματης αναγνώρισης προσώπου και εκφράσεων πριν θεμελιωθούν τα FACS. Γενικά, δεν υπήρξε ένα ενιαίο πλαίσιο και αυτό είχε ως συνέπεια τα συστήματα που αναπτύχθηκαν για λόγους κινουμένων σχεδίων να έχουν διαφορετικές παραμέτρους. Για να αντιμετωπιστεί αυτό το θέμα, το MPEG είχε ίδιες προδιαγραφές με το animation προσώπου στο πρότυπο MPEG-4. Το MPEG-4 έχει τις παραμέτρους animation προσώπου (FAPs) που είναι αλληλένδετες με τις Action Units του FACS.

Κατά το MPEG-4, ένα μοντέλο προσώπου που βρίσκεται σε ουδέτερη κατάσταση έχει ένα συγκεκριμένο σύνολο ιδιαιτεροτήτων, δηλαδή, όλοι οι μύες του προσώπου χαλαροί, τα βλέφαρα εφάπτονται με την ίριδα και η κόρη έχει διάμετρο που ισούται με το 1/3 της διαμέτρου της ίριδας.

Το πρότυπο αυτό ορίζεται από 84 σημεία (FP) που είναι σημαντικά χαρακτηριστικά προσώπων που βρίσκονται στην ουδέτερη φάση. Για την κατανόηση καθώς και την αναγνώριση κινήσεων προσώπου και εκφράσεων, χρησιμοποιείται η μετακίνηση των σημείων FP. Η παρακάτω εικόνα 4 αφορά στις θέσεις 84 σημείων προσώπου που έχουν οριστεί με βάση το πρότυπο MPEG-4.



Εικόνα 4: Τα FPs του MPEG-4.

Οι FAPs είναι ένα σύνολο από παραμέτρους που αναπαριστούν τις συνολικές κινήσεις του προσώπου και την κίνηση του κεφαλιού, του στόματος, της γλώσσας και των ματιών.

Κάθε FAP αποτελεί κίνηση του προσώπου στην ουδέτερη κατάσταση που παραμορφώνεται. Η τιμή που έχει κάθε FAP δείχνει το πόσο μεγάλη είναι η κίνηση, που με την σειρά της δείχνει το πόσο μεγάλη είναι η παραμόρφωση που έχει προκληθεί, όπως, για παράδειγμα, ένα μικρό γέλιο σε σχέση με ένα πλατύ γέλιο. Τα 68 FAPs ορίζονται από το πρότυπο MPEG-4. Παρακάτω, στον πίνακα 1 που ακολουθεί, φαίνονται οι FAPs και το πώς αυτές περιγράφονται.

| 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 1: FAPs

Οι έξι βασικές εκφράσεις: χαράς, θυμού, φόβου, λύπης, έκπληξης και αηδίας, ορίζονται από το πρότυπο MPEG-4. Παρότι οι FAPs είχαν αρχικά δημιουργηθεί για εξυπηρέτηση σκοπών που αφορούσαν τα animations, τα τελευταία χρόνια οι ερευνητές τα χρησιμοποιούν για σκοπούς αναγνώρισης συναισθημάτων και έκφρασης.

Η σχέση των AUs του FACS και των FAPs του MPEG4 είναι σημαντική. Οι αντιστοιχίες τους παρουσιάζονται στον πίνακα 2.

| 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: FACS AUs σε αντιστοίχιση με FAPs.

2.5 Παγκόσμιες εκφράσεις προσώπου

Ένα ερώτημα που εξετάζεται από ερευνητές είναι εάν οι εκφράσεις που περιγράφουν τα συναισθήματα έχουν σχέση με την κουλτούρα και το πολιτισμικό περιβάλλον του κάθε ατόμου.

Αυτό που υποστηρίζει ο Δαρβίνος, βάσει της θεωρίας της εξέλιξης, είναι πως οι εκφράσεις των προσώπων μπορούν να θεωρούνται παγκόσμιες. 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. Έκτοτε, παρόλες τις διαφωνίες, η άποψη αυτή έχει επικρατήσει.

Σε έρευνα σε άτομα από διαφορετικές κουλτούρες, προβλήθηκαν 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. Αποτελέσματα ερευνών δείχνουν ότι οι άνθρωποι ερμηνεύουν τη συναισθηματική κατάσταση με μεγάλη ακρίβεια που φτάνει μέχρι και το 80%. Αυτά τα αποτελέσματα έκαναν τους ερευνητές να συμπεράνουν πως οι άνθρωποι απ' όλο τον κόσμο προσλαμβάνουν τις εκφράσεις του θυμού, της περιφρόνησης, του φόβου, της λύπης, της αηδίας, και της έκπληξης με σχεδόν ίδιο τρόπο.

Η ακρίβεια της έρευνας αμφισβητήθηκε λόγω του ότι υπήρχε λίστα συναισθημάτων από την οποία οι συμμετέχοντες έπρεπε να επιλέξουν την απάντησή τους και του ότι τα συναισθήματα που παρουσιάζονταν δεν ήταν αυθόρμητα. Ωστόσο, οι επόμενες έρευνες έδειξαν πως τα αποτελέσματα δεν διαφέρουν όταν οι συμμετέχοντες περιγράφουν εικόνες με δικά τους λόγια και όταν τους παρουσιάζονται εικόνες από αυθόρμητα συναισθήματα.

Ο Ekman, το 1973, παρατήρησε το πώς αντιδρούν οι φοιτητές στις ΗΠΑ και την Ιαπωνία καθώς έβλεπαν κάποιο βίντεο. Καθόλη τη διάρκεια της έρευνας, οι φοιτητές δεν είχαν ιδέα πως καταγράφονται οι αντιδράσεις τους και φάνηκε να έχουν παρόμοιες ή ίδιες αρνητικές εκφράσεις στο πρόσωπό τους ενώ παρακολουθούσαν βίντεο που είχε σχεδιαστεί με σκοπό να προκαλέσει το άγχος τους. Έπειτα από έρευνα το '92 των Camras, Campos, Oster, Miyake & Bradshaw, φάνηκε πως βρέφη μικρότερα του ενός έτους από τις ΗΠΑ και Ιαπωνία είχαν παρόμοιες συμπεριφορές και εκφράσεις, στάση σώματος και ήχους, όταν έδειχναν θυμό σε συνθήκες που αισθάνονταν ότι προκαλούν πόνο.

Έρευνες που πραγματοποιήθηκαν το 2003 από τον Horstmann και από τους Lawrence et al. το 2005, δείχνουν ότι η λογική με την οποία οι άνθρωποι εκφράζουν και αναγνωρίζουν τα συναισθήματα γίνεται με τον ίδιο τρόπο χωρίς να παίζει ρόλο η καταγωγή τους.

Άλλη έρευνα συγκέντρωσε αναφορές διαφορετικών ανθρώπων που σχετίζονται με εκφράσεις και συναισθήματα και το συμπέρασμα που προέκυψε ήταν πως άνθρωποι από όλα τα μέρη του κόσμου δείχνουν να συμφωνούν στο πως εκφράζουν το συναίσθημα αυτό. Παράδειγμα αποτελεί η έκφραση της αμηχανίας που μπορεί να εκφραστεί με νευρικό γέλιο και με στροφή του βλέμματος.

Μπορεί οι εκφράσεις του προσώπου να είναι παγκόσμιες και ίδιες σε ό,τι έχει να κάνει με την έκφραση συναισθημάτων, αλλά φαίνεται να υπάρχουν και λίγες διαφορές από κουλτούρα σε κουλτούρα στο συναίσθημα. Οι εκφράσεις είναι παγκόσμιες επειδή είναι πολύ σημαντικές.

2.6 Εκφράσεις προσώπου αποτέλεσμα συναισθηματικών αντιδράσεων

Οι εκφράσεις του προσώπου είναι άμεσα συνδεδεμένες με τη συναισθηματική εμπειρία του ατόμου και συντονίζονται με άλλες αντιδράσεις όπως είναι οι νευρο-ενδοκρινικές αλλαγές. Οι ερευνητές αναφέρουν το γεγονός αυτό ως «συναισθηματικό πακέτο».

Οι εκφράσεις του προσώπου συσχετίζονται με φυσιολογικές αλλαγές που έχει ο οργανισμός. Παράδειγμα, το συναίσθημα τρόμου συνοδεύεται από την αντίδραση της εφίδρωσης, του χλωμιάσματος ή την διαστολή της κόρης, η αμηχανία έπειτα από το κοκκίνισμα, ενώ η λύπη έπειτα από το κλάμα. Στις έρευνες αυτές οι συμμετέχοντες ήταν άνθρωποι από διαφορετική καταγωγή (ΗΠΑ, Ινδονησία).

Παρατηρείται, επίσης έντονη σχέση των εκφράσεων του προσώπου με την συμπεριφορά του ατόμου.

Το 1972, οι Ekman, Liebert et al. έλεγξαν τη σχέση που έχουν οι εκφράσεις και η συμπεριφορά παιδιών ενώ έβλεπαν προγράμματα στην τηλεόραση. Οι αντιδράσεις τους καταγράφηκαν καθώς παρακολουθούσαν σκηνές βίας την ταινία και σκηνές από ένα αθλητικό πρόγραμμα. Εν συνέχεια, τα παιδιά αυτά είχαν την δυνατότητα να επιλέξουν είτε να βλάψουν, είτε να βοηθήσουν κάποιο άλλο παιδί. Φάνηκε ότι τα παιδιά που γέλασαν με αυτό στην συνέχεια είχαν βίαιη συμπεριφορά ενώ αυτά που στεναχωρήθηκαν βοήθησαν άλλο παιδί. Οι Matsumoto, Gary, Haan, Theodorou και

Σχόλιο [u2]: Αυτά σχετίζονται με τον τρόπο ή κάτι άλλο;

Μορφοποιήθηκε: Γραμματοσειρά: 12 pt, Ελληνικά (Ελλάδα)

Μορφοποιήθηκε: Γραμματοσειρά: 12 pt, Ελληνικά (Ελλάδα)

Μορφοποιήθηκε: Γραμματοσειρά: 12 pt, Ελληνικά (Ελλάδα)

Μορφοποιήθηκε: Γραμματοσειρά: 12 pt, Ελληνικά (Ελλάδα)

Μορφοποιήθηκε: Γραμματοσειρά: 12 pt, Ελληνικά (Ελλάδα)

Μορφοποιήθηκε: Γραμματοσειρά: 12 pt, Ελληνικά (Ελλάδα)

Cooke-Carney πραγματοποίησαν έρευνες σε παιδιά που βρίσκονταν σε προσχολική ηλικία το '86 και οι έρευνες από τους Moffit, Keltner, και Stouthamer-Loeber σε εφήβους το '95 είχαν τα ίδια συμπεράσματα.

Σύμφωνα με όσα έχουν αναφερθεί, τα συμπεράσματα για τις εκφράσεις του προσώπου είναι τα παρακάτω:

- Είναι παγκόσμιοι δείκτες των συναισθημάτων όταν αυτά είναι αυθόρμητα και όχι προσποιητά.
- Αλλάζουν ανάλογα με τη συναισθηματική εμπειρία του ανθρώπου.
- Αποτελούν μέρος πιο σύνθετου πακέτου που αφορά στις συναισθηματικές αντιδράσεις.
- Άνθρωποι από διαφορετικές κουλτούρες τα αναγνωρίζουν εύκολα.
- Διευκολύνουν τις διαπροσωπικές και κοινωνικές επαφές.

3. Στατιστικά Μοντέλα

3.1 Εισαγωγή

Η όραση των υπολογιστών έχει ως στόχο την κατανόηση των δομών της εικόνας αλλά και τι ακριβώς αυτές οι εικόνες αναπαριστούν. Για να επιτευχθεί αυτός ο στόχος χρησιμοποιούνται μοντέλα. Για τα μοντέλα αυτά, δίνεται μια περιγραφή ή μια ετικέτα κατηγορίας των εικόνων στις οποίες βασίζονται, και περιγράφονται έτσι τα αντικείμενα που έχουν αυτή τη δομή. Συγκεκριμένα, αυτά που έχουν την ίδια δομή όπως περιγράφονται από τον άνθρωπο. Όταν πρόκειται για εικόνες πολύπλοκων και ευμετάβλητων αντικειμένων, η ερμηνεία τους με μεθόδους που βασίζονται στη χρήση μοντέλων είναι πολύ δυσκολότερη. Σε αυτή την κατηγορία βρίσκονται οι εικόνες του ανθρώπινου προσώπου. Το πρόβλημα είναι στο γεγονός πως, για να είναι ένα μοντέλο επιτυχημένο και χρήσιμο, πρέπει να έχει δύο χαρακτηριστικά τα οποία είναι σχεδόν αμοιβαία αποκλειόμενα μεταξύ τους. Μόνο αποδεκτά στιγμιότυπα του αντικειμένου μπορούν να αναπαρίστανται από το μοντέλο. Επιπλέον, είναι απαραίτητο να επιτρέπεται μέχρι κάποιο βαθμό η μεταβλητότητα (variation) που θα περιγράφει τα φυσικά αντικείμενα συνολικά. Η αντιμετώπιση αυτού του προβλήματος βασίζεται στα στατιστικά μοντέλα που μπορούν και χρησιμοποιούνται απευθείας για να ερμηνεύσουν τις εικόνες. Τα μοντέλα AAM ανήκουν στα στατιστικά μοντέλα. Τα ενεργά μοντέλα εμφάνισης [5] μελετώνται στη συνέχεια. Η περιγραφή των μοντέλων ενεργού περιγράμματος αποτελεί τη βάση σε εφαρμογές όπου γίνεται αναγνώριση των εκφράσεων του προσώπου. Αυτές οι αρχές έχουν εφαρμογή σε πολλούς τομείς.

Υπό πραγματικές συνθήκες, οι δομές τις οποίες καλείται να αναγνωρίσει ένα μοντέλο είναι πολύπλοκες και ευμετάβλητες. Κάποιες φορές που οι εικόνες εμπεριέχουν θόρυβο και η πληροφορία δεν είναι επαρκής, αυτό έχει ως αποτέλεσμα να είναι σχεδόν αδύνατο να ερμηνευθεί η εικόνα δίχως να υπάρχει προηγουμένως γνώση που είναι σχετική με το εξεταζόμενο αντικείμενο όπως, για παράδειγμα, ιατρική απεικόνιση όταν αφορά στα όργανα του ανθρώπινου σώματος. Οι παραπάνω δυσκολίες αντιμετωπίζονται χρησιμοποιώντας μεθόδους που στηρίζονται στη χρήση μοντέλων.

Ιδιαίτερο ενδιαφέρον έχουν τα γενετικά μοντέλα. Τα γενετικά μοντέλα έχουν την ικανότητα να εξάγουν ρεαλιστικές εικόνες από τα υπό εξέταση αντικείμενα. Παράδειγμα γενετικού μοντέλου μπορεί να εξάγει εικόνες προσώπου ενός ανθρώπου έχοντας αλλάξει την έκφρασή του. Χρησιμοποιώντας ένα τέτοιο μοντέλο, η ερμηνεία των εικόνων μετατρέπεται αυτομάτως σε ένα

πρόβλημα ταιριάσματος. Η μεταβολή των παραμέτρων του μοντέλου μπορεί να δημιουργήσει μια εικόνα που είναι όσο γίνεται πιο όμοια με την πραγματική.

Η μεταβλητότητα που παρουσιάζουν τα αντικείμενα όπως είναι το ανθρώπινο πρόσωπο μπορεί να αντιμετωπιστεί. Αρκετά χρήσιμα είναι τα παραμορφώσιμα μοντέλα. Τα μοντέλα αυτά έχουν την δυνατότητα να διατηρούν τα βασικά χαρακτηριστικά της κατηγορίας των αντικειμένων που αυτά αναπαριστούν. Επιπλέον, όμως, μπορούν να παραμορφωθούν σε τέτοιο βαθμό ώστε να ταιριάζουν σε ένα μεγάλο φάσμα από στιγμιότυπα. Αυτού του είδους τα μοντέλα έχουν δύο κύρια χαρακτηριστικά. Πρώτον, να είναι γενικά ώστε να παράγουν το κάθε πιθανό στιγμιότυπο της κατηγορίας στην οποία ανήκουν. Δεύτερον, να είναι ειδικά ώστε τα στιγμιότυπα που παράγονται να είναι αποδεκτά στιγμιότυπα της κατηγορίας. Η δεύτερη αυτή ιδιότητα που έχουν είναι και η περισσότερο σημαντική, αφού όλο το νόημα της χρήσης των τεχνικών που στηρίζονται στα μοντέλα είναι να επικεντρώνεται το σύστημα αυτό σε αποδεκτά στιγμιότυπα.

Για να δημιουργηθεί ένα στατιστικό μοντέλο, απαιτείται η ανάλυση της δομής του συνόλου παραδειγμάτων που είναι το σύνολο εκπαίδευσης (training dataset). Τα παραδείγματα έχουν διαφορές στο σχήμα και την υφή τους, και αυτό έχει ως αποτέλεσμα το σύστημα να μπορεί να εκπαιδευτεί σε σχέση με το ποιες αλλαγές θεωρούνται αποδεκτές και ποιες μη αποδεκτές.

Τα πλεονεκτήματα μιας τέτοιας μεθόδου είναι:

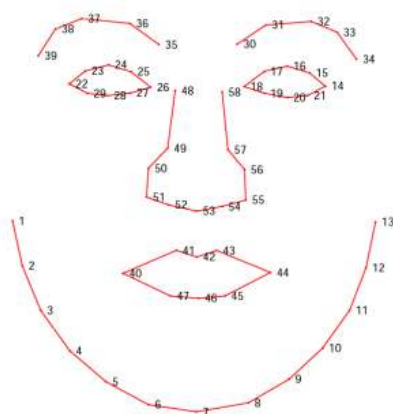
- Έχει ευρεία εφαρμογή. Μπορεί να χρησιμοποιηθεί ο ίδιος αλγόριθμος σε πολλές διαφορετικές περιπτώσεις προβλημάτων με τη χρήση διαφορετικών παραδειγμάτων για τη δημιουργία του μοντέλου.
- Το σύστημα που δημιουργείται αποκτά γνώση που σχετίζεται με τα παραδείγματα με τα οποία εκπαιδεύεται.
- Τα μοντέλα επιτρέπουν αυθαίρετες μεταβολές ακόμα και όταν αυτές δεν περιέχονται στα δεδομένα του συνόλου εκπαίδευσης.
- Το σύστημα δεν κάνει αρχικές υποθέσεις σχετικά με τα αντικείμενα που μελετώνται, παρά μόνο αυτά που μαθαίνει από τα δεδομένα του συνόλου εκπαίδευσης.

Απαραίτητη είναι η ύπαρξη χαρακτηριστικών σημείων (landmark points) σε κάθε παράδειγμα του συνόλου εκπαίδευσης. Πρέπει τα σημεία αυτά να υπάρχουν σε όλα τα παραδείγματα του συνόλου εκπαίδευσης και, επιπλέον, να αποτελούν μέρος της δομής της εικόνας που εξετάζεται. Για παράδειγμα, αν εξετάζεται το ανθρώπινο πρόσωπο, landmark points θεωρούνται οι δύο γωνίες του ματιού. Η τοπολογία και η δομή των αντικειμένων που υπάρχει στην εύρεση καλών landmark points πρέπει να μη μεταβάλλεται.

Τα ενεργά μοντέλα εμφάνισης βασίζονται σε ένα μοντέλο που είναι ικανό να εξάγει νέες εικόνες του αντικειμένου που εξετάζεται. Ο αλγόριθμος ενεργού μοντέλου εμφάνισης (AAM) αναζητά εκείνες τις παραμέτρους του μοντέλου που παράγουν μια συνθετική εικόνα που να είναι όσο γίνεται πιο όμοια με την αρχική εικόνα.

3.2 Στατιστικά Μοντέλα Σχήματος

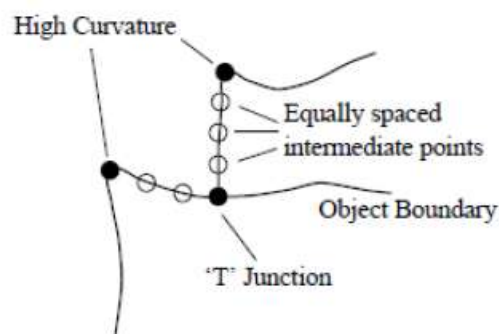
Τα στατιστικά μοντέλα σχήματος [6] είναι χρήσιμα για την αναπαράσταση αντικειμένων στις εικόνες. Η βάση των μοντέλων σχήματος είναι τα 58 landmark points του προσώπου όπως φαίνονται στην εικόνα 5. Το σχήμα του αντικειμένου αποτελείται από n συνολικά σημεία στο χώρο.



Εικόνα 5: Τα landmark points του προσώπου

Για τη δημιουργία του μοντέλου, αρχικά δημιουργείται το σύνολο εκπαίδευσης με προσημειωμένα τα landmark points στα δεδομένα που χρησιμοποιούνται για την εκπαίδευσή τους. Ο πιο απλός τρόπος να γίνει αυτό είναι η χειροκίνητη σημείωση τους, όμως αυτό απαιτεί χρόνο, για αυτό το λόγο χρησιμοποιούνται ημι-αυτόματες διαδικασίες ορισμού των σημείων landmarks.

Όσον αφορά στις δύο διαστάσεις, επιλέγονται σημεία των γωνιών, δηλαδή οι διασταυρώσεις των ορίων τους, ή τα ευδιάκριτα σημεία των δομών του αντικειμένου. Στην περίπτωση που τα σημεία δεν είναι αρκετά για να περιγραφεί το σχήμα του αντικειμένου που εξετάζεται, προστίθενται και άλλα σημεία του περιγράμματος που είναι κατανεμημένα ομοιόμορφα στα καλά ορισμένα σημεία που αρχικά έχουν επιλεγεί (Εικόνα 5). Επιλέγοντας συνολικά σημεία σε d διαστάσεις, τότε προκύπτει ένα διάνυσμα x με μέγεθος nd . Για παράδειγμα $x=(x_1, \dots, x_n, y_1, \dots, y_n)^T$, στις 2 διαστάσεις. Αν το dataset εκπαίδευσης αποτελείται από s δεδομένα, ορίζονται s διανύσματα x_j . Τα διανύσματα αυτά χρησιμοποιούνται στη στατιστική ανάλυση από την οποία θα προκύψει το μοντέλο.



Εικόνα 6: Καλά ορισμένα landmark points

Για να αναλυθούν τα διανύσματα, τα σχήματα ευθυγραμμίζονται και, τότε, η μεταβλητότητά τους, που τυχόν να υπάρχει ως αποτέλεσμα των μετασχηματισμών, αφαιρείται. Είναι πολλές οι μέθοδοι με τις οποίες μπορεί ένα σχήμα να ευθυγραμμιστεί σε ένα κοινό σύστημα συντεταγμένων. Η πιο προσεγγιστική μέθοδος είναι η ανάλυση Προκρούστη. Όλα τα σχήματα ευθυγραμμίζονται, με αποτέλεσμα να ελαχιστοποιείται το άθροισμα της απόστασης του κάθε σχήματος από το μέσο σχήμα ($D=\sum |x_i - \bar{x}|^2$). Μια εύκολη επαναληπτική προσέγγιση είναι:

1. Μετατόπιση κάθε εικόνας, με το κέντρο βάρους της να είναι στην αρχή των αξόνων.

2. Επιλογή μιας εικόνας σαν αρχική εκτίμηση του μέσου σχήματος και κλίμακας $|\bar{x}|=1$.
3. Σημείωση της πρώτης εκτίμησης ως \bar{x} ώστε να οριστεί το καρέ αναφοράς.
4. Όλα τα σχήματα σε ευθυγράμμιση σε σχέση με την τρέχουσα εκτίμηση του μέσου σχήματος.
5. Για τα ευθυγραμμισμένα σχήματα επανεκτίμηση του μέσου.
6. Εφαρμόζονται περιορισμοί στην τρέχουσα εκτίμηση του μέσου και ευθυγραμμίζονται ώστε $|\bar{x}|=1$.
7. Επιστροφή στο βήμα 4 έως ότου να προκύψει σύγκλιση. (Σύγκλιση υπάρχει όταν η εκτίμηση του μέσου παραμένει ίδια ή χωρίς μεγάλες αλλαγές μετά από την επόμενη επανάληψη).

Στην συνέχεια, γίνεται η μοντελοποίηση της μεταβλητότητας των σχημάτων. Από τα προηγούμενα βήματα, υπάρχει το σύνολο s διανυσμάτων που έχουν κατανομηθεί ομοιόμορφα στο χώρο διαστάσεων nd . Έπειτα από μοντελοποίηση της κατανομής, παράγονται νέα δεδομένα σχεδόν όμοια με αυτά του συνόλου της εκπαίδευσης και εξετάζονται νέα σχήματα για να αποφασιστεί εάν είναι αποδεκτά. Δηλαδή, γίνεται αναζήτηση ενός μοντέλου της μορφής $x=M(b)$, όπου b το διάνυσμα των παραμέτρων του μοντέλου. Η μεταβολή των παραμέτρων b μπορεί να παράγει νέα διανύσματα x . Η μοντελοποίηση της κατανομής $p(b)$ των παραμέτρων μπορεί να περιοριστεί ούτως ώστε τα διανύσματα x που έχουν παραχθεί να είναι παρόμοια με εκείνα του συνόλου εκπαίδευσης.

Για λόγους απλότητας, μειώνονται οι διαστάσεις του παραπάνω προβλήματος. Μία προσέγγιση είναι να εφαρμοστεί Principal Component Analysis (PCA) στα δεδομένα που αποτελούν ένα σύννεφο σημείων στον nd -διάστατο χώρο. Η PCA υπολογίζει κύριους άξονες του σύννεφου αυτού, δίνοντας τη δυνατότητα προσέγγισης οποιουδήποτε αρχικού σημείου με τη χρήση ενός μοντέλου που αποτελείται από λιγότερες από nd παραμέτρους.

Τα μοντέλα σχημάτων αποτελούνται από γραμμικούς συνδυασμούς σχημάτων από το σύνολο της εκπαίδευσης. Για καλό ταίριασμα σε ένα νέο σχήμα, το σύνολο εκπαίδευσης πρέπει να αναπαριστά όλη την μεταβλητότητα της κλάσης του σχήματος που μοντελοποιείται. Σε αντίθετη περίπτωση, το μοντέλο αυτό δεν είναι επαρκώς γενικό για να ταιριάζει με το σύνολο των αποδεκτών παραδειγμάτων. Παράδειγμα, όταν το μοντέλο εκπαιδεύεται μόνο σε τετράγωνα, δε θα έχει την ικανότητα γενίκευσης σε ορθογώνιο, αφού αυτό δεν αντιπροσωπεύει το σύνολο της εκπαίδευσης.

3.3 Ενεργά μοντέλα εμφάνισης

Παρακάτω γίνεται ανάλυση του μοντέλου και αναφορά στον αλγόριθμο ενεργού μοντέλου εμφάνισης που χρησιμοποιήθηκε. Τα AAMs οφείλουν την δημιουργία τους στους Cootes, Edwards και Taylor και η παρουσίαση τους έγινε σε συνέδριο της IEEE το 1998.

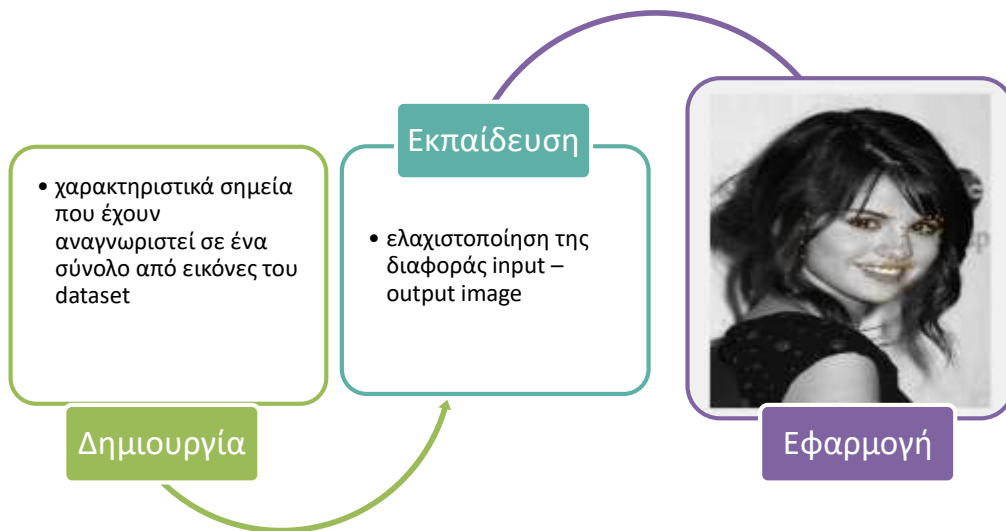
3.3.1 Γενικά

Τα ενεργά μοντέλα εμφάνισης είναι στατιστικά μοντέλα που συνδυάζουν δεδομένα από σχήμα και υφή σε κάποια εικόνα. Το AAM αντιπροσωπεύει έναν αλγόριθμο που ανήκει στον κλάδο της όρασης υπολογιστών. [5]. Η φωτεινότητα των pixels ή το χρώμα του σχήματος αποτελούν την υφή της εικόνας που ερμηνεύεται. Απαραίτητη προϋπόθεση είναι η ύπαρξη ενός dataset εκπαίδευσης δηλαδή μια βάση δεδομένων με εικόνες που έχουν προσημειωμένα σημεία (landmarks). Στην μελέτη του ανθρώπινου προσώπου τα μαρκαρισμένα σημεία βρίσκονται σε θέσεις κλειδιά περιγράφοντας τα χαρακτηριστικά του. Η δημιουργία του μοντέλου γίνεται στην φάση εκπαίδευσης (training phase). Για να δημιουργηθεί το μοντέλο σχήματος χρησιμοποιείται το ASM (στατιστικό μοντέλο σχήματος).

Τα AAMs ερευνήθηκαν από τους Cootes, Edwards και Taylor και παρουσιάστηκαν στην εργασία τους [2], που είχε ως σκοπό να εντοπίζει παραμορφώσεις αντικειμένων σε εικόνες. Τα AAMs χρησιμοποιούνται από τότε για εντοπισμό προσώπων ή και άλλων αντικειμένων μέσω ιατρικών εξετάσεων μαγνητικής τομογραφίας (MRI), όπου γίνεται ταυτοποίηση κάποιας κακοήθειας.

Οι φάσεις του ενεργού μοντέλου εμφάνισης χωρίζονται σε τρεις κατηγορίες:

- Δημιουργία
- Εκπαίδευση
- Εφαρμογή



Εικόνα 7. Φάσεις ενός AAM

Στη φάση της δημιουργίας, παράγεται ένα μοντέλο χρησιμοποιώντας κάποια χαρακτηριστικά σημεία που έχουν επισημανθεί σε ένα σύνολο από εικόνες εκπαίδευσης. Το μοντέλο αυτό εμπεριέχει τα χαρακτηριστικά του σχήματος και της υφής που ανήκουν στο μέσο πρόσωπο της βάσης εκφράσεων που χρησιμοποιείται. Έπειτα, το μοντέλο εκπαιδεύεται ώστε να μπορεί να τροποποιήσει τις παραμέτρους του για να εφαρμοστεί σε μια νέα εικόνα. Στη φάση αυτή, εφαρμόζεται το μοντέλο σε συνθετικές εικόνες, οι οποίες προκύπτουν από το ίδιο το μοντέλο, με σκοπό τον έλεγχο και τη διόρθωση των παραμέτρων του μοντέλου, ώστε να είναι ικανό να εφαρμοστεί σε μη συνθετικές εικόνες. Η τρίτη φάση είναι η φάση της εφαρμογής του μοντέλου μέσω προσαρμογής του στις εικόνες της βάσης που χρησιμοποιείται, ώστε να προκύψουν για κάθε εικόνα οι συντεταγμένες των χαρακτηριστικών σημείων που έχουν επιλεγεί.

3.3.2 Δημιουργία μοντέλου

Στη φάση της δημιουργίας του μοντέλου, παράγεται το στατικό μοντέλο το οποίο θα χρησιμοποιηθεί για εφαρμογή στις νέες εικόνες. Η δημιουργία αυτή βασίζεται σε κάποια

χαρακτηριστικά σημεία που αναγνωρίζονται από το σύνολο των εικόνων εκπαίδευσης. Τα σημεία αυτά αντιπροσωπεύουν τα σημεία που εντοπίζονται από το μοντέλο σε κάποια νέα εικόνα και τοποθετούνται με μη αυτόματο τρόπο.

Αρχικά, τα σημεία εκφράζονται σε ένα κοινό σύστημα συντεταγμένων και ευθυγραμμίζονται ώστε να αποφευχθεί η εισαγωγή σφάλματος. Ο αλγόριθμος που εκτελεί τη διαδικασία αυτή είναι αυτός της ανάλυσης του Προκρούστη.

Έπειτα, δημιουργείται ένα μοντέλο σχήματος. Τα σημεία που έχουν επιλεγεί σε μία εικόνα σχηματίζουν ένα διάνυσμα v :

$$v = \begin{bmatrix} x1 \\ y1 \\ x2 \\ y2 \\ \cdot \\ \cdot \\ \cdot \\ xN \\ yN \end{bmatrix} \quad (1.1)$$

όπου x_i, y_i το ζεύγος τιμών που προσδιορίζουν τις συντεταγμένες του σημείου και N ο αριθμός των σημείων.

Η διάσταση του παραπάνω διανύσματος είναι $2N$ καθώς κάθε σημείο εκφράζεται με δύο συντεταγμένες. Έπειτα, εφαρμόζεται Ανάλυση Κύριων Συνιστωσών (PCA) στα δεδομένα. Κάθε σύνολο σημείων για μια εικόνα μπορεί να παρασταθεί ως:

$$v = \bar{u} + P_s b_s \quad (1.2)$$

Όπου \bar{u} το μέσο σχήμα, P_s ένας ορθογώνιος πίνακας που περιέχει τα ιδιοδιανύσματα των διασπορών και b_s ένα σύνολο από παραμέτρους σχήματος.

Σε κάθε παράδειγμα (σύνολο σημείων μιας εικόνας εκπαίδευσης), εφόσον έχει παραμορφωθεί το σχήμα ώστε να ταυτιστεί με το μέσο σχήμα, γίνεται δειγματοληψία της πληροφορίας υφής (συνήθως είναι σε κλίμακα του γκρι). Η πληροφορία της υφής είναι ευάλωτη στις συνθήκες του φωτισμού, επομένως εφαρμόζεται κάποιου είδους κανονικοποίηση. Μετά από την εφαρμογή Ανάλυσης Κύριων Συνιστωσών (PCA), εξάγεται το διάνυσμα υφής που μπορεί να οριστεί με την εξίσωση:

$$g = \bar{g} + P_g b_g \quad (1.3)$$

όπου \bar{g} το διάνυσμα μέσης υφής, P_g ένας ορθογώνιος πίνακας που περιέχει τα ιδιοδιανύσματα των διασπορών και b_g ένα σύνολο από παραμέτρους υφής.

Οι πληροφορίες συνοψίζονται σε ένα κοινό διάνυσμα το οποίο έχει τη μορφή:

$$b = \begin{pmatrix} W_s b_s \\ b_g \end{pmatrix} = \begin{pmatrix} W_s P_s^T (v - \bar{v}) \\ P_g^T (g - \bar{g}) \end{pmatrix} \quad (1.4)$$

όπου W_s ο διαγώνιος πίνακας με τα βάρη για κάθε παράμετρο σχημάτων.

Στα δεδομένα, ενδέχεται να υπάρχουν συσχετίσεις ανάμεσα σε αυτά του σχήματος και της υφής, και γίνεται εφαρμογή πάλι της PCA (Ανάλυση Κύριων Συνιστωσών), για να εξαχθεί ένα μοντέλο που να έχει την μορφή:

$$B = Qc \quad (1.5)$$

όπου Q ο πίνακας που εμπεριέχει ιδιοδιανύσματα και στο διάνυσμα που περιέχει πληροφορία για το σχήμα και την υφή, δηλαδή οι παράμετροι εμφάνισης.

Λόγω της γραμμικότητας του μοντέλου, μπορεί να εκφραστεί η πληροφορία του σχήματος και του χρώματος συναρτήσει του c :

$$v = \bar{v} + P_s W_s Q_s c \quad (1.6)$$

$$g = \bar{g} + P_g W_g Q_g c \quad (1.7)$$

όπου

$$\begin{pmatrix} Q_s \\ Q_g \end{pmatrix} \quad (1.8)$$

Τα δεδομένα σχήματος εκφράζουν απόσταση, ενώ τα δεδομένα υφής χρωματικότητα, επομένως δεν μπορούν να συγκριθούν άμεσα. Για να μπορούν να γίνουν ανάλογα, πρέπει να εντοπιστεί η

επίδραση του σχήματος στην υφή. Για να γίνει αυτό, μεταβάλλεται κάθε στοιχείο σχήματος στα παραδείγματα και παρατηρείται η επίδραση αυτής της μεταβολής στα δεδομένα της υφής.

3.3.3 Εκπαίδευση μοντέλου

Η χρήση των AAM αποσκοπεί στον εντοπισμό χαρακτηριστικών σημείων στις εικόνες όπου δεν έχει πραγματοποιηθεί η παραπάνω διαδικασία. Η βασική ιδέα του αλγορίθμου είναι η ελαχιστοποίηση της διαφοράς μεταξύ της αρχικής εικόνας και αυτής που συντέθηκε από το μοντέλο εμφάνισης. Μετά τη δημιουργία του μοντέλου, αυτό εκπαιδεύεται για να έχει εφαρμογή σε μια νέα εικόνα. Τροποποιείται για να εντοπίζει τα χαρακτηριστικά σημεία προσώπων που απεικονίζονται σε μια συγκεκριμένη εικόνα κάνοντας ελαχιστοποίηση ενός συγκεκριμένου δείκτη σφάλματος. Σχηματίζεται το διάνυσμα διαφορών:

$$\delta I = I_i - I_m \quad (1.9)$$

όπου I_i το διάνυσμα υφής της αρχικής εικόνας και I_m το διάνυσμα του τρέχοντος συνθετικού μοντέλου.

Για να βρεθεί το καλύτερο ταίριασμα μεταξύ του μοντέλου και της αρχικής εικόνας, ελαχιστοποιείται το διάνυσμα διαφορών $\Delta = |\delta I|^2$ μεταβάλλοντας τις παραμέτρους c του μοντέλου. Η πιο απλή προσέγγιση, που δίνει καλά αποτελέσματα, είναι η εξάρτηση αυτή να είναι γραμμική.

Δηλαδή:

$$\delta c = A\delta I \quad (1.10)$$

Για να βρεθεί ο πίνακας A εφαρμόζεται γραμμική παλινδρόμηση μεταξύ των μετατοπίσεων δc και ακολούθως των διαφορών μεταξύ των εικόνων δI . Τέτοια σύνολα προκύπτουν με τυχαία τροποποίηση των παραμέτρων του συνθετικού μοντέλου κατά γνωστό, πάντα, τρόπο, ώστε να προκύψουν συνθετικές εικόνες. Εκτός από την τροποποίηση των παραμέτρων με τυχαίο τρόπο, μοντελοποιούνται και γραμμικές μετατοπίσεις των παραμέτρων, όπως και στροφή και κλιμάκωσή τους. Με τον τρόπο αυτό, γίνεται το μοντέλο πιο εύρωστο.

Η διαδικασία που ακολουθείται περιγράφεται ως εξής: Δημιουργείται ένα συνθετικό μοντέλο με γνωστές παραμέτρους, έστω C_0 . Εισάγεται γνωστή παραμόρφωση στο μοντέλο, έστω δC .

Επομένως το νέο μοντέλο θα έχει παραμέτρους:

$$c = \delta C + C_0 \quad (1.11)$$

Με βάση αυτές τις παραμέτρους παράγεται η πληροφορία σχήματος v και υφής g_m . Με βάση τα σημεία v γίνεται δειγματοληψία της εικόνας και λαμβάνεται η πληροφορία υφής g_s . Έπειτα υπολογίζεται το σφάλμα:

$$\delta g = g_s - g_m \quad (1.12)$$

Με συνεχόμενη αλλαγή των παραμέτρων και καταγραφή των δc , δg εξάγονται τα στοιχεία του πίνακα A για να προκύψει η σχέση:

$$\delta c = A \delta g \quad (1.13)$$

3.3.4 Εφαρμογή μοντέλου σε εικόνα

Μέχρι τώρα έγινε η δημιουργία του μοντέλου και εκπαιδεύτηκε για να τροποποιεί τις παραμέτρους του και να εφαρμόζεται σε μια νέα εικόνα. Η διαδικασία της εφαρμογής του μοντέλου σε νέα εικόνα περιγράφεται παρακάτω.

Έστω c_0 οι τρέχουσες παράμετροι του μοντέλου και g_s η εικόνα που έχει υποστεί δειγματοληψία στην τρέχουσα εκτίμηση. Η επαναληπτική διαδικασία που εκτελείται σε κάθε βήμα για να ληφθεί μια καλύτερη προσέγγιση είναι η εξής:

1. Υπολογίζεται το διάνυσμα του σφάλματος

$$\delta g_0 = g_s - g_m \quad (1.14)$$

Όπου g_m το διάνυσμα που περιέχει την πληροφορία υφής του μέσου συνθετικού μοντέλου που έχει δημιουργηθεί.

2. Υπολογίζεται το συνολικό σφάλμα ως:

$$\delta E_0 = |\delta g_0|^2 (1.15)$$

3. Υπολογίζεται το διάνυσμα τροποποίησης των παραμέτρων του μοντέλου με την σχέση που έχει προκύψει παραπάνω:

$$\delta c = A \delta g_0 (1.16)$$

4. Ορίζεται η παράμετρος:

$$k = 1 \quad (1.17)$$

που ορίζει τον ρυθμό διόρθωσης παραμέτρων του μοντέλου, για να ικανοποιεί το κριτήριο της σύγκλισης που είναι η ελαχιστοποίηση σφάλματος.

5. Η επανεκτίμηση των παραμέτρων του μοντέλου πραγματοποιείται με την σχέση:

$$c_1 = c_0 - k \delta c (1.18)$$

η οποία παρέχει τις παραμέτρους του νέου μοντέλου που επιχειρεί να περιγράψει τα δεδομένα της εικόνας.

6. Γίνεται δειγματοληψία της εικόνας στα σημεία που υπολογίστηκαν και υπολογίζεται το νέο διάνυσμα σφάλματος g_1 το οποίο, όπως περιγράφηκε προηγουμένως, περιέχει την απόκλιση από την πληροφορία υφής του μέσου συνθετικού μοντέλου.

7. Υπολογίζεται το νέο συνολικό σφάλμα E_1 . Αν ισχύει η σχέση:

$$E_1 < E_0 (1.19)$$

η οποία αποτελεί τον κανόνα σύγκλισης του αλγορίθμου, τότε οι νέες παράμετροι γίνονται αποδεκτές.

8. Αλλιώς, γίνεται δοκιμή για $k = 1.5$, $k = 0.5$, $k = 0.25$ κτλ.

3.3.5 Πειραματικά αποτελέσματα εφαρμογής μοντέλου

Εφόσον έχει δημιουργηθεί και εκπαιδευτεί το μοντέλο, εφαρμόζεται σε όλες τις εικόνες της βάσης, ώστε να προκύψουν τα χαρακτηριστικά σημεία κάθε προσώπου της βάσης. Είναι προφανές ότι τα σημεία αυτά ταυτίζονται με τα χαρακτηριστικά σημεία που έχουν προσημειωθεί για ορισμένες εικόνες της βάσης και χρησιμοποιήθηκαν για τη δημιουργία του μοντέλου. Επομένως, αναλόγως της βάσης που χρησιμοποιείται, γίνεται η δημιουργία του αντίστοιχου μοντέλου.

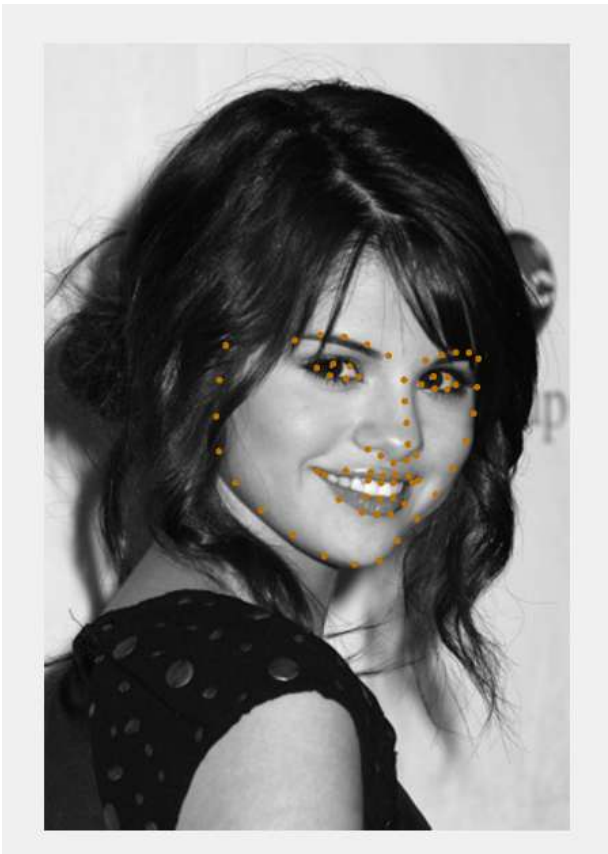


Εικόνα 8. Μαρκιασμένα σημεία του προσώπου από τη βάση εκφράσεων LFPW που χρησιμοποιείται για εκπαίδευση AAM

Ένα παράδειγμα εικόνας με προσημειωμένα χαρακτηριστικά σημεία μπορεί να φανεί στην εικόνα 8. Το πρόσωπο αυτό προέρχεται από τη βάση εκφράσεων LFPW dataset, η οποία παρέχεται δωρεάν με μια αίτηση στην αντίστοιχη σελίδα [12]. Παρουσιάστηκε για πρώτη φορά το 2011 από τους Belhumer, Jacobs, Kriegman, Kumar. Περιέχει 1,432 φωτογραφίες προσώπων κατεβασμένες από

το διαδίκτυο. Περιλαμβάνει το σύνολο εκπαίδευσης και το σύνολο δοκιμής. Οι εικόνες που χρησιμοποιούνται για να δημιουργηθεί το μοντέλο θα είναι οργανωμένες με πληροφορίες ποικιλομορφίας στα πρόσωπα, στο πόσο εκφραστικότητα έχουν, ώστε το μοντέλο που δημιουργείται να είναι αρκετά ικανό για εφαρμογή του σε νέα πρόσωπα.

Ακολουθείται ο ίδιος τρόπος αντιμετώπισης κατά την διαδικασία εφαρμογής μοντέλου. Στις περισσότερες περιπτώσεις η εφαρμογή του μοντέλου σε μια εικόνα επιφέρει τα επιθυμητά αποτελέσματα εφόσον υπολογίζει με μεγάλη ακρίβεια τα χαρακτηριστικά σημεία, όπως φαίνονται στην εικόνα 9. Το υπολογιζόμενο λάθος είναι μικρό (0,02). Άρα, η εφαρμογή του μοντέλου θεωρείται επιτυχής.



Εικόνα 9. Εύρεση χαρακτηριστικών σημείων με εφαρμογή του AAM σε ένα καρέ της βάσης LFPW

Σε κάποιες περιπτώσεις, όμως, η εφαρμογή του μοντέλου δεν είναι επιτυχής στο να εντοπίσει τα χαρακτηριστικά σημεία και αυτό οφείλεται στους παράγοντες που οφείλονται σε έντονες μετακινήσεις κεφαλιού καθώς και στροφής του. Το Σχήμα 1.3 αποτελεί ένα τέτοιο παράδειγμα.



Εικόνα 10. Παράδειγμα κακής εκτίμησης του AAM σε ένακαρέ της βάσης LFPW

Το παραπάνω φαινόμενο μπορεί να αντιμετωπιστεί με την χρήση ενός αλγορίθμου υπολογισμού της θέσης και στροφής του κεφαλιού καθώς, στην εκπαίδευση του μοντέλου, θεωρήθηκαν μόνο μικρές μετακινήσεις, τις οποίες είναι ικανό να συνυπολογίσει στην εφαρμογή του. Τα σημεία που υπολογίζονται σε εικόνες στις οποίες η εφαρμογή του μοντέλου δεν είναι ικανοποιητική, κρατούνται ως έχουν για δύο λόγους. Δεν υπάρχει κάποιος εύκολος τρόπος να εντοπιστούν ώστε να ληφθούν τα κατάλληλα αντίμετρα, και χρησιμοποιούνται ως μια μέθοδος να εξεταστεί η απόδοση του υποσυστήματος λήψης απόφασης, εάν το υποσύστημα εξαγωγής χαρακτηριστικών δημιουργεί σφάλματα.

3.3.6 Δημιουργία ακολουθιών παρατηρήσεων και ακολουθιών σημείων

Τα σημεία που υπολογίζονται κατά την εφαρμογή του μοντέλου είναι δύο διαστάσεων. Επομένως, κάθε σημείο είναι της μορφής (x_i, y_i) όπου ο δείκτης i αντιπροσωπεύει το αντίστοιχο χαρακτηριστικό σημείο. Ο αριθμός N των σημείων αυτών ισούται με τον αριθμό των σημείων που χρησιμοποιούνται στην εκπαίδευση του μοντέλου. Δημιουργείται ένα διάνυσμα παρατηρήσεων το οποίο εκπροσωπεί μια εικόνα παραθέτοντας τις συντεταγμένες όλων των σημείων σε ένα διάνυσμα στήλης το οποίο, όπως έχει αναφερθεί ξανά, θα έχει την παρακάτω μορφή:

$$v = \begin{bmatrix} x_1 \\ y_1 \\ x_2 \\ y_2 \\ \vdots \\ \vdots \\ \vdots \\ x_N \\ y_N \end{bmatrix} \quad (1.20)$$

Κάθε ακολουθία εικόνων έχει διαφορετικό μήκος (αριθμό καρτέ), έστω K . Για να δημιουργηθεί ο πίνακας που θα περιέχει τα χαρακτηριστικά μια ακολουθίας εικόνων, παρατίθενται μαζί όλα τα διανύσματα παρατηρήσεων που έχουν προκύψει από όλα τα καρτέ της ακολουθίας αυτής με τη σειρά που εμφανίζονται αυτά. Δηλαδή:

$$P = [v_1 v_2 \dots v_k] = \begin{pmatrix} x_1^1 & \dots & x_1^K \\ y_1^1 & \dots & y_1^K \\ x_2^1 & \dots & x_2^K \\ y_2^1 & \dots & y_2^K \\ \dots & & \dots \\ x_N^1 & \dots & x_N^K \\ y_N^1 & \dots & y_N^K \end{pmatrix} \quad (1.21)$$

ο δείκτης N αντιπροσωπεύει το κάθε χαρακτηριστικό σημείο, ενώ ο K δείκτης τον αριθμό καρτέ.

Επομένως, η διάσταση των δεδομένων είναι $2N$. Οι πίνακες αυτής της μορφής, οι οποίοι ονομάζονται ακολουθίες σημείων, περιέχουν τις συντεταγμένες (στο σύστημα συντεταγμένων μιας εικόνας) των χαρακτηριστικών σημείων του προσώπου που απεικονίζεται σε κάθε καρτέ μιας ακολουθίας εικόνων. Σε όλες τις εκφράσεις, τα σημεία αυτά μετατοπίζονται άλλα σε μεγαλύτερο και άλλα σε μικρότερο βαθμό. Η μετακίνηση αυτή εξαρτάται από την κίνηση των μυών του προσώπου και είναι περίπου ίδια σε κάθε έκφραση. Μέσω της εξαγωγής χαρακτηριστικών,

αποκτούνται πληροφορίες για αυτήν την κίνηση. Οι παραπάνω ακολουθίες δίνουν πληροφορίες για αυτήν την κίνηση και είναι εξαρτώμενες από κάθε πρόσωπο εξαιτίας των διαφορετικών φυσικών χαρακτηριστικών κάθε προσώπου, όπως για παράδειγμα του σχήματός του.

3.3.7 Εντοπισμός Προσώπου

Το πρώτο βήμα σε κάθε σύστημα επεξεργασίας εκφράσεων προσώπου είναι να εφαρμοστεί ο εντοπισμός των θέσεων στην εικόνα όπου εμφανίζονται πρόσωπα. Στα περισσότερα συστήματα, οι συνθήκες κατά τις οποίες παρακολουθείται ο άνθρωπος είναι σχετικά ελεγχόμενες. Ο συνηθέστερος τρόπος απεικόνισης είναι η εικόνα να περιέχει μόνο ένα πρόσωπο το οποίο βρίσκεται ανφάς. Έτσι, εξασφαλίζεται η παρουσία μόνο προσώπου και δίνεται ταυτόχρονα η πληροφορία της πιθανής του θέσης.

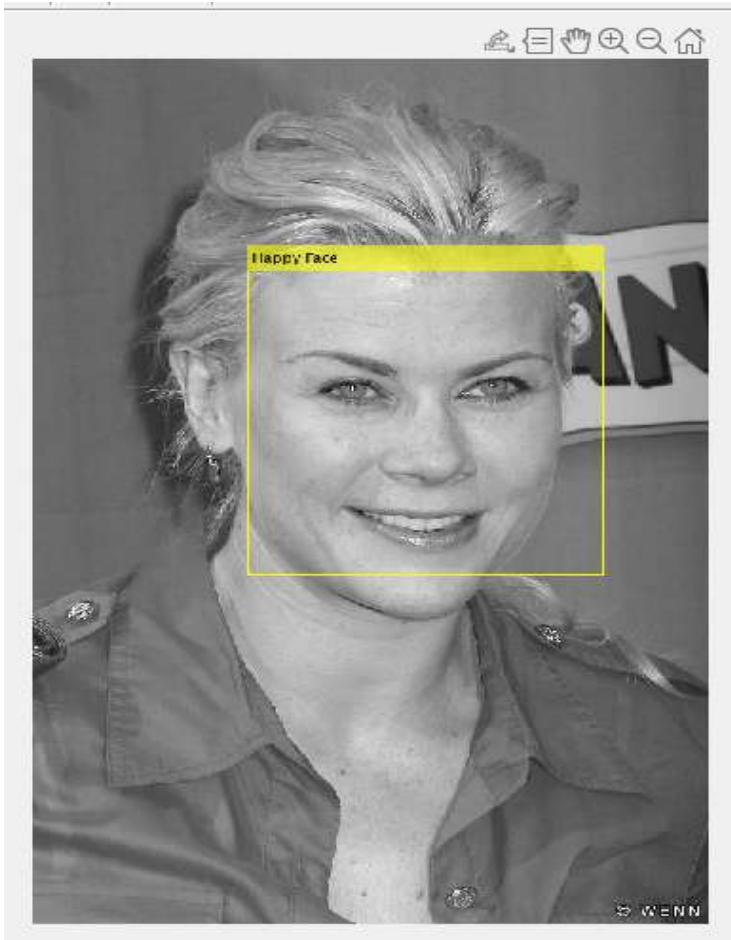
Ωστόσο, ο εντοπισμός της ακριβούς θέσης του προσώπου είναι ένα πρόβλημα σύνθετο. Από εικόνα σε εικόνα, η κλίμακα και ο προσανατολισμός του προσώπου μπορεί να διαφέρουν. Έτσι, η αναζήτηση συγκεκριμένου προτύπου εικόνας γίνεται δύσκολη. Η επικάλυψη των προσώπων από αντικείμενα και η παρουσία θορύβου κάνουν το πρόβλημα πιο δύσκολο.

3.3.8 Ανιχνευτής προσώπου Viola-Jones

Οι Viola και Jones περιέγραψαν ένα πλαίσιο εργασίας για την αναγνώριση προσώπου, το οποίο συνδυάζει τόσο τη γρήγορη επεξεργασία των εικόνων όσο και τα υψηλά ποσοστά αναγνώρισης [7]. Στο γεγονός αυτό συνεισφέρουν τα χαρακτηριστικά του συστήματος.

Προτείνεται μία νέα αναπαράσταση της εικόνας που επιτρέπει την πολύ γρήγορη αποτίμηση της αναγνώρισης προσώπου ή και άλλων χαρακτηριστικών του. Η αναπαράσταση αυτή μπορεί να υπολογιστεί πολύ εύκολα από την εικόνα, μόνο με τη χρήση μερικών απλών γραμμών κώδικα στο Matlab.

Στην παρακάτω εικόνα, απεικονίζεται ένα πρόσωπο της βάσης δεδομένων LFPW, το οποίο είναι μαρκαρισμένο με ένα πλαίσιο. Αυτό το πλαίσιο είναι το περιγεγραμμένο ορθογώνιο της τεχνικής Viola-Jones και αντιστοιχεί στην αναγνώριση προσώπου. Με τον ίδιο τρόπο, μπορούμε να απομονώσουμε και άλλα στοιχεία όπως μάτια, μύτη, στόμα.



Εικόνα 11. Τοποθέτηση πλαισίου για αναγνώριση προσώπου

4. Τεχνητή νοημοσύνη

Η τεχνητή νοημοσύνη καλύπτει ένα φάσμα επιμέρους πεδίων που συμπεριλαμβάνει την προσομοίωση της ανθρώπινης σκέψης και την εκπαίδευση, καθώς επίσης και εργασίες σύνθετες όπως μαθηματικά και αναγνώριση συναισθημάτων [8]. Η τεχνητή νοημοσύνη θεωρείται ως αυτόματος τρόπος σκέψης. Σημειώνονται στον ακόλουθο πίνακα [1] ορισμοί που έχουν δοθεί για την τεχνητή νοημοσύνη.

| | |
|---|---|
| Συστήματα που σκέπτονται σαν τον άνθρωπο. | Συστήματα που σκέπτομαι ορθολογικά. |
| « Η συναρπαστική νέα προσπάθεια για να κάνουμε τους υπολογιστές να σκέπτονται. » (Haugeland, 1985) | « Η μελέτη των νοητικών ικανοτήτων με τη χρήση υπολογιστικών μοντέλων. » (Charniak και McDermott, 1985) |
| « Η αυτοματοποίηση των δραστηριοτήτων που σχετίζονται με την ανθρώπινη σκέψη, όπως η λήψη αποφάσεων, η επίλυση προβλημάτων και η μάθηση. » (Bellman, 1978) | « Η μελέτη των υπολογιστικών εργασιών που μας δίνουν τη δυνατότητα να αντιλαμβανόμαστε και να ενεργούμε. » (Winston, 1992) |
| Συστήματα που ενεργούν σαν τον άνθρωπο. | Συστήματα που ενεργούν ορθολογικά |
| « Η τέχνη της δημιουργίας μηχανών που πραγματοποιούν λειτουργίες που απαιτούν νοημοσύνη όταν πραγματοποιούνται από ανθρώπους. » (Kurzweil, 1990) | « Υπολογιστική νοημοσύνη είναι η μελέτη της σχεδίασης ευφυών πρακτόρων. » (Poole, 1998) |
| « Η μελέτη του πως μπορούμε να κάνουμε τους υπολογιστές να κάνουν πράγματα, στα οποία, προς το παρόν, οι άνθρωποι είναι καλύτεροι. » (Rick και Knight, 1991) | « Η τεχνητή νοημοσύνη ασχολείται με την ευφυή συμπεριφορά των τεχνουργημάτων. » (Nilsson, 1998) |

Πίνακας 1. Περιγραφές τεχνητής νοημοσύνης

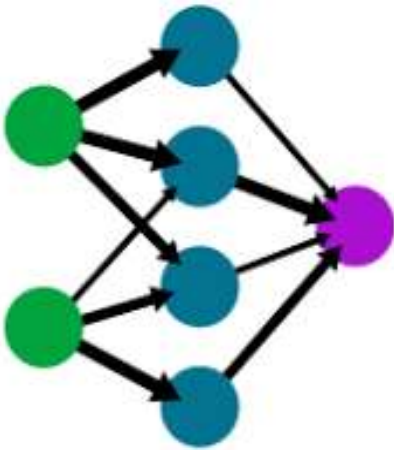
Ο T. M. Mitchell περιγράφει την τεχνητή νοημοσύνη ως εξής: «Ένα πρόγραμμα υπολογιστικό μαθαίνει από την εμπειρία E σχετικά με μια διεργασία T και αποδίδει το P όταν το T, με βάση το P, γίνεται καλύτερο με εμπειρία E» .

4.1 Μηχανική μάθηση

Η μηχανική μάθηση συμπεριλαμβάνεται στην τεχνητή νοημοσύνη. Με σωστή μηχανική μάθηση από δεδομένα γίνεται η διοχέτευση πληροφοριών και η βελτίωσή τους αυτοματοποιημένα [9]. Διακρίνουμε μηχανική μάθηση με επίβλεψη, χωρίς επίβλεψη και ενισχυτική.

4.2 Νευρωνικά δίκτυα

Το νευρωνικό δίκτυο συναντά τη λογική του ανθρώπινου νευρικού συστήματος και συνίσταται από κόμβους που συνδέονται. Οι κόμβοι συνιστούν τα θεμέλια του δικτύου αυτού. Οι εισοδοί είναι διάφορες πληροφορίες και οι έξοδοι υπολογίζονται με την κατάλληλη τροφοδότηση εισόδου [10]. Έπειτα, η έξοδος μπορεί να διοχετευθεί ως είσοδος σε άλλο δίκτυο και στη συνέχεια σε άλλο.



Εικόνα 12. Απλό νευρωνικό δίκτυο

Οι έρευνες στο επιστημονικό πεδίο που αφορά στα νευρωνικά δίκτυα βασίζονται στο ότι ο ανθρώπινος εγκέφαλος αντιλαμβάνεται και υπολογίζει πληροφορίες διαφορετικά από έναν ηλεκτρονικό υπολογιστή. Ο εγκέφαλος του ανθρώπου αποτελεί πεδίο έρευνας και θεωρείται περίπλοκος σε σχέση με την αντίληψη των πληροφοριών. Η λειτουργία του εγκεφάλου είναι η οργάνωση των νευρώνων με τρόπο που να εκτελούνται συγκεκριμένοι υπολογισμοί με υψηλότερη ταχύτητα από τον ταχύτερο ηλεκτρονικό υπολογιστή του σήμερα.

Τα νευρωνικά συστήματα είναι τομέας αιχμής και μπορούν να προσαρμοστούν αναλόγως του περιβάλλοντός τους. Η λειτουργία των νευρώνων ως μονάδες επεξεργασίας πληροφοριών στον ανθρώπινο εγκέφαλο είναι εξίσου σημαντική όσο και για τα νευρωνικά δίκτυα που αποτελούνται από τεχνητούς νευρώνες. Το νευρωνικό δίκτυο αποτελεί μία μηχανή που είναι σχεδιασμένη για να προσομοιώνει τους τρόπους με τους οποίους ο εγκέφαλος εκτελεί συγκεκριμένη εργασία και λειτουργία.

Ένα τεχνητό νευρωνικό δίκτυο μπορεί να οριστεί ως ένας παράλληλος επεξεργαστής κατανεμημένης αρχιτεκτονικής. Ο υπολογιστής αυτός αποτελείται από απλές επεξεργαστικές μονάδες και μπορεί να αποθηκεύσει γνώση από εμπειρίες καθώς επίσης και να τη διαθέτει για κάθε μελλοντική χρήση. Μοιάζει με τον εγκέφαλο του ανθρώπου στα εξής δύο σημεία:

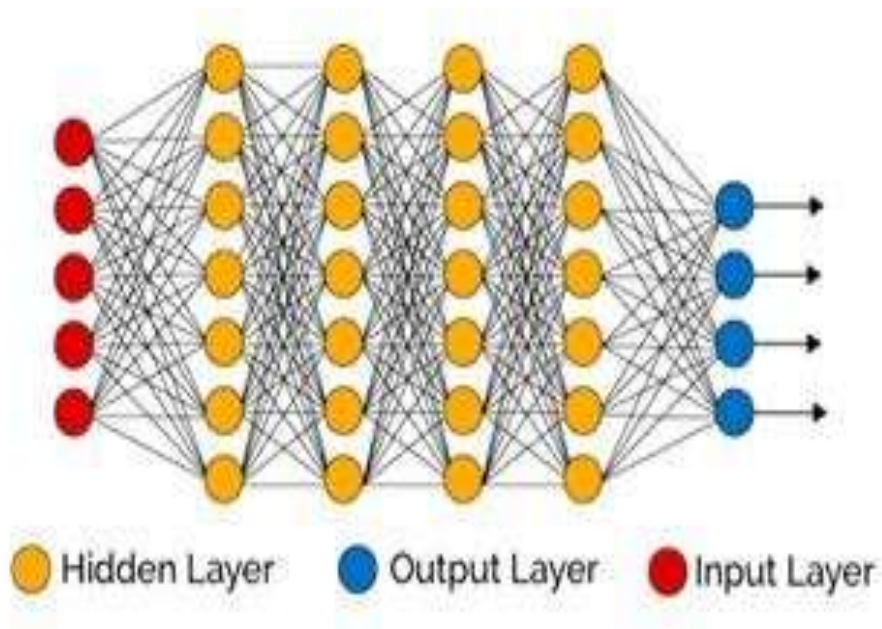
1. Το δίκτυο λαμβάνει από το περιβάλλον του τη γνώση, μέσω της διαδικασίας μάθησης.
2. Η ισχύς που έχουν οι συνδέσεις των νευρώνων μεταξύ τους είναι το συναπτικό βάρος και αποθηκεύει τη γνώση που αποκτάται.

Η διαδικασία που χρησιμοποιείται για την επίτευξη της μάθησης είναι ο αλγόριθμος μάθησης και η λειτουργία του είναι η τροποποίηση των συναπτικών βαρών του δικτύου με κατάλληλο τρόπο ώστε να επιτευχθεί ο επιθυμητός στόχος.

Ένα νευρωνικό δίκτυο αντιστοιχεί σε μεγάλη υπολογιστική ισχύ, λόγω παράλληλης, κατανεμημένης δομής και έχει την ικανότητα να μαθαίνει και να γενικεύει. Γενίκευση του νευρωνικού δικτύου υπάρχει κατά την παραγωγή λογικών εξόδων για εισόδους που δεν συνάντησε κατά τη φάση εκπαίδευσής του. Αυτές οι δυνατότητες κάνουν τα νευρωνικά δίκτυα ικανά στο να βρίσκουν προσεγγιστικές και σωστές λύσεις στα περίπλοκα προβλήματα που μπορεί να μην είναι επιδεκτικά λύσεων.

4.3 Βαθιά νευρωνικά δίκτυα

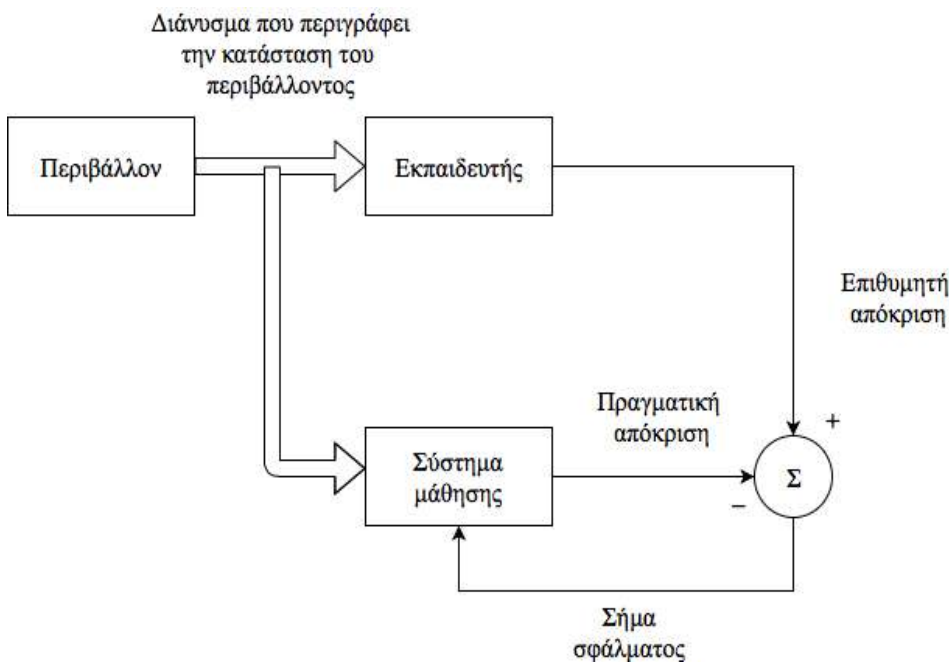
Βαθιά μηχανική μάθηση είναι η διαδικασία με την οποία εφαρμόζονται μαθησιακές διεργασίες σε νευρωνικά δίκτυα πολλών επιπέδων. Είναι μέρος μίας οικογένειας μεθόδων μηχανικής μάθησης βασισμένων σε αναπαράσταση δεδομένων, εν αντιθέσει με αλγορίθμους που επικεντρώνονται σε υπολογιστικές εργασίες [11]. Αρχιτεκτονικές βαθιάς μηχανικής μάθησης βρίσκουν εφαρμογή σε πολλούς τομείς της τεχνητής νοημοσύνης, όπως είναι η αναγνώριση φωνής, η υπολογιστική όραση και η βιοϊατρική.



Εικόνα 13. Μορφή ενός πολυεπίπεδου νευρωνικού δικτύου

4.4 Επιβλεπόμενη μάθηση νευρωνικών δικτύων

Η επιβλεπόμενη μάθηση είναι βασισμένη στην υπόθεση ότι ένα σύνολο δεδομένων με ετικέτες (labels) είναι διαθέσιμο κατά τη διάρκεια εκπαίδευσης του νευρωνικού δικτύου. Το διάγραμμα 1 περιγράφει αυτή τη μορφή μάθησης.



Σχήμα 1. Σχηματικό διάγραμμα της μάθησης με εκπαίδευση

Ο εκπαιδευτής οφείλει να γνωρίζει το περιβάλλον του και η γνώση αυτή είναι το σύνολο των παραδειγμάτων εισόδου-εξόδου. Άλλωστε, το περιβάλλον θεωρείται άγνωστο προς το νευρωνικό δίκτυο. Χάρη στη γνώση που έχει ήδη, ο εκπαιδευτής είναι σε θέση να παρέχει στο νευρωνικό δίκτυο μία επιθυμητή απόκριση για το κάθε διάνυσμα εκπαίδευσης. Η απόκριση αποτελεί τη βέλτιστη ενέργεια που πρέπει να εκτελείται από το νευρωνικό δίκτυο. Οι παράμετροι του δικτύου επηρεάζονται από το διάνυσμα εκπαίδευσης και το σήμα σφάλματος. Το σήμα σφάλματος είναι η διαφορά της επιθυμητής απόκρισης από την πραγματική απόκριση του δικτύου. Η προσαρμογή γίνεται με επανάληψη, βήμα προς βήμα, που έχει ως στόχο να θέσει τελικά το νευρωνικό δίκτυο σε κατάσταση που θα προσομοιώσει συμπεριφορές του εκπαιδευτή. Η προσομοίωση θεωρείται

βέλτιστη με στατιστική έννοια. Ο εκπαιδευτής διοχετεύει γνώση που την μεταφέρει στο νευρωνικό δίκτυο μέσω της διαδικασίας της εκπαίδευσης και, εν συνεχεία, αποθηκεύεται με τη μορφή σταθερών συναπτικών βαρών, που αντιπροσωπεύουν τη μακροπρόθεσμη μνήμη.

Η διαδικασία της επιβλεπόμενης μάθησης αποτελείται από ένα σύστημα ανάδρασης κλειστού βρόχου, όμως το άγνωστο περιβάλλον παραμένει εκτός βρόχου. Ως μέτρο απόδοσης για το σύστημα μπορεί να χρησιμοποιηθεί το μέσο τετραγωνικό σφάλμα ή το άθροισμα των τετραγώνων των σφαλμάτων επί του δείγματος εκπαίδευσης, ορισμένο ως συνάρτηση των ελεύθερων παραμέτρων (δηλαδή των συναπτικών βαρών) του συστήματος. Αυτή η συνάρτηση είναι μία πολυδιάστατη επιφάνεια σφάλματος-απόδοσης, ή μία απλή επιφάνεια σφάλματος που έχει τις ελεύθερες παραμέτρους ως συντεταγμένες. Η πραγματική επιφάνεια σφάλματος είναι ο μέσος όρος επί όλων των παραδειγμάτων εισόδου-εξόδου. Όποια άλλη λειτουργία του συστήματος υπό την επίβλεψη του εκπαιδευτή αναπαρίσταται σαν ένα σημείο της επιφάνειας του σφάλματος.

Για να έχει τη δυνατότητα το σύστημα να βελτιώνει την απόδοσή του με το πέρας του χρόνου και να εκπαιδευτεί μέσω του εκπαιδευτή, το σημείο της συνάρτησής του μετακινείται προς τα κάτω σταδιακά προς ένα ελάχιστο σημείο της επιφάνειας του σφάλματος. Αυτό το ελάχιστο σημείο είναι τοπικό ελάχιστο ή ολικό ελάχιστο. Ένα σύστημα επιβλεπόμενης μάθησης μπορεί να το καταφέρει αυτό διαθέτοντας τη χρήσιμη πληροφορία του που σχετίζεται με την κλίση της επιφάνειας του σφάλματος, το οποίο αντιστοιχεί στην τρέχουσα συμπεριφορά του συστήματος. Η κλίση υπολογίζεται από την παράγωγο του σφάλματος.

4.5 Συνελκτικά νευρωνικά δίκτυα

Το συνελκτικό δίκτυο αποτελείται από μία αρχιτεκτονική πολλών νευρωνικών επιπέδων, η οποία είναι κατάλληλα σχεδιασμένη για να αναγνωρίζει με υψηλή ακρίβεια δισδιάστατα σχήματα αμετάβλητα σε κλιμάκωση, μετατόπιση, στρέβλωση και άλλου είδους παραμορφώσεις. Αυτή η εργασία γίνεται με επιβλεπόμενο τρόπο, μέσω δικτύου που η δομή του διαμορφώνεται με κάποιους περιορισμούς:

1. Εξαγωγή χαρακτηριστικών: Οι είσοδοι των συνάψεων λαμβάνονται από το τοπικό δεκτικό πεδίο κάθε νευρώνα του προηγούμενου επιπέδου, που το υποχρεώνουν να εξάγει τοπικά χαρακτηριστικά. Εφόσον γίνει εξαγωγή ενός χαρακτηριστικού, η θέση του δεν είναι τόσο σημαντική, αφού διατηρεί προσεγγιστικά τη σχετική του θέση σε σχέση με άλλα χαρακτηριστικά.

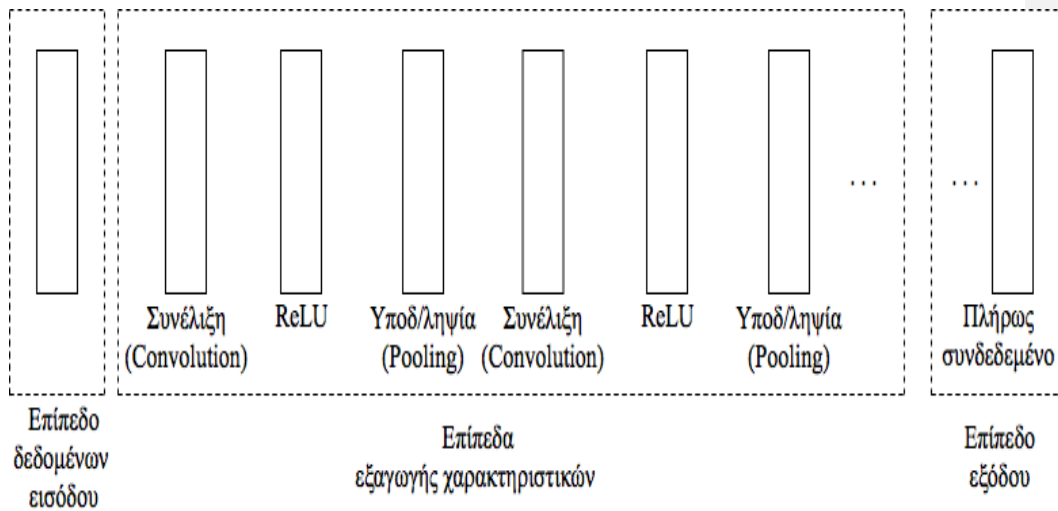
2. Αντιστοίχιση χαρακτηριστικών: Οι πολλαπλοί χάρτες χαρακτηριστικών αποτελούν το υπολογιστικό επίπεδο του δικτύου όπου ο κάθε χάρτης χαρακτηριστικών είναι σε μορφή ενός επιπέδου μέσα στο οποίο γίνεται ο έλεγχος των μεμονωμένων νευρώνων για τον διαμοιρασμό του ίδιου συνόλου συναπτικών βαρών. Αυτή η μορφή δομικού περιορισμού έχει τα εξής επακόλουθα:

- i. δεν έχει ευαισθησία στη μετατόπιση που επιβάλλεται για τη λειτουργία μιας κατηγορίας χαρακτηριστικών με τη βοήθεια της συνέλιξης και ενός πυρήνα μικρού μεγέθους, που ακολουθείται από την συνάρτηση ενεργοποίησης (πχ. ReLU).
- ii. μειώνεται ο αριθμός των ελεύθερων παραμέτρων, που επιτυγχάνεται κατά τον διαμοιρασμό βαρών.

3. Υποδειγματοληψία: Το κάθε συνελκτικό επίπεδο αποτελείται από ένα υπολογιστικό επίπεδο που κάνει υποδειγματοληψία. Έτσι, μειώνεται και η ανάλυση της κατηγορίας χαρακτηριστικών. Η υποδειγματοληψία μειώνει την ευαισθησία της εξόδου της κατηγορίας των χαρακτηριστικών σε μετατοπίσεις και παραμορφώσεις.

4. Εξαγωγή προβλέψεων: Στο τέλος της αρχιτεκτονικής υπάρχει ένα ή και περισσότερα πλήρως συνδεδεμένα επίπεδα για να εξάγεται η τελική πρόβλεψη.

Υπάρχουν αρκετές κατηγορίες στις αρχιτεκτονικές που βασίζονται τα συνελκτικά νευρωνικά δίκτυα και η γενική μορφή τους είναι:



Σχήμα 2.: Σχηματική περιγραφή ενός CNN

Τα τρία βασικά επίπεδα της αρχιτεκτονικής CNN είναι:

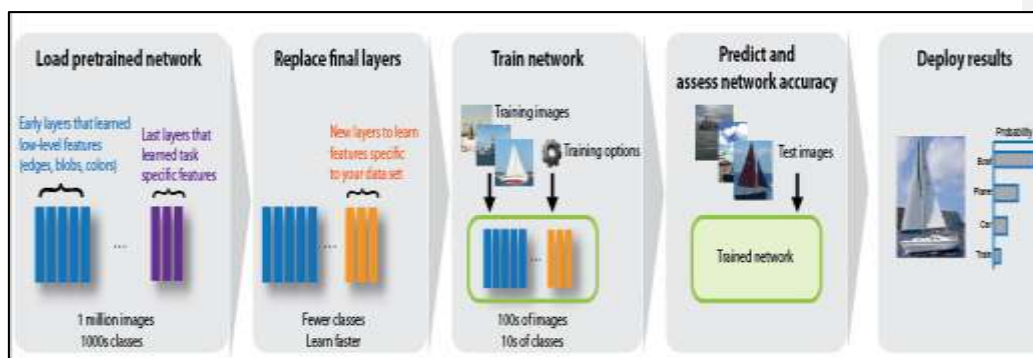
- Επίπεδο Συνέλιξης (CONV)
- Επίπεδο Υπόδειγματοληψίας (POOL)
- Πλήρως Συνδεδεμένο Επίπεδο (FC)

Κάθε επίπεδο ενός CNN λαμβάνει ως είσοδο μία πληροφορία που τη μετασχηματίζει σε μία άλλη. Τα CNN μοντελοποιούν τα μικρά αυτά τμήματα πληροφορίας και στη συνέχεια ενώνονται με σκοπό να δημιουργηθεί υψηλότερου επιπέδου πληροφορία. Για παράδειγμα, σε ένα μοντέλο CNN το πρώτο επίπεδο προσπαθεί να εντοπίσει ακμές, το δεύτερο επίπεδο παίρνει την πληροφορία των ακμών και προσπαθεί να εντοπίσει περιγράμματα, και ούτω καθεξής.

Σχόλιο [u3]: Τι μεταφράζεις εδώ;

4.6 Προ-εκπαιδευμένα Νευρωνικά δίκτυα (Pre-trained NN)

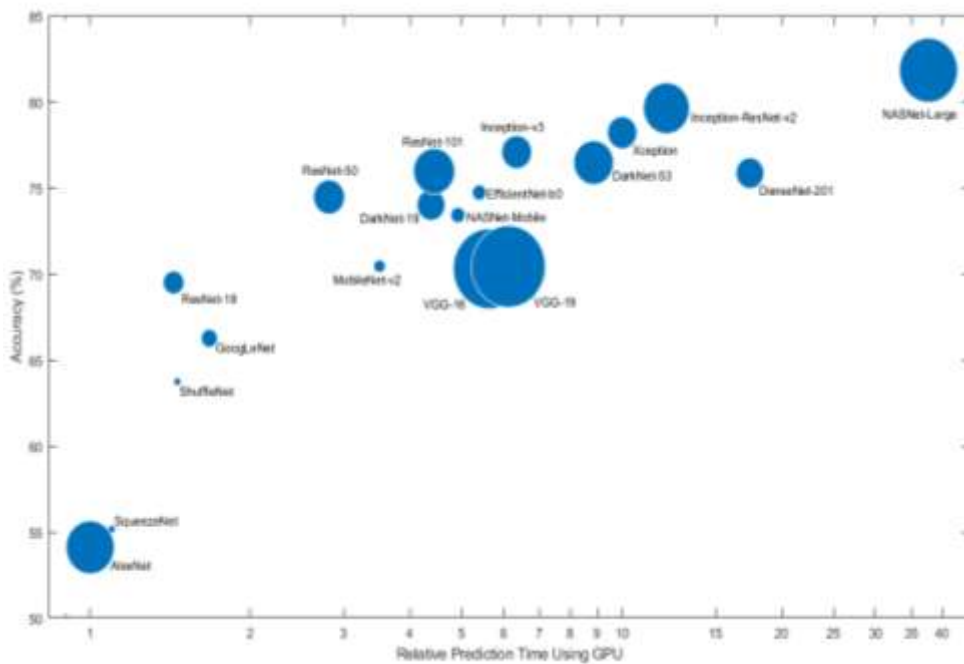
Το προ-εκπαιδευμένο νευρωνικό δίκτυο είναι ένα ήδη εκπαιδευμένο δίκτυο το οποίο χρησιμοποιείται, για παράδειγμα, με την εντολή `net=alexnet` στο MATLAB και, επεμβαίνοντας στα τελευταία Layers, εκπαιδύεται με τις επιλογές που έχουν γίνει για να εμφανίσει στο τέλος την πρόβλεψη [12].



Εικόνα 14. Διαδικασία pre-trained

Τα προ-εκπαιδευμένα νευρωνικά δίκτυα είναι δίκτυα ταξινόμησης εικόνων που ήδη έχουν μάθει να εξάγουν από εικόνες ακριβή αποτελέσματα. Η εκπαίδευση γίνεται από ένα υποσύνολο της βάσης εκφράσεων. Είναι εκπαιδευμένα σε πολλές διαφορετικές εικόνες και ταξινομούν εικόνες σε κατηγορίες. Αυτή η μέθοδος μεταφοράς μάθησης μέσω προ-εκπαιδευμένων δικτύων είναι πιο εύκολη και πιο γρήγορη από το να εκπαιδύεται το δίκτυο από την αρχή.

Στην εικόνα 15 συγκρίνονται μεταξύ τους τα προ-εκπαιδευμένα NN που κυκλοφορούν, και πόση ακρίβεια έχουν σε σχέση με το χρόνο που χρειάζονται. Στη συγκεκριμένη εργασία επιλέχθηκε το GoogleNet [13].

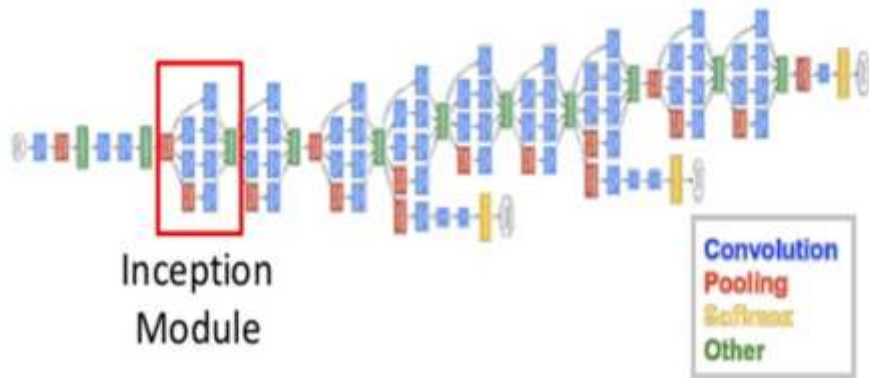


Εικόνα 15. Σύγκριση προ-εκπαιδευμένων νευρωνικών δικτύων

4.7 Μοντέλο GoogleNet

Τα βαθιά συνελκτικά δίκτυα χρησιμοποιούνται για ταξινόμηση εικόνων. Οι ερευνητές προσπαθούν συνεχώς να επεκτείνουν σε βάθος τα δίκτυα αυτά για την επίλυση περίπλοκων και πιο σύνθετων προβλημάτων. Η επέκταση σε βάθος καθιστά δύσκολη την εκπαίδευσή του δικτύου και σε αυτό χρησιμοποιείται η υπολειπόμενη μάθηση. Το GoogLeNet είναι ένα προ-εκπαιδευμένο βαθύ συνελκτικό νευρωνικό δίκτυο με πολλά επίπεδα (layers), τα οποία είναι σε στοίβες και εκπαιδεύονται. Το δίκτυο, στο τέλος των επιπέδων του, μαθαίνει χαρακτηριστικά χαμηλού, μεσαίου και υψηλού επιπέδου. Στην εικόνα 16 φαίνεται η αρχιτεκτονική GoogLeNet.

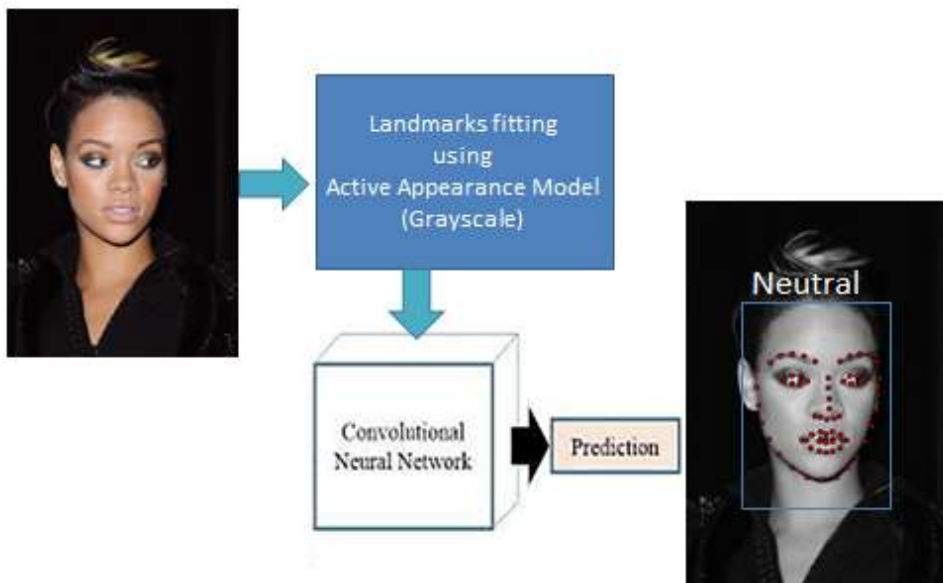
22 layers deep network



Εικόνα 16. GoogleNet επίπεδα

5. Εκπαίδευση μοντέλων και αποτελέσματα

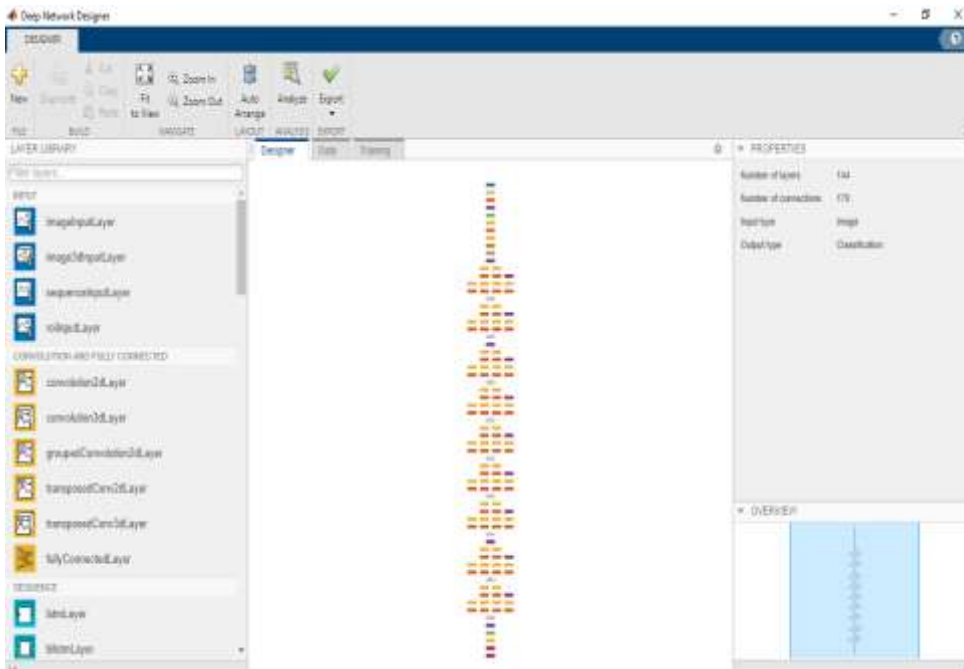
Το λογισμικό που χρησιμοποιήθηκε είναι το Matlab 2020a. Η διαδικασία εκπαίδευσης που χρησιμοποιήθηκε ήταν ίδια για όλα τα μοντέλα. Άλλαζαν οι υπερπαραμέτροι ανάλογα με τον τύπο του προβλήματος. Το διάγραμμα της εικόνας 17, που ακολουθεί, περιγράφει την πορεία εκπαίδευσης του αλγορίθμου AAM και την εισαγωγή του στο νευρωνικό δίκτυο.



Εικόνα 17. Πορεία του μοντέλου για εξαγωγή πρόβλεψης

5.1 Αρχικοποίηση μοντέλου

Στο στάδιο αυτό καθορίστηκε η αρχιτεκτονική του νευρωνικού δικτύου. Η αρχιτεκτονική του μοντέλου, όπως φαίνεται στο περιβάλλον του Matlab μέσω του Deep Network Designer, είναι η παρακάτω.

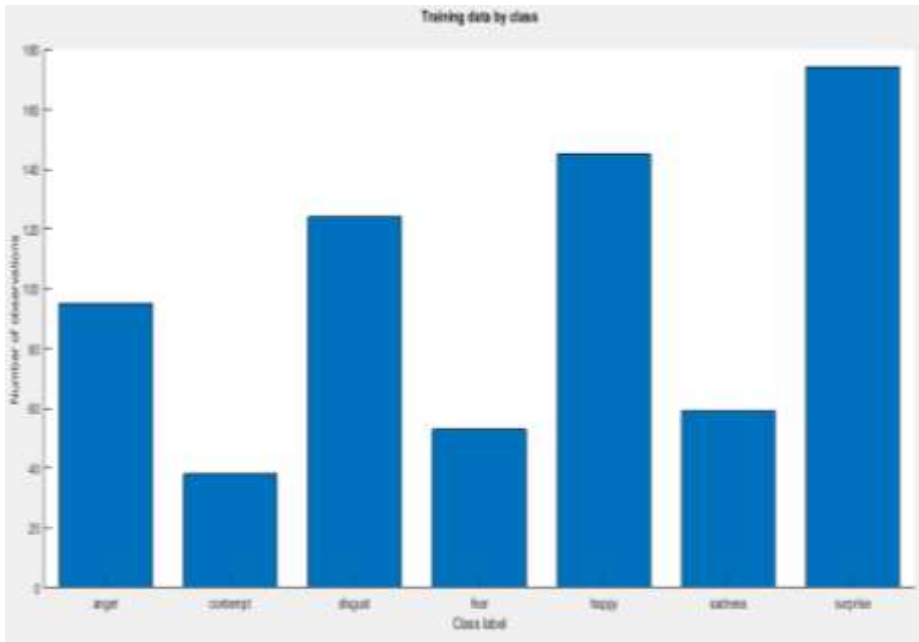


Εικόνα 18. Deep Network Designer (Matlab 2020a)

5.2 Προετοιμασία δεδομένων

Για τη διαδικασία της εκπαίδευσης τα δεδομένα χωρίστηκαν σε δύο κατηγορίες: τα δεδομένα εκπαίδευσης (trainingset) και τα δεδομένα επικύρωσης (validation set). Εξαιτίας των περιορισμών της βάσης LFPW, η οποία δεν ήταν κατηγοριοποιημένη σε κλάσεις με ετικέτες, χρησιμοποιήθηκε η βάση CK+ (CohnKanade) [15][16] στο νευρωνικό δίκτυο. Η βάση αυτή παρουσιάστηκε το 2000 και περιλαμβάνει φωτογραφίες προσώπων με διαφορετικές εκφράσεις. Είναι συνολικά 981 φωτογραφίες σε φακέλους με ετικέτες. Δηλαδή χρησιμοποιήσαμε 7 κατηγορίες εκφράσεων όπως φαίνεται στην εικόνα 19.

Η εικόνα 20 δείχνει το πλήθος των εικόνων της βάσης CK+ ανά κατηγορία. Κατηγορία θυμού: 135 εικόνες, κατηγορία περιφρόνησης: 54 εικόνες, κατηγορία αηδίας: 177 εικόνες, κατηγορία φόβου: 75 εικόνες, κατηγορία χαράς: 84 εικόνες και, τέλος, η κατηγορία έκπληξης: 254 εικόνες.



Εικόνα 19. Κατηγορίες εκφράσεων της βάσης CK+

| Command Window | |
|----------------|-------|
| Label | Count |
| anger | 135 |
| contempt | 54 |
| disgust | 177 |
| fear | 75 |
| happy | 207 |
| sadness | 84 |
| surprise | 249 |

Εικόνα 20. Πλήθος κατηγοριών εκφράσεων της βάσης CK+

5.3 Εκπαίδευση μοντέλου

Στο τελευταίο στάδιο της διαδικασίας συνδυάζονται όλα τα προηγούμενα στάδια που περιγράφηκαν για να γίνει η τελική προσαρμογή του μοντέλου. Το πλήθος των εποχών

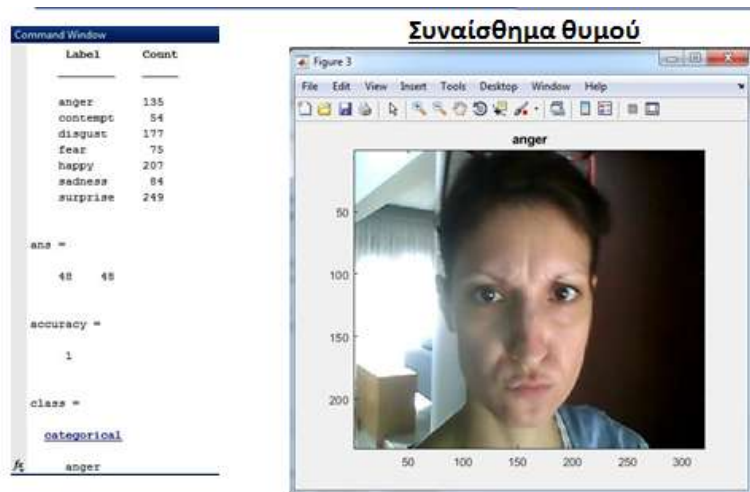
εκπαίδευσης ήταν 10 και οι επαναλήψεις 6. Στο τέλος του προγράμματος ενεργοποιείται η κάμερα, που ρυθμίστηκε κατάλληλα, και γίνεται μια λήψη για εξαγωγή της πρόβλεψης της έκφρασης του προσώπου που απεικονίζεται.

Στις παρακάτω εικόνες, φαίνεται η διαδικασία εκπαίδευσης και τα αποτελέσματα από την φωτογραφία που λήφθηκε με το άνοιγμα της κάμερας που χρησιμοποιήθηκε.



Εικόνα 21. Διαδικασία εκπαίδευσης

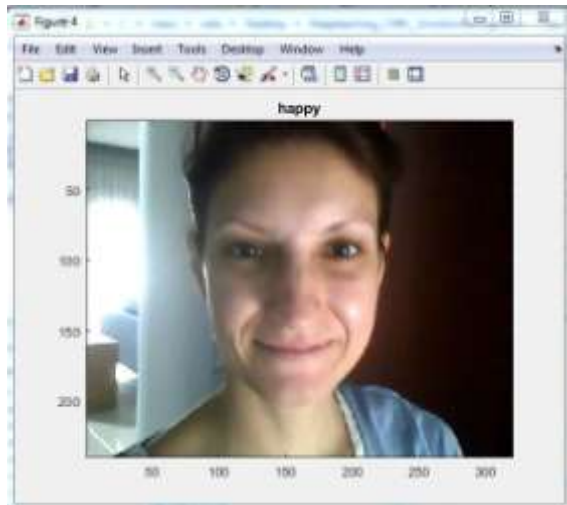
Οι παρακάτω εικόνες 22, 23, 24, 25 είναι το αποτέλεσμα που λάβαμε από την κάμερα αλλά και από την βάση δεδομένων CK+ που χρησιμοποιήθηκε [15][16].



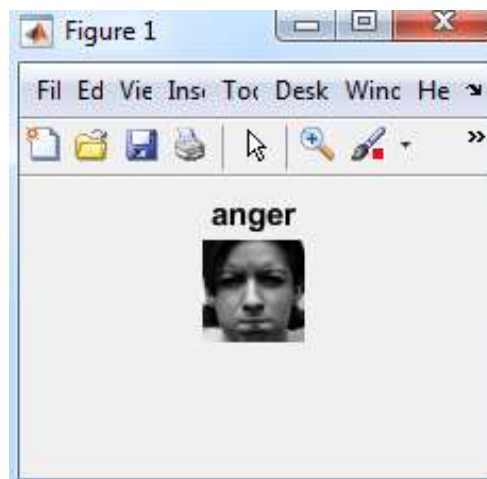
Εικόνα 22. Αποτέλεσμα λήψης με ετικέτα συναίσθηματος θυμού



Εικόνα 23. Αποτέλεσμα λήψης με ετικέτα συναίσθηματος φόβου



Εικόνα 24. Αποτέλεσμα λήψης με ετικέτα συναίσθηματος χαράς



Εικόνα 25. Αποτέλεσμα βάσης εκφράσεων CK+ με ετικέτα συναίσθηματος θυμού

6. Σύνοψη και Συμπεράσματα

Στη παρούσα εργασία παρουσιάστηκε ένα αυτόματοσύστημα για την αναγνώριση ανθρώπινων εκφράσεων. Το σύστημα αυτό μπορεί να χρησιμοποιηθεί σε εφαρμογές στις οποίες απαιτείται αναγνώριση, μοντελοποίηση και ανάλυση της συμπεριφοράς του προσώπου χωρίς καμία ανθρώπινη παρέμβαση.

Παρουσιάστηκε το ενεργό μοντέλο εμφάνισης με επιτυχές ταίριασμα του αλγορίθμου σε νέα εικόνα. Ο στόχος της αναγνώρισης της συναισθηματικής κατάστασης του ανθρώπου μέσα από εκφράσεις του προσώπου έγινε μέσα από ένα προεκπαιδευμένο νευρωνικό δίκτυο.

Μέσα στην περιοχή του προσώπου εντοπίζονται 68 χαρακτηριστικά σημεία του και με χρήση του αλγορίθμου AAM γίνεται η εφαρμογή τους σε νέα εικόνα.

Τα πειράματα που εκτελέστηκαν στον αλγόριθμο AAM με τη βοήθεια βάσης δεδομένων απέδειξαν τη σταθερότητα του μοντέλου τόσο ποιοτικά όσο και ποσοτικά. Το ποσοστό επιτυχίας ήταν αρκετά καλό, με μικρό περιθώριο λάθους. Τα δεδομένα που χρησιμοποιήθηκαν στην παρούσα εργασία περιείχαν φωτογραφίες από πλαϊνές γωνίες λήψης. Ο αλγόριθμος ενεργού μοντέλου εμφάνισης απέδωσε σωστά και σε αυτήν την περίπτωση.

Το πρόβλημα της αναγνώρισης των συναισθημάτων του ανθρώπου μέσα από εκφράσεις του προσώπου επιλύθηκε από το νευρωνικό δίκτυο. Όλα τα μοντέλα εκπαιδεύτηκαν έτσι ώστε να μεγιστοποιείται η πιθανότητα να εξάγουν ορθές προβλέψεις.

7. Μελλοντικές κατευθύνσεις

Τα μοντέλα που χρησιμοποιήθηκαν στην παρούσα εργασία είχαν καλή απόδοση, παρόλα αυτά μπορούν να επεκταθούν περισσότερο προς άλλες κατευθύνσεις ώστε το σύστημα να βελτιωθεί.

Για να είναι αποτελεσματικό το σύστημα σε μικρές αλλαγές των χαρακτηριστικών του προσώπου, θα μπορούσαμε να προσεγγίσουμε τις ανθρώπινες εκφράσεις με τη βοήθεια των FACs.

Πιθανή βελτίωση του συστήματος θα ήταν η εισαγωγή μοντέλων SVMs που, ενδεχομένως, να βελτιώνει την ακρίβεια του συστήματος και την ανοχή του σε σφάλματα.

Όσον αφορά την βάση εκφράσεων χωρίς ετικέτες, θα μπορούσαν να χρησιμοποιηθούν εναλλακτικά οι deep sparse autoencoders στο μοντέλο AAM, οι οποίοι θα έδιναν τη δυνατότητα εξαγωγής χαρακτηριστικών σημείων που θα κρίνουν την αξιολόγηση των συναισθημάτων.

Μία πιθανή επέκταση του συστήματος θα ήταν η προσπάθεια αναγνώρισης της ηλικίας του προσώπου που απεικονίζεται καθώς και το φύλλο ή η εθνικότητά του. Θα ήταν επιθυμητή η εύρεση βάσης εικόνων με ετικέτες και χαρακτηριστικά σημεία, ώστε να γίνει η εκπαίδευση του μοντέλου, καθώς η διαδικασία χειροκίνητης δημιουργίας ετικετών είναι χρονοβόρα.

Επίσης, θα μπορούσε να χρησιμοποιηθεί αρχιτεκτονική με ανατροφοδοτούμενα νευρωνικά δίκτυα (RNN). Ενισχύεται, έτσι, η πιθανότητα να επιτευχθούν καλύτερα αποτελέσματα με συστήματα υβριδικά που περιλαμβάνουν αρχιτεκτονικές CNN και RNN νευρωνικών δικτύων. Τα ανατροφοδοτούμενα νευρωνικά δίκτυα (RNN) διαθέτουν μνήμη. Τα δίκτυα αυτά έχουν την ικανότητα να εξάγουν συσχετίσεις μέσα από ακολουθίες δεδομένων, επομένως η χρήση τους μπορεί να οδηγήσει σε εξαγωγή σημαντικών χαρακτηριστικών που εκφράζουν την αλλαγή της συναισθηματικής κατάστασης του χρήστη κατά τη διάρκεια της εκάστοτε συνεδρίας.

Ένα άλλο σημείο που θα ενίσχυε την απόδοση θα ήταν η αύξηση του συνόλου δεδομένων. Στην παρούσα εργασία, το μέγεθος του εξαγόμενου συνόλου δεδομένων ήταν ικανοποιητικό, έτσι ώστε να εκτελεστούν επιτυχώς τα πειράματα και να δημιουργηθούν τα μοντέλα. Σε περίπτωση, όμως, που προστεθούν σε μελλοντική δουλειά ανατροφοδοτούμενα νευρωνικά δίκτυα, το ολικό σύστημα θα γίνει πιο σύνθετο με αποτέλεσμα να απαιτούνται περισσότερα δεδομένα. Οι τεχνικές αύξησης του συνόλου δεδομένων με χρήση των αρχικών δεδομένων (dataaugmentation) θα μπορούσαν να προσφέρουν λύσεις.

Βιβλιογραφία

- [1] Bettadapura, Vinay ‘Face expression recognition & analysis: the state of art’arXiv preprint arXiv:1203.6722 (2012)
- [2] T.F. Cootes and C.J. Taylor. Statistical Models of Appearance for Computer Vision. March 2004, University of Manchester
- [3] Mathworks help, GoogleNet, <https://www.mathworks.com/help/deeplearning/ref/googlenet.html>
- [4] Neural Networks and Learning Machines by S. O. Haykin
- [5] Wikipedia: ‘Active Appearance Models’
- [6] Wikipedia: ‘Active Shape Models’
- [7] P. Viola, M. Jones: ‘Rapid object detection with a boosted cascade of simple features’, 2001
- [8] Wikipedia: ‘Artificial Intelligence’
- [9] Wikipedia: ‘Machine Learning’
- [10] Wikipedia: ‘Neural Networks’
- [11] Wikipedia: ‘Deep Neural Networks’
- [12] Wikipedia: ‘Deep Learning’
- [13] Wikipedia: ‘GoogleNet’
- [14] LFPW dataset ‘<https://neerajkumar.org/databases/lfpw/>’
- [15] Kanade, Cohn, & Tian (2000). ‘Comprehensive database for face expressions. To the 4th International Conference of IEEE in Face & Gesture Automatic Recognition’, Grenoble, France

[16] Lucey, Cohn, Kanade, Saragih, Ambadar, & Matthews (2010). Extended Cohn-KanadeDataset (CK+): 'A complete expression dataset in actional unit & emotion-specified expression. To the 3rd International Workshop (CVPR) in Human Communicative Behavior Analysis', San Francisco

Παράρτημα κώδικας main_AAM

```
clear; clc; close all;
addpath functions

%% Train
% should you change any of the parameters below, set flag_train = 1;

addpath(genpath('trainset')) %to read subfolders from trainset folder
flag_train = 1;
where = '.';
folder = 'trainset';
%folder = fileparts(which(trainset));

what = 'png';

AAM.num_of_points = 68;
% scales refers to the resolution that fitting is taking place.
% if scale is 1, then we fit in  $1/(2^{(1-1)}) = 1$  i.e. in the original image
resolution
% if scale is 2, then we fit in  $1/(2^{(2-1)}) = 1/2$  i.e. half the original
resolution
% Multi-resolution fitting is a heuristic for improving fitting.
AAM.scales = [1 2];
% max_n and max_m refers to the number of components
% that we keep after we apply PCA on the similarity-free shapes and shape-
free textures
AAM.shape.max_n = 136;
num_of_scales = length(AAM.scales);
AAM.texture = cell(1, num_of_scales);
for ii = 1:num_of_scales
    AAM.texture{ii}.max_m = 550;
end

% Create the AAM
if flag_train
    AAM = train_AAM(where, folder, what, AAM);
    save([where '/' folder '/AAM.mat'], 'AAM');
end

%% Precompute
% This step precomputes all precomputable quantities required during fitting
% should you change any of the parameters below, set flag_precompute = 1;
% The code below creates a "chopped AAM" used in Fast-SIC algorithm
% n_all and m refers to the number of model parameters
% i.e. the number of components for the shape and texture model that we use
for fitting
% at each scale (these are usually much smaller than AAM.shape.max_n
% and AAM.texture{ii}.max_m). So these are the total number of
% parameters that Fast-SIC algorithm is aimed to recover. In the example
below we fit
% n_all = 3+4 shapes in half resolution and n_all = 10+4 shapes in the
original resolution.
% 4 is the number of similarity eigenvectors and is always fixed.
% Exactly the same applies for the texture parameters.
flag_precompute = 1;
if flag_train
    flag_precompute = 1;
end
cAAM.shape{1}.n = 10;
```

```

cAAM.shape{2}.n = 3;
cAAM.shape{1}.num_of_similarity_eigs = 4;
cAAM.shape{2}.num_of_similarity_eigs = 4;
cAAM.shape{1}.n_all = cAAM.shape{1}.n +
cAAM.shape{1}.num_of_similarity_eigs;
cAAM.shape{2}.n_all = cAAM.shape{2}.n +
cAAM.shape{2}.num_of_similarity_eigs;
cAAM.texture{1}.m = 200;
cAAM.texture{2}.m = 50;

if flag_precompute
    if ~flag_train
        load([where '/' folder '/AAM.mat']);
    end

    cAAM.num_of_points = AAM.num_of_points;
    cAAM.scales = AAM.scales;
    cAAM.coord_frame = AAM.coord_frame;

    for ii = 1:num_of_scales
        % shape
        cAAM.shape{ii}.s0 = AAM.shape.s0;
        cAAM.shape{ii}.S = AAM.shape.S(:, 1:cAAM.shape{ii}.n);
        cAAM.shape{ii}.Q = AAM.shape.Q;

        % texture
        cAAM.texture{ii}.A0 = AAM.texture{ii}.A0;
        cAAM.texture{ii}.A = AAM.texture{ii}.A(:, 1:cAAM.texture{ii}.m);
        cAAM.texture{ii}.AA0 = AAM.texture{ii}.AA0;
        cAAM.texture{ii}.AA = AAM.texture{ii}.AA(:, 1:cAAM.texture{ii}.m);

        % warp jacobian
        [cAAM.texture{ii}.dW_dp, cAAM.coord_frame{ii}.triangles_per_point] =
create_warp_jacobian(cAAM.coord_frame{ii}, cAAM.shape{ii});
    end
    save([where '/' folder '/cAAM.mat'], 'cAAM');

else
    load([where '/' folder '/cAAM.mat']);
end

%% fitting related parameters
num_of_scales_used = 2;
num_of_iter = [50 50];

%% landmark initializations
load initializations_LFPW

%% get images and ground truth shapes
names1 = dir('./testset/*.png');
names2 = dir('./testset/*.pts');

gg = 18; % choose image gg to fit
input_image = imread(['./testset/' names1(gg).name]);
pts = read_shape(['./testset/' names2(gg).name], cAAM.num_of_points);
if size(input_image, 3) == 3
    input_image = double(rgb2gray(input_image));
else
    input_image = double(input_image);
end

%% ground_truth

```



```

gt_s = (pts);
face_size = (max(gt_s(:,1)) - min(gt_s(:,1)) + max(gt_s(:,2)) -
min(gt_s(:,2)))/2;

%% initialization
s0 = cAAM.shape{1}.s0;
current_shape = scl(gg)*reshape(s0, cAAM.num_of_points, 2) +
repmat(trans(gg, :), cAAM.num_of_points, 1);
input_image = imresize(input_image, 1/scl(gg));
current_shape = (1/scl(gg))*(current_shape);
% uncomment to see initialization
% figure;imshow(input_image, []); hold on; plot(current_shape(:,1),
current_shape(:,2), '.', 'MarkerSize', 11);

%% Fitting an AAM using Fast-SIC algorithm
sc = 2.^(cAAM.scales-1);
for ii = num_of_scales_used:-1:1
    current_shape = current_shape /sc(ii);

    % indices for masking pixels out
    ind_in = cAAM.coord_frame{ii}.ind_in;
    ind_out = cAAM.coord_frame{ii}.ind_out;
    ind_in2 = cAAM.coord_frame{ii}.ind_in2;
    ind_out2 = cAAM.coord_frame{ii}.ind_out2;
    resolution = cAAM.coord_frame{ii}.resolution;

    A0 = cAAM.texture{ii}.A0;
    A = cAAM.texture{ii}.A;
    AA0 = cAAM.texture{ii}.AA0;
    AA = cAAM.texture{ii}.AA;

    for i = 1:num_of_iter(ii)

        % figure(1);clf;
        % imshow(imresize(input_image, [size(input_image, 1)/sc(ii)
size(input_image, 2)/sc(ii)]), []); hold on;
        % trimesh(cAAM.coord_frame{ii}.triangles,
current_shape(:,1),current_shape(:,2),'Color',(i/num_of_iter(ii)).*[0 1
1],'LineStyle','-');hold off;

        % Warp image
        Iw = warp_image(cAAM.coord_frame{ii}, current_shape*sc(ii),
input_image);
        I = Iw(:); I(ind_out) = [];
        II = Iw(:); II(ind_out2) = [];

        % compute reconstruction Irec
        if (i == 1)
            c = A*(I - A0) ;
        else
            c = c + dc;
        end
        Irec = zeros(resolution(1), resolution(2));
        Irec(ind_in) = A0 + A*c;

        % compute gradients of Irec
        [Irecx Irecy] = gradient(Irec);
        Irecx(ind_out2) = 0; Irecy(ind_out2) = 0;
        Irec(ind_out2) = [];
        Irec = Irec(:);

        % compute J from the gradients of Irec

```

```

        J = image_jacobian(Irecx, Irecy, cAAM.texture(ii).dW_dp,
cAAM.shape{ii}.n_all);
        J(ind_out2, :) = [];

        % compute Jfsic and Hfsic
        Jfsic = J - AA*(AA'*J);
        Hfsic = Jfsic' * Jfsic;
        inv_Hfsic = inv(Hfsic);

        % compute dp (and dq) and dc
        dq = inv_Hfsic * Jfsic*(II-AA0);
        dc = AA*(II - Irec - J*dq);

        % This function updates the shape in an inverse compositional
fashion
        current_shape = compute_warp_update(current_shape, dq,
cAAM.shape{ii}, cAAM.coord_frame{ii});
        end
        current_shape(:,1) = current_shape(:, 1) * sc(ii) ;
        current_shape(:,2) = current_shape(:, 2) * sc(ii) ;
end

figure;imshow(input_image, []); hold on; plot(current_shape(:,1),
current_shape(:,2), '.', 'MarkerSize',11);
current_shape = current_shape*scl(gg);

%% error metric used, a value of approx 0.03 shows very good fitting
pt_pt_err1 = [];
for ii = 1:cAAM.num_of_points
    pt_pt_err1(ii) = norm(gt_s(ii,:) - current_shape(ii,:));
end
pt_pt_err = mean(pt_pt_err1)/face_size

%% neural network

convnet = googlenet

% Create Image Datstore
imds = imageDatastore('gesture', ...
    'IncludeSubfolders',true,'LabelSource','foldernames')

% Split Data for Training and validation
[imdstrain, imdsvalid]=splitEachLabel(imds,.8,'randomize');

CountLabel = imds.countEachLabel
aa=read(imds);
size(aa)

% Define the Network Layers
layers = [
    imageInputLayer([size(aa), [], 1])

    convolution2dLayer(3,8,'Padding','same')
    batchNormalizationLayer
    reluLayer

```

```

maxPooling2dLayer(2,'Stride',2)

convolution2dLayer(3,16,'Padding','same')
batchNormalizationLayer
reluLayer

maxPooling2dLayer(2,'Stride',2)

convolution2dLayer(3,32,'Padding','same')
batchNormalizationLayer
reluLayer

fullyConnectedLayer(7)
softmaxLayer
classificationLayer];

% Define Option for Training

options = trainingOptions('sgdm', ...
    'InitialLearnRate',0.01, ...
    'MaxEpochs',10, ...
    'Shuffle','every-epoch', ...
    'ValidationFrequency',10, ...
    'Verbose',false, ...
    'Plots','training-progress');

% Train Model

convnet = trainNetwork(imdstrain, layers, options);

% Calaculate Accuracy using Validation Dataset

YPred = classify(convnet, imdsvalid);
YValidation = imdsvalid.Labels;

accuracy = sum(YPred == YValidation)/numel(YValidation)

% Confusion Matrix

plot(YValidation, YPred)

% Read an image from datastore and predict the class
%dimage = double(input_image);
a=imread(double(input_image, []));
%a=imread(input_image, []);

%imshow(input_image, [])
class=classify(convnet,a)
imshow(a)
title(string(class))

```