

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

Νικόλας Α. Τσαπατσούλης

Διπλ. Ηλεκτρολόγος Μηχανικός ΕΜΠ

**ΠΡΟΗΓΜΕΝΕΣ ΤΕΧΝΙΚΕΣ ΑΝΑΓΝΩΡΙΣΗΣ ΠΡΟΣΩΠΩΝ
ΚΑΙ ΑΝΑΛΥΣΗΣ ΕΚΦΡΑΣΕΩΝ**

Επιβλέπων Καθηγητής:

Στέφανος Κόλλιας

Τριμελής Συμβουλευτική Επιτροπή:

Στέφανος Κόλλιας

Ανδρέας Σταφυλοπάτης

Παναγιώτης Τσανάκας

Επταμελής Εξεταστική Επιτροπή



Ανδρέας Σταφυλοπάτης
Καθηγητής ΕΜΠ



Παναγιώτης Τσανάκας
Καθηγητής ΕΜΠ

Στέφανος Κόλλιας
Καθηγητής ΕΜΠ



Γεώργιος Παπακωνσταντίνου
Καθηγητής ΕΜΠ



Γεώργιος Καραγιάννης
Καθηγητής ΕΜΠ



Πέτρος Μαραγκός
Καθηγητής ΕΜΠ



Χρήστος Σχίζας
Καθηγητής Παν. Κύπρου

Περίληψη

ΠΡΟΗΓΜΕΝΕΣ ΤΕΧΝΙΚΕΣ ΑΝΑΓΝΩΡΙΣΗΣ ΠΡΟΣΩΠΩΝ ΚΑΙ ΑΝΑΛΥΣΗΣ ΕΚΦΡΑΣΕΩΝ

του Νικόλα Τσαπατσούλη

Η παρούσα διδακτορική διατριβή έχει κεντρικό θέμα την επικοινωνία ανθρώπου μηχανής όπως αυτή εκφράζεται μέσα από δύο ιδιαίτερα σημαντικές διεργασίες που εκτελούνται από τα ανθρώπινα όντα: την αναγνώριση προσώπων και την ανάλυση ανθρωπίνων εκφράσεων. Στόχος της είναι η συνεισφορά προς την κατεύθυνση της ανάπτυξης ενός πλήρως αυτοματοποιημένου υπολογιστικού συστήματος το οποίο να είναι ικανό να προσομοιώνει σε κάποιο αποδεκτό επίπεδο τις παραπάνω ικανότητες των ανθρώπων. Η διατριβή απαρτίζεται από τέσσερα μέρη. Στο πρώτο μέρος γίνεται μια εισαγωγή στην επικοινωνία ανθρώπου μηχανής και στη συνέχεια πραγματοποιείται μια ανασκόπηση των υφιστάμενων τεχνικών για την αναγνώριση προσώπων και την ανάλυση εκφράσεων όπως αυτές εκφράζονται μέσω τεσσάρων υποπροβλημάτων: της ανίχνευσης και εντοπισμού προσώπων, του εντοπισμού των σημαντικών χαρακτηριστικών του προσώπου, της ταυτοποίησης προσώπων και της αναγνώρισης εκφράσεων με χρήση οπτικών στοιχείων. Το δεύτερο μέρος αναφέρεται στον αυτόματο εντοπισμό προσώπων και χαρακτηριστικών τους σε φωτογραφίες ή ακολουθίες βίντεο και σε εφαρμογές πέρα από την αναγνώριση προσώπων στις οποίες η διαδικασία εντοπισμού είναι χρήσιμη. Γίνεται εκτεταμένη αναφορά στις τεχνικές που αναπτύχθηκαν προς την κατεύθυνση αυτή κατά τη διάρκεια της διδακτορικής διατριβής. Στο τρίτο μέρος της διατριβής παρουσιάζονται οι προτεινόμενες τεχνικές για ταυτοποίηση προσώπων και πώς αυτές συνδέονται με προηγούμενες εργασίες καθώς και με τη διαδικασία εντοπισμού προσώπων του δεύτερου μέρους. Η ανάλυση εκφράσεων μέσω της οπτικής πληροφορίας εξετάζεται στο τέταρτο μέρος της διατριβής. Η διατριβή τερματίζεται με έκθεση συμπερασμάτων και προτάσεις για κάποια ζητήματα τα οποία θα μπορούσαν να αποτελέσουν τη βάση για μελλοντική έρευνα.

ΕΥΧΑΡΙΣΤΙΕΣ-ΑΦΙΕΡΩΣΗ

Η σελίδα στην οποία κάποιος εκφράζει την εκτίμηση του, όσον αφορά τη βοήθεια και τις συμβουλές των οποίων έτυχε, σε τρίτα πρόσωπα είναι από τις δυσκολότερες στη συγγραφή. Εκτός του ότι κινδυνεύεις να ξεχάσεις κάποιους ανθρώπους, ξεκινάς πάντοτε με τη δυσκολία της απόδοσης στα Ελληνικά της λέξης Acknowledgments.

Σε αυτό το κομμάτι της διατριβής θα ήθελα να εκφράσω την εκτίμηση μου για πολλούς ανθρώπους. Θα αρχίσω όμως από κάποιο ίδρυμα: Το Ίδρυμα Κρατικών Υποτροφιών -ΙΚΥ, η υποτροφία του οποίου μου έδωσε την ευκαιρία να ξεκινήσω την τρέχουσα διατριβή.

Όσον αφορά τους ανθρώπους, θα πρέπει να εκφράσω τις ευχαριστίες μου στον επιβλέποντα της διατριβής, Καθηγητή κ. Κόλλια, η βοήθεια του οποίου ήταν πολύπλευρη. Ξεχωρίζω τις συμβουλές του όσον αφορά το επιστημονικό πεδίο αλλά και την επικοινωνιακή (constructive) στάση απέναντι στη ζωή. Θέλω επίσης να ευχαριστήσω την σύζυγο μου Γεωργία για την ανοχή και αντοχή που επέδειξε όλα αυτά τα χρόνια.

Τέλος πρέπει να ευχαριστήσω και να συγχαρώ όλους τους συναδέλφους, παλαιότερους και νεότερους, του Εργαστηρίου Ψηφιακής Επεξεργασίας Εικόνας Βίντεο και Πολυμέσων τόσο για τη βοήθεια τους όσο και για τη δημιουργία ενός ιδιαίτερα ευχάριστου εργασιακού και ερευνητικού χώρου. Με τους περισσότερους από αυτούς συνεργαστήκαμε και στον επιστημονικό τομέα και η σχέση μας είναι τουλάχιστον φιλική.

Η παρούσα διατριβή αφιερώνεται στη μνήμη του πατέρα μου Ανδρέα και του γαμπρού μου Δημήτρη, ο οποίος ήταν περισσότερο από αδελφός για μένα.

ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ ΚΑΙ ΣΧΗΜΑΤΩΝ

Σχήμα A1.1: Μπλοκ διάγραμμα της επικοινωνίας ανθρώπου μηχανής μέσω της οπτικής πληροφορίας	1
Εικόνα A2.1: Παραδείγματα mug shot φωτογραφιών και φωτογραφιών από δημόσια έγγραφα	9
Εικόνα A2.2: Φωτογραφίες προσώπων σε ανομοιόμορφο φόντο	9
Εικόνα A3.1: Ανθρώπινο πρόσωπο σε διάφορες χωρικές αναλύσεις	16
Σχήμα A4.1: Σύστημα συντεταγμένων με βάση τη φωτογραφία	19
Σχήμα A4.2: Ολοκληρωματικές προβολές	22
Σχήμα A4.3: Αναπαράσταση προσώπου ως άθροισμα ιδιοδιανυσμάτων	23
Εικόνα A4.1: Σημεία σε εικόνες προφίλ και εμπρόσθια όψη	26
Σχήμα A5.1: Ο μετασχηματισμός Radon	44
Εικόνα A5.1: Χρονική εξέλιξη της έκφρασης «οργή»	48
Εικόνα A5.2: Χρονική εξέλιξη της έκφρασης «χαρά»	48
Εικόνα A5.3: Χρονική εξέλιξη της έκφρασης «έκπληξη»	49
Εικόνα A5.4: Προβολή αποτελεσμάτων ταξινόμησης στο χώρο ενεργοποίησης - επαλήθευσης	49
Εικόνα B1.1: Υπολογισμός της εικόνας διαφορών	57
Εικόνα B1.2: Κατάτμηση προσώπου με ανίχνευση των pixels που ικανοποιούν την εξίσωση της έλλειψης	59
Εικόνα B1.3: Κατάτμηση προσώπου από ανομοιογενές φόντο και υπό περιστροφή	59
Εικόνα B1.4: Εφαρμογή της ανίχνευσης ματιών	60
Εικόνα B1.5: Εντοπισμός της οριζόντιας θέσης της μύτης και του στόματος	62
Εικόνα B1.6: Εντοπισμός χαρακτηριστικών σημείων στο προφίλ	64
Εικόνα B1.7: Χαρακτηριστικά σημεία στη εμπρόσθια όψη και στο προφίλ	65
Σχήμα B2.1: Αποτελέσματα εντοπισμού τμημάτων δέρματος για τέσσερις διαφορετικές κατηγορίες εικόνων	70
Εικόνα B2.1: Εντοπισμός και απομόνωση των πιθανών τμημάτων δέρματος	72
Εικόνα B2.2: Παραδείγματα ανίχνευσης προσώπου σε μονόχρωμες εικόνες	76
Εικόνα B2.3: Παραδείγματα εντοπισμού προσώπου σε μονόχρωμες ακολουθίες βίντεο	77
Εικόνα B2.4: Παραδείγματα εντοπισμού προσώπου σε έγχρωμες φωτογραφίες	78
Εικόνα B2.5: Εντοπισμός πολλαπλών προσώπων	78
Σχήμα B3.1: Το μπλοκ διάγραμμα της τεχνικής ανάλυσης τηλεοπτικών δελτίων ειδήσεων	82
Σχήμα B3.2: Συναρτήσεις συμμετοχής για την κυκλικότητα και την επιμήκυνση προσώπου	85
Εικόνα B3.1: Εντοπισμός και κατάτμηση προσώπου σε πλάνο του studio	86
Εικόνα B3.2: Ένα τυπικό απόσπασμα από ακολουθία δελτίου ειδήσεων	87

Εικόνα B3.3: Χρονική ανάλυση μιας ακολουθίας από δελτίο ειδήσεων	88
Σχήμα B4.1: Μπλοκ διάγραμμα του σχήματος ανάκλησης εικόνων προσώπου με βάση το περιεχόμενο	91
Εικόνα B4.1: Ανάκληση με βάση την ομοιότητα χρώματος	96
Εικόνα B4.2: Ανάκληση με βάση την ομοιότητα στην κλίμακα	96
Εικόνα B4.3: Ανάκληση με βάση τον αριθμό των τμημάτων δέρματος	97
Εικόνα B4.4: Ανάκληση με βάση τις κεντρικές ροπές φωτεινότητας	98
Σχήμα Γ1.1: Η αρχιτεκτονική ενός γενικού συστήματος αναγνώρισης προσώπων	104
Εικόνα Γ1.1: Η συνολική διασπορά του πληθυσμού σε σχέση με τη διασπορά των επιμέρους κλάσεων	107
Εικόνα Γ2.1: Εντός κλάσης διακυμάνσεις για τις εικόνες προσώπου	115
Σχήμα Γ2.1: Συγκρίσεις όσον αφορά την μετατόπιση ανάμεσα στους KLT και VKLT	115
Σχήμα Γ2.2: Μέσο σφάλμα ανακατασκευής των μετασχηματισμών KL και VKL	116
Σχήμα Γ2.3: Συγκρίσεις όσον αφορά τις μεταβολές φωτισμού ανάμεσα στους KLT και VKLT	116
Σχήμα Γ3.1: Σχήμα για ταυτοποίηση με τη χρήση του μετασχηματισμού DCT	118
Σχήμα Γ3.2: Δημιουργία του διανύσματος αναπαράστασης από τους συντελεστές DCT των μπλοκ της εικόνας	118
Εικόνα Γ4.1: Πρόσωπα υπό διαφορετικές συνθήκες φωτισμού και τα αντίστοιχα διανύσματα ιδιαιδιών τιμών	123
Εικόνα Γ4.2: Η επίδραση του διανύσματος ιδιαιδιών τιμών στην υφή της εικόνας	124
Εικόνα Γ4.3: Προσαρμογή της φωτεινότητας προσώπων της βάσης σύμφωνα με αυτήν της εικόνας εισόδου	125
Σχήμα Γ4.1: Μέση απόσταση ανάμεσα σε πρόσωπα σε σχέση με τα κριτήρια s και s^*	130
Σχήμα Γ5.1: Η αρχιτεκτονική του SOM	133
Σχήμα Γ5.2: Η αρχιτεκτονική του αλγόριθμου ταυτοποίησης	134
Σχήμα Γ5.2: Ταυτοποίηση μέσω του δικτύου SOM	135
Εικόνα Δ2.1: Πρότυπα των εκφράσεων “ουδέτερη”, “οργή”, “χαρά” και “λύπη”	146
Εικόνα Δ2.2: Τα ιδιοδιανύσματα εικόνων που αναπαριστούν εκφράσεις	148
Σχήμα Δ2.1: Εξαγωγή χαρακτηριστικών για περιγραφή εκφράσεων με συνδυασμό δύο νευρωνικών δικτύων	150
Εικόνα Δ2.3: Απεικόνιση των βαρών που αντιστοιχούν στους 20 νευρώνες του εσωτερικού επιπέδου	151
Εικόνα Δ2.4: Απεικόνιση των βαρών που αντιστοιχούν στους 5 νευρώνες του εσωτερικού επιπέδου	152
Εικόνα Δ3.1: Οπτική αναπαράσταση των συναισθημάτων “delighted” και “eager”	155
Σχήμα Δ3.1: Συνεισφορά των επιμέρους χαρακτηριστικών για την ταξινόμηση των πρωτεύουσών εκφράσεων	161
Σχήμα Δ3.2: Συνεισφορά των χαρακτηριστικών για την ταξινόμηση παραλλαγών της έκφρασης “χαρά”	161
Σχήμα Δ3.3: Τα 19 FDP σημεία που χρησιμοποιούνται για την περιγραφή των FAPs του Πίνακα Δ3.3	162
Σχήμα Δ3.4: Η δομή του ασαφούς συστήματος	162
Σχήμα Δ3.5: Συναρτήσεις συμμετοχής τραπέζοειδούς μορφής	163
Σχήμα Δ4.1: Η αρχιτεκτονική του σχήματος ταξινόμησης με τη χρήση HMM	170

Εικόνα Δ4.1: Κατάτμηση προσώπου σε υποπεριοχές	172
Εικόνα Δ4.2: Ενέργεια στις οκτώ κύριες κατευθύνσεις για κάποιες από τις περιοχές του προσώπου	173
Εικόνα Δ4.3: Διαγράμματα κίνησης σχετιζόμενα με την έκφραση “χαρά”	174
Σχήμα Δ4.2: Διάγραμμα του ταξινομητή HMM	175

ΟΡΟΛΟΓΙΑ

Mug shot φωτογραφίες: Φωτογραφίες που λαμβάνονται από αστυνομικές αρχές και απεικονίζουν το πρόσωπο σε κοντινό πλάνο. Αντίστοιχες με τις φωτογραφίες δημόσιων εγγράφων όπως διαβατήρια, ταυτότητες κλπ.

False alarm: Ταξινόμηση κάποιου προτύπου σε διαφορετική κατηγορία από αυτήν που ανήκει

Dismissal: Αποτυχία ταξινόμησης κάποιου προτύπου στη κατηγορία που πραγματικά ανήκει

FACS - Facial Action Coding System: Σύστημα κωδικοποίησης το οποίο περιγράφει όλες τις οπτικά διαχωρίσιμες κινήσεις του προσώπου

Facial Definition Parameter set – FDP: Σύνολο παραμέτρων για τον ορισμό 3D συνθετικών μοντέλων προσώπου στο πλαίσιο του MPEG-4

Facial Animation Parameter set – FAP: Σύνολο παραμέτρων για την απόδοση κίνησης προσώπου στο πλαίσιο του MPEG-4

Άψευ: Η κορύφωση της έκφρασης

3-Δ βίντεο: Σύλληψη με πολλαπλές κάμερες και δυνατότητα παρατήρησης μιας σκηνής από τυχαία θέση

Πλάνο –shot: Το σύνολο των καρέ τα οποία έχουν ληφθεί από την ίδια κάμερα

Φορμάτ προσώπου- head format: Οριοθέτηση των προσώπων σε μηδενική στροφή, ως προς και στους τρεις άξονες, ομοιόμορφη κλίμακα και με κάποια βασικά χαρακτηριστικά προσώπου –συνήθως τα μάτια- είναι ευθυγραμμισμένα.

ΠΕΡΙΕΧΟΜΕΝΑ

ΜΕΡΟΣ Α	<i>Error! Bookmark not defined.</i>
ΜΕΡΟΣ Β	<i>Error! Bookmark not defined.</i>
ΜΕΡΟΣ Γ	<i>Error! Bookmark not defined.</i>
ΜΕΡΟΣ Δ	<i>Error! Bookmark not defined.</i>
1. Εισαγωγή	6
2. Εφαρμογές της αναγνώρισης προσώπων	8
3. Προβλήματα στατικής ταύτισης	9
4. Προβλήματα δυναμικής ταύτισης	11
5. Τα βασικά υποπροβλήματα στην αναγνώριση προσώπων	12
Αναπαράσταση Προσώπων.....	12
Ανίχνευση και Εντοπισμός Προσώπων.....	12
Ταυτοποίηση Προσώπων.....	13
Ταξινόμηση με βάση τα φυσικά χαρακτηριστικά.....	13
1. Αποτελέσματα απο τον τομεα της ψυχοφυσικής και νευροφυσιολογιας	14
Η αναγνώριση προσώπου είναι μια τυποποιημένη διαδικασία.....	14
Η αντίληψη του προσώπου γίνεται ολιστικά ή με βάση μεμονωμένα χαρακτηριστικά.....	14
Η σημαντικότητα των επιμέρους χαρακτηριστικών.....	14
Πρωτοτυπία.....	14
Ο ρόλος των χωρικών συχνοτήτων.....	15
Ανάπτυξη του ανθρώπινου συστήματος αναγνώρισης κατά τη διάρκεια της παιδικής ηλικίας.....	15
Ο ρόλος του φύλου-φυλής.....	15
Συμπεράσματα.....	15
2. Οι δυνατότητες του οπτικού συστήματος των ανθρώπων	16
Αναπαράσταση.....	16
Εντοπισμός και κατάτμηση προσώπου.....	16
Ταυτοποίηση.....	17
Ταξινόμηση με βάση τα φυσικά χαρακτηριστικά.....	18
1. Εισαγωγή	19
2. Αναπαράσταση	21
2.1 Αναπαράσταση προσώπων - Εξαγωγή χαρακτηριστικών	21
(β) Στατιστικά Χαρακτηριστικά.....	23
(γ) Χαρακτηριστικά συντελεστών μετασχηματισμού.....	23
(δ) Αλγεβρικά Χαρακτηριστικά.....	23
2.2 Αναπαράσταση προσώπων με χρήση των τιμών της κλίμακας του γκρι	24
Ο ρόλος της χωρικής ανάλυσης σε διαδικασίες ταύτισης με βάση τις gray scale τιμές.....	24
2.3 Γεωμετρικά χαρακτηριστικά από εικόνες εμπρόσθιας όψης και προφύ	27
2.4 Υβριδική αναπαράσταση	27
3. Εντοπισμός - Κατάτμηση προσώπου	28
4. Ταυτοποίηση προσώπων	32
4.1 Επίλογή των χαρακτηριστικών	35
4.2 Σχήματα ταύτισης	35
1. Εισαγωγή	37
2. Προσωπα και συναισθηματικες καταστασεις	38
2.1 Ένα πλαίσιο περιγραφής της απεικόνισης των συναισθημάτων στα πρόσωπα.....	38
3. Απεικονίσεις που σχετίζονται με την συναισθηματική έκφραση	39
4. Νευροφυσιολογία και συναισθήματα	40

5. Τεχνικές αναγνώρισης εκφράσεων.....	41
Παρακολούθηση της κίνησης του προσώπου.....	41
5.1 Στατικές προσεγγίσεις.....	42
5.2 Δυναμικές προσεγγίσεις.....	42
(α) Εκτίμηση της κίνησης των σημείων στην επιφάνεια του προσώπου.....	42
(β) Η ολική κατανομή των διανυσμάτων κίνησης στην επιφάνεια του προσώπου και οι εκφράσεις.....	44
(γ) Πλήρως δυναμικές προσεγγίσεις.....	46
(δ) Η χρονική εξέλιξη του ενεργειακού πεδίου των διανυσμάτων κίνησης και οι εκφράσεις.....	48
1. Εντοπισμός περιγράμματος προσώπου.....	56
2. Εντοπισμός σημείων σε εικόνες εμπρόσθιας οψής.....	59
2.1 Εντοπισμός επιπλέον σημείων με χρήση των φίλτρων Gabor.....	62
3. Εντοπισμός σημείων σε εικόνες προφίλ.....	63
4. 3-δ αναπαράσταση.....	65
1. Εισαγωγή.....	67
2. Εντοπισμός των πιθανότερων τμημάτων δερματος.....	68
3. Απομονώση των τμημάτων δερματος.....	71
4. Φίλτρασμα με βάση το σχήμα.....	73
5. Εντοπισμός προσώπων εντός των τμημάτων.....	74
6. Πειραματικά αποτελέσματα.....	75
6.1 Εντοπισμός προσώπου σε μονόχρωμες εικόνες.....	75
6.2 Εντοπισμός προσώπου σε μονόχρωμες ακολουθίες βίντεο.....	76
6.3 Εντοπισμός προσώπου σε έγχρωμες φωτογραφίες.....	77
6.4 Εντοπισμός προσώπου σε έγχρωμες ακολουθίες βίντεο.....	78
7. Συμπεράσματα.....	79
1. Εισαγωγή.....	80
2. Ανάλυση του προτεινομένου συστήματος.....	82
2.1 Βαθμός μεταβολής ιστογράμματος.....	84
2.2 Μεταβολή μεγέθους και μετατόπιση τμημάτων.....	84
3. Εντοπισμός και περιγραφή προσώπων στα καρέ.....	84
4. Πειραματικά αποτελέσματα.....	86
1. Εισαγωγή.....	90
2. Ένα σενάριο ανακλήσης.....	91
3. Χαρακτηριστικά περιγραφής τμημάτων προσώπου.....	92
4. Ανακλήση με βάση το περιεχόμενο.....	93
4.1 Ανάκληση με βάση τις μέσες χρωματικές συνιστώσες.....	93
4.2 Ανάκληση με βάση τις κανονικοποιημένες κεντρικές ροπές.....	94
4.3 Ανάκληση με βάση την κλίμακα.....	95
4.4 Ανάκληση με βάση τον αριθμό των προσώπων.....	95
5. Πειραματικά αποτελέσματα.....	95
6. Συμπεράσματα.....	97
1. Εισαγωγή.....	101

2. Εντοπισμός των πυρήνων.....	102
3. Ταξινόμηση κυτταρών – σύστημα προγνώσης.....	104
Χαρακτηριστικά για την ταξινόμηση των πυρήνων.....	104
Ταξινόμηση των πυρήνων.....	105
Υπολογισμός του δείκτη διάγνωσης.....	105
4. Διαδικασία επαλήθευσης – συμπεράσματα.....	105
1. Εισαγωγή.....	108
2. Συστήματα Αναγνώρισης Προσώπων.....	109
3. Μαθηματική περιγραφή της αναγνώρισης προσώπων.....	109
3.1 Ο μετασχηματισμός Karhunen-Loeve.....	111
3.2 Fisher Linear Discriminant –FLD.....	114
1. Εισαγωγή.....	116
2. Η προσέγγιση του KLT με διανύσματα.....	117
3. Διαδικασία επαλήθευσης.....	118
4. Πειραματικά Αποτελέσματα.....	120
1. Εισαγωγή.....	122
2. Χονδρική περιγραφή της προτεινόμενης τεχνικής.....	122
3. Δημιουργία του διανύσματος αναπαράστασης.....	123
4. Το στάδιο ταυτοποίησης.....	124
5. Πειραματικά αποτελέσματα.....	124
6. Συμπεράσματα.....	125
1. Εισαγωγή.....	126
2. Αναλυση Ιδιαζουσών Τιμών.....	126
2.1 Ανασθησία του διανύσματος ιδιαζουσών τιμών σε αναλογικές μεταβολές στη φωτεινότητα.....	127
3. Προσαρμογή των εικόνων προσώπου σε άλλες συνθήκες φωτισμού.....	130
4. Διαδικασία ταυτοποίησης.....	130
5. Πειραματικά αποτελέσματα.....	134
6. Συμπεράσματα.....	135
1. Εντοπισμός και κατάτμηση του προσώπου.....	136
2. Σύνολα περιγραφής.....	136
2.1 Αναπαράσταση ανεξάρτητη κλίμακας.....	136
2.2 Αναπαράσταση με βάση την υφή και το φωτισμό.....	138
2.3 Αναπαράσταση ανεξάρτητη από το γενικό φωτισμό - Luminance Invariant Representation.....	139
3. Αλγόριθμος Ταυτοποίησης.....	139
3.1 Ταυτοποίηση ανεξάρτητη από κλίμακα.....	139
3.2 Ταυτοποίηση ανεξάρτητη από φωτισμό.....	140
3.3 Ταυτοποίηση βασισμένη στην πληροφορία υφής.....	141
4. Μηχανή λήψης απόφασης.....	141
5. Πειραματικά Αποτελέσματα.....	143
6. Συμπεράσματα.....	144

1. Εισαγωγή.....	148
2. Τα προτυπα μεσησ τιμησ.....	148
3. Η μεθοδος FLD.....	149
3.1 Ανάλυση PCA για τη δημιουργία συμπαγούς αναπαράστασης.....	150
3.2 Επιλογή των χαρακτηριστικών με τη μεγαλύτερη ικανότητα διαχωρισμού.....	151
4. Τεχνικές επιβλεπομενης μαθησης.....	152
4.1 Ταξινόμηση μέσω του MLP δικτύου εφαρμοσμένη άμεσα στα δεδομένα εισόδου.....	153
4.2 Συμπύεση και αναπαράσταση.....	154
4.3 Ταξινόμηση των αναπαραστάσεων της PCA ανάλυσης.....	155
5. Συμπερασματα.....	155
1. Εισαγωγή.....	157
2. Συνολα παραμετρων για τον ορισμο προσωπων και της κινησησ σε αυτα.....	159
3. Συσχετιση των FDPs και των προτυπων εκφρασεων.....	160
3.1 Ανίχνευση των χαρακτηριστικών σημείων προσώπου.....	163
3.2 Αξιολόγηση του διανύσματος χαρακτηριστικών.....	163
4. Το συστημα ασαφους λογικης.....	164
4.1 Ασαφής διαμέριση των διαστημάτων διακύμανσης των στοιχείων της εισόδου.....	165
4.2 Οι κανόνες απόφασης.....	168
4.3 Υπολογισμός της εξόδου του συστήματος ασαφούς λογικής.....	169
5. Αναγνωριση παραλλαγων των προτυπων συναισθηματων.....	171
6. Πειραματικα αποτελεσματα.....	171
7. Συμπερασματα.....	172
1. Ο προτεινομενος αλγοριθμος.....	175


```

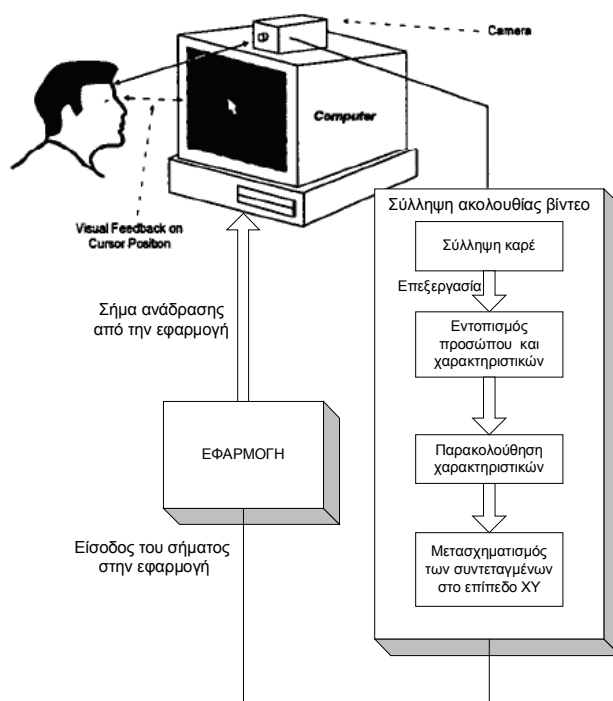
graph LR
    A[Ακολουθία βίντεο] --> B[Κατάμηση προσώπου και ευθυγράμμιση χαρακτηριστικών]
    B --> C[Εκτίμηση κίνησης και φιλτράρισμα]
    C --> D[Δημιουργία του διανύσματος αναπράστασης]
    D --> E[Ταξινομητής HMM]
    E --> F[Έκφραση]
  
```

.....	175
1.1 Εκτίμηση κίνησης.....	175
1.2 Φιλτράρισμα των διανυσμάτων κίνησης.....	176
2 Δημιουργία του διανύσματος αναπαράστασης.....	176
3.Ο ταξινομητής HMM.....	179
4. Πειραματικα αποτελεσματα.....	181
5. Συμπερασματα.....	181
Εισαγωγη.....	190

<i>Το preg-4 και η αναγνώριση χαρακτηριστικών</i>	191
Αναλυτικά για το FAPs.....	191
Αναλυτικά για το FDPs.....	192

ΕΙΣΑΓΩΓΗ ΣΤΗΝ ΕΠΙΚΟΙΝΩΝΙΑ ΑΝΘΡΩΠΟΥ ΜΗΧΑΝΗΣ

Η ανάλυση και αναγνώριση προσώπου είχε αρχικά προταθεί ως μια εναλλακτική πρόταση ταυτοποίησης ανθρώπων χωρίς τη χρήση εγγράφων. Η πράξη δείχνει ότι άλλα ανθρωπομετρικά χαρακτηριστικά όπως τα δακτυλικά αποτυπώματα είναι περισσότερο αποδοτικά για το σκοπό αυτό. Από την άλλη πλευρά η συνεχής ανάπτυξη των εφαρμογών πολυμέσων κατέστησε την αναγνώριση προσώπων αλλά και την ανάλυση εκφράσεων εξαιρετικά σημαντικές εφαρμογές στον τομέα της επικοινωνίας ανθρώπου μηχανής αλλά και της πλοήγησης σε πολυμεσιές βάσεις δεδομένων.



Σχήμα Α1.1: Μπλοκ διάγραμμα της επικοινωνίας ανθρώπου μηχανής μέσω της οπτικής πληροφορίας

Η επικοινωνία των ανθρώπινων όντων πραγματοποιείται μέσω δύο διαφορετικών καναλιών: Είτε μέσω σαφών λεκτικών μηνυμάτων, τα οποία μπορούν να αφορούν σε οτιδήποτε, είτε μέσω εσωτερικών μηνυμάτων τα οποία εκφράζουν πληροφορία σχετικά με την συναισθηματική κατάσταση των ομιλητών. Παρόλο που το πρώτο κανάλι επικοινωνίας έχει προσελκύσει τεράστιο ενδιαφέρον από ένα ευρύ φάσμα επιστημών, όπως η γλωσσολογία, η επεξεργασία φωνής κοκ, το ίδιο δεν έχει συμβεί με το δεύτερο κανάλι. Η αναγνώριση προσώπων και των μορφοποιήσεων που λαμβάνουν χώρα σε αυτό, η αναγνώριση

νοηματικών κινήσεων καθώς και η κατανόηση συναισθημάτων είναι κάποια από τα ουσιαστικότερα θέματα που σχετίζονται με την επικοινωνία μέσω του δεύτερου καναλιού.

Η αναγνώριση προσώπου και των νοηματικών κινήσεων σε αυτό είναι ιδιαίτερα σημαντική διαδικασία γιατί παρέχει μεταξύ άλλων και εναλλακτικούς τρόπους επικοινωνίας με τον υπολογιστή σε ανθρώπους με κάποια μορφή αναπηρίας. Ο παραδοσιακός τρόπος επικοινωνίας ανθρώπου και υπολογιστή πραγματοποιείται με τη χρήση πληκτρολογίου και ποντικιού. Εναλλακτικά μπορεί να πραγματοποιηθεί διαβίβαση εντολών με τη χρήση μικροφώνου και με διαδικασίες λεκτικής ανάλυσης. Για ανθρώπους με σοβαρά κινητικά προβλήματα αλλά και με προβλήματα ομιλίας και ακοής η επικοινωνία μέσω κινήσεων του κεφαλιού ή περιοχών του προσώπου όπως τα χείλη, τα μάτια και τα βλέφαρα είναι ιδιαίτερα επιθυμητή. Στο Σχήμα A1.1 παρουσιάζεται με ποιο τρόπο ένα σύστημα επικοινωνίας ανθρώπου υπολογιστή πραγματοποιείται με τη χρήση μιας απλής ατομικής βιντεοκάμερας. Στο σύστημα αυτό εμφανίζονται κάποια επιμέρους προβλήματα τα οποία απαιτούν αξιόπιστη λύση και αφορούν στο τμήμα σύλληψης και επεξεργασίας της ακολουθίας βίντεο. Η επίλυση συγκεκριμένων θεμάτων όπως ο εντοπισμός της θέσης και η παρακολούθηση της κίνησης του προσώπου, ο εντοπισμός των βασικών χαρακτηριστικών του προσώπου και η παρακολούθηση της κίνησης τους καθώς και η εύρεση της θέσης συγκεκριμένων σημείων στην επιφάνεια του αποτελούν τον κορμό του ανωτέρω σχήματος επικοινωνίας. Με βάση την πληροφορία που εξάγεται στα στάδια αυτά, δίνονται οι εντολές για την κατάλληλη αντίδραση του υπολογιστή στα δεδομένα της εισόδου του.

Η κατανόηση των συναισθημάτων είναι επίσης ένα εξαιρετικά σημαντικό θέμα στην επικοινωνία ανθρώπου μηχανής. Δεδομένης της δυσκολίας με την οποία οι άνθρωποι ερμηνεύουν την εσωτερική κατάσταση των συνανθρώπων τους είναι αντιληπτό από όλους ότι η κατανόηση των ανθρωπίνων συναισθημάτων από τον υπολογιστή αποτελεί μια πρόκληση. Η ανάλυση και αναδημιουργία του μη λεκτικού καναλιού επικοινωνίας μπορεί να οδηγήσει σε μια πληθώρα εφαρμογών μερικές από τις οποίες συνοψίζονται στη συνέχεια:

Δημιουργία ομιλίας με συναισθηματικό χρωματισμό

Η οικειότητα ή συμπάθεια ανάμεσα σε δύο επικοινωνούντα άτομα εκφράζεται κατά το μεγαλύτερο ποσοστό της από μη λεκτικού περιεχομένου πληροφορία. Το ίδιο ισχύει και για την επιφυλακτικότητα και την τήρηση αποστάσεων. Το συναισθηματικό χρώμα στην ομιλία αλλά και οι εκφράσεις του προσώπου είναι οι κλασικοί τρόποι με τους οποίους εκφράζεται η ευαρέσκεια ή δυσαρέσκεια κατά τη διάρκεια της επικοινωνίας. Οι ηλεκτρονικοί ομιλητές χαρακτηρίζονται από την απουσία του συναισθηματικού χρώματος στη φωνή που πολλές φορές λειτουργεί ανασταλτικά στους ανθρώπους που επικοινωνούν μαζί τους. Η κατανόηση της συναισθηματικής κατάστασης του ανθρώπου από τον υπολογιστή μπορεί να οδηγήσει στον κατάλληλο χρωματισμό της ηλεκτρονικής ομιλίας οδηγώντας είτε στην τήρηση αποστάσεων είτε στην έκφραση οικειότητας και συμπάθειας.

Αλληλεπίδραση του λεκτικού και μη λεκτικού καναλιού επικοινωνίας

Το μη λεκτικό κανάλι επικοινωνίας καθορίζει σε πολλές περιπτώσεις το πως πρέπει να ερμηνευθούν τα λεκτικά δεδομένα. Με απλά λόγια η κατανόηση του αν τα λεγόμενα πρέπει να εληφθούν κυριολεκτικά ή μεταφορικά είναι κρίσιμος παράγοντας για την αποδοτική συνέχιση της επικοινωνίας. Προφανώς και η

αντίδραση του υπολογιστή πρέπει να είναι προσαρμοσμένη στην αλληλεπίδραση του λεκτικού και μη λεκτικού καναλιού επικοινωνίας.

Επέκταση της ανθρώπινης κρίσης

Η ανίχνευση των συναισθημάτων του συνομιλητή μας είναι πολλές φορές συνυφασμένη με την εξαγωγή συμπερασμάτων σχετικά με τον χαρακτήρα και την προσωπικότητα. Μια κλασική εφαρμογή που υπάγεται στο πλαίσιο αυτό είναι η ανίχνευση ψεύδους. Η ικανότητα των ανθρώπων όσον αφορά τη διαδικασία αυτή μοιάζει ασύλληπτη για τους υπολογιστές. Υπάρχουν ωστόσο εφαρμογές στις οποίες η επέκταση της ανθρώπινης κρίσης είναι εφικτή για τον υπολογιστή. Χαρακτηριστικό παράδειγμα αποτελεί η διάγνωση στοιχείων που δηλώνουν προχωρημένη σχιζοφρένεια μέσω του συναισθηματικού χρώματος της ομιλίας. Με δεδομένο ότι οι άνθρωποι δεν αντιλαμβάνονται τα στοιχεία αυτά με την ευκολία που θα περιμέναμε κανείς, η χρησιμότητα του υπολογιστή στο συγκεκριμένο πεδίο είναι τουλάχιστον σημαντική.

Δημιουργία συνθετικών μοντέλων

Η κατανόηση των συναισθημάτων είναι βασική απαίτηση των σύγχρονων προτύπων κωδικοποίησης στα οποία η δημιουργία συνθετικών μοντέλων αποτελεί δομικό στοιχείο. Ορθή κατανόηση οδηγεί σε ορθή μοντελοποίηση και φυσικά σε ορθή αναπαραγωγή στο δέκτη. Αποτυχία μετάδοσης των συναισθηματικών καταστάσεων μπορεί να έχει καταστροφικές συνέπειες.

Εκπαίδευση

Μια προφανής εφαρμογή για τους υπολογιστές που μπορούν να κατανοούν τα συναισθήματα είναι η αυτόματη εκπαίδευση. Ένας αποδοτικός εκπαιδευτής πρέπει να μπορεί να αντιλαμβάνεται τις αντιδράσεις του ακροατηρίου και να προσαρμόζει αντίστοιχα τη διδασκαλία του. Η ανάλυση των εκφράσεων και συναισθημάτων δίνει την απαιτούμενη ανατροφοδότηση ώστε να γίνεται έγκαιρα αντιληπτή η κούραση, η αδυναμία παρακολούθησης, η ανησυχία κ.ο.κ.

Μεταφορά χειρισμού από τον υπολογιστή στο άνθρωπο

Η φοβία και η αμηχανία πολλών ανθρώπων όταν αναγκάζονται να επικοινωνήσουν με κάποιο μηχάνημα – τόσο απλό όσο ένα τηλεφωνικός πληροφοριοδότης- είναι κάτι το οποίο απαιτείται να ανιχνεύεται ώστε να διαβιβάζεται ο χειρισμός του προβλήματος σε ανθρώπους. Με απλά λόγια ο υπολογιστής πρέπει να μπορεί μέσω της συναισθηματικής κατάστασης του ατόμου που επικοινωνεί με αυτόν να καταλαβαίνει πότε είναι ικανός να χειριστεί την κατάσταση και πότε να διαβιβάσει τον έλεγχο σε άλλους.

Προσαρμογή ρυθμίσεων λογισμικού

Σε πάρα πολλές περιπτώσεις ο τρόπος με τον οποίο αντιδρά ο υπολογιστής σε κάποιες από τις ενέργειες στις οποίες προβαίνουμε μας απογοητεύει. Ένας υπολογιστής ή λογισμικό το οποίο μπορεί να κατανοήσει τα συναισθήματα του χρήστη μπορεί να προσαρμόσει αντίστοιχα τις αντιδράσεις του. Κλασικό παράδειγμα των περιπτώσεων αυτών είναι η ταχύτητα με την οποία δίνονται οι οδηγίες κατά την εγκατάσταση κάποιου λογισμικού. Αν ο χρήστης αδυνατεί να τις παρακολουθήσει ο υπολογιστής πρέπει να προχωρήσει σε απλούστερες διαδικασίες ώστε να επιτευχθεί το ζητούμενο αποτέλεσμα. Αντίστοιχο παράδειγμα αποτελεί η αυτόματη ενεργοποίηση κάποιων χαρακτηριστικών σε ορισμένα προγράμματα

λογισμικού. Ακόμα και κατά τη συγγραφή της συγκεκριμένης διατριβής ορισμένα χαρακτηριστικά του επεξεργαστή κειμένου προκαλούσαν τη δυσaréσκεια του συγγραφέα –για παράδειγμα η μετατροπή πεζών σε κεφαλαία, η αυτόματη αρίθμηση κ.ο.κ. Ένα έξυπνο λογισμικό θα πρέπει να μπορεί να αντιλαμβάνεται τη δυσaréσκεια του χρήστη και να προσαρμόζει αντίστοιχα τις ρυθμίσεις του.

Διασκέδαση

Ο τομέας της διασκέδασης μέσω υπολογιστών δεν είναι καθόλου αμελητέος. Κανένας δεν μπορεί να αμφισβητήσει την πρόοδο που έχει επιτευχθεί, όσον αφορά στις ικανότητες των υπολογιστών, εξαιτίας της να προσφέρουν ελκυστικά παιχνίδια. Οι συνθετικοί «ηθοποιοί» για να είναι καλοί στο ρόλο τους πρέπει να μπορούν να εκφράζουν σωστά τις κατάλληλες για την περίπτωση συναισθηματικές καταστάσεις. Ένας πραγματικά κακός συνθετικός αντιήρωας πρέπει να θυμώνει με πειστικό τρόπο!

Η παρούσα διατριβή εξετάζει τα θέματα της αυτόματης ανάλυσης και αναγνώρισης προσώπων και εκφράσεων στο πλαίσιο της επικοινωνίας ανθρώπου μηχανής όπως αυτό καθορίστηκε στις παραγράφους που προηγήθηκαν. Επειδή τόσο η ανάλυση εκφράσεων όσο και -κυρίως- η αναγνώριση προσώπων δεν είναι συσχετισμένες μόνο με την επικοινωνία ανθρώπου και υπολογιστή αλλά και πληθώρα άλλων εφαρμογών γίνεται προσπάθεια μιας όσο το δυνατόν σφαιρικότερης κάλυψης των δύο αυτών θεμάτων.

ΤΟ ΠΡΟΒΛΗΜΑ ΤΗΣ ΑΝΑΓΝΩΡΙΣΗΣ ΠΡΟΣΩΠΩΝ

Οι άνθρωποι εντοπίζουν και αναγνωρίζουν πρόσωπα σε μια σκηνή με χαρακτηριστική ευκολία. Η ικανότητα αυτή αναπτύσσεται σταδιακά κατά τη διάρκεια της παιδικής ηλικίας και μαζί με τη δυνατότητα εκτίμησης της έκφρασης ενός προσώπου παίζουν ένα πολύ σημαντικό ρόλο στη κοινωνική ζωή των ανθρώπων. Η ευκολία εντούτοις με την οποία εκτελείται η παραπάνω διαδικασία στον άνθρωπο δεν πρέπει να μας παραπλανεί. Τα πρόσωπα αντιπροσωπεύουν σύνθετες, πολυδιάστατες, γεμάτες πληροφορία διεγέρσεις, πράγμα που καθιστά την ανάπτυξη ενός υπολογιστικού συστήματος αναγνώρισης προσώπων εξαιρετικά δύσκολη.

1. ΕΙΣΑΓΩΓΗ

Το πρόσωπο είναι ένα μοναδικό χαρακτηριστικό για τα ανθρώπινα όντα. Ακόμα και τα πρόσωπα δύο ομοζυγωτικών διδύμων διαφέρουν. Εντοπίζουμε, αναγνωρίζουμε και διακρίνουμε τα πρόσωπα με χαρακτηριστική ευκολία και απλή παρατήρηση. Η ευκολία αναγνώρισης των προσώπων από τον άνθρωπό ήταν ένας από τους λόγους για την επιλογή γραμματοσήμων το οποία περιλάμβαναν πρόσωπα: ήταν πολύ δυσκολότερο να γίνουν πλαστογραφίες γραμματοσήμων γιατί αυτές γίνονταν εύκολα αντιληπτές. Η μοναδικότητα των προσώπων είναι επίσης ο βασικός λόγος για την ευρεία χρήση του σε εφαρμογές όπου η ταυτοποίηση ανθρώπων είναι ιδιαίτερα σημαντική.

Στο σημείο αυτό πρέπει να γίνει διάκριση μεταξύ ταυτοποίησης προσώπων και ταυτοποίησης ανθρώπων με βάση τη συνολική τους συμπεριφορά. Στην ταυτοποίηση προσώπων χρησιμοποιείται αποκλειστικά το οπτικό σύστημα του ανθρώπου. Στην γενικότερη ταυτοποίηση ανθρώπων χρησιμοποιούνται και άλλα στοιχεία, όπως το βάδισμα, η στάση του σώματος, η φωνή ή το ντύσιμο. Παρόλο που το πρόσωπο και τα χαρακτηριστικά του χρησιμοποιούνται ευρύτατα για την αναγνώριση ανθρώπων με κανένα τρόπο δεν αποτελούν τους μοναδικούς τρόπους αναγνώρισης. Στο τέλος του προηγούμενου αιώνα, ένα σύστημα βασισμένο σε μετρήσεις άλλων χαρακτηριστικών του σώματος, όπως το μήκος και εύρος του κεφαλιού, το μήκος του αριστερού πέλματος, το μήκος του μεσαίου δακτύλου του αριστερού χεριού και ποδιού, είχε χρησιμοποιηθεί στη Γαλλία για την ταυτοποίηση φυλακισμένων. Πρόσφατα αναπτύχθηκαν τεχνικές για ταυτοποίηση ανθρώπων βασισμένες στη δομή των γονιδίων.

Ένα σύστημα εντοπισμού και ταυτοποίησης προσώπων θα ήταν χρήσιμο σε μια πλειάδα εφαρμογών όπως αναγνώριση εγκληματιών, πρόσβαση σε συστήματα ασφαλείας όπως ταμεία τραπεζών (ATM) και γενικά σε κάθε εφαρμογή που απαιτεί πρόσβαση βάση κάποιου αποδεικτικού στοιχείου ταυτότητας. Έτοιμα συστήματα στη παρούσα φάση χρησιμοποιούν στοιχεία όπως κάρτες, PINs και κωδικούς πρόσβασης. Πρόσβαση βασισμένη σε αναγνώριση προσώπων θα μπορούσε να είναι περισσότερο αξιόπιστη και πιο απλή. Η αλήθεια πάντως είναι ότι σε εφαρμογές που απαιτούν υψηλή αξιοπιστία όσον αφορά την ταυτοποίηση, η αναγνώριση προσώπου δεν επαρκεί –τουλάχιστον με τα υπάρχοντα δεδομένα. Αντίθετα σε

εφαρμογές όπως τα συστήματα επικοινωνίας ανθρώπου μηχανής, ανάκλησης στοιχείων από βάσεις δεδομένων με βάση το περιεχόμενο, τηλεδιάσκεψη κλπ, η αναγνώριση προσώπου είναι πιο εύκολα υλοποιήσιμη και σε πολλές περιπτώσεις απαραίτητη.

Η ανάπτυξη ενός υπολογιστικού μοντέλου αναγνώρισης προσώπων είναι μια πρόκληση για την ανθρώπινη εφευρετικότητα και φιλοδοξία. Οι πιθανές εφαρμογές ενός τέτοιου μοντέλου, ενδεχομένως σε μελλοντικό χρόνο, χρησιμοποιούμενου για ασφάλεια κτιρίων και σταθμών εργασίας, αναγνώριση κακοποιών, επαλήθευση πιστωτικών καρτών, βελτίωσης της επικοινωνίας ανθρώπου - μηχανής κλπ, το καθιστούν ιδιαίτερα σημαντικό.

Το πρόβλημα χονδρικά τίθεται ως εξής: Δοθέντων φωτογραφιών ή ακολουθιών βίντεο από μια σκηνή, αναγνώρισε ένα ή περισσότερα άτομα χρησιμοποιώντας μια υπάρχουσα βάση δεδομένων από πρόσωπα. Συμπληρωματικές πληροφορίες, όπως ηλικία, φύλο, φυλή, μπορεί να είναι διαθέσιμες για κάθε αποθηκευμένο στη βάση πρόσωπο, ούτως ώστε να περιορίζεται η έρευνα. Διάφορα υποπροβλήματα πηγάζουν για την αυτοματοποίηση της παραπάνω διαδικασίας:

- (1) Απομόνωση και κατάτμηση του προσώπου μέσα από ένα –γενικά- θορυβώδες περιβάλλον.
- (2) Εξαγωγή ευσταθών χαρακτηριστικών από την περιοχή του προσώπου. Ευσταθή χαρακτηριστικά είναι εκείνα, τα οποία δεν επηρεάζονται από την κλίμακα, τον φωτισμό, την κλίση-στροφή του προσώπου, την αμφίεση και τη γήρανση.
- (3) Χρήση χαρακτηριστικών ή άλλων παρόμοιων τεχνικών με σκοπό τη μείωση της διάστασης του προβλήματος και τη δημιουργία προτύπων. Το πρόβλημα τελικά ανάγεται στην αναγνώριση προτύπων και τη χρήση κατάλληλων μετρικών για ταυτοποίηση-κατηγοριοποίηση.

Επιπλέον το πρόβλημα μπορεί να διαχωριστεί σε δύο κατηγορίες με βάση τη πηγή πληροφορίας, αν δηλαδή αναφέρεται σε σταθερές ή κινούμενες εικόνες. Η αναγνώριση προσώπου από σταθερές εικόνες έχει πλεονεκτήματα και μειονεκτήματα σε σχέση με την αναγνώριση από ακολουθίες βίντεο, ως εκ τούτου οι χρησιμοποιούμενες τεχνικές στις δύο αυτές κατηγορίες ποικίλουν.

Ο τρόπος με τον οποίο γίνεται η αναγνώριση προσώπων από τον άνθρωπο έχει απασχολήσει τους τομείς της Ψυχοφυσικής και της Νευροφυσιολογίας για περισσότερα από 20 χρόνια. Η μοναδικότητα των προσώπων, το ερώτημα αν η αναγνώριση γίνεται ολιστικά ή με χρήση επιμέρους χαρακτηριστικών, η ανάλυση των εκφράσεων του προσώπου για σκοπούς αναγνώρισης, η οργάνωση της μνήμης για τα πρόσωπα και διάφορα άλλα θέματα απασχόλησαν τις παραπάνω επιστήμες. Οι μηχανικοί έδειξαν ενδιαφέρον για την έρευνα αυτή με σκοπό την υλοποίηση ενός αποδοτικού συστήματος αναγνώρισης. Μερικά από τα αποτελέσματα είναι πράγματι ενδιαφέροντα αν και τελικά η αναγνώριση προσώπων από μηχανές αναπτύχθηκε ανεξάρτητα από τις μελέτες της Ψυχοφυσικής και Νευροφυσιολογίας.

Σημαντική προσπάθεια στον τομέα της αναγνώρισης προσώπων από μηχανές είχε γίνει τα πρώτα χρόνια της δεκαετίας του '70. Κατά τη διάρκεια της δεκαετίας του '80 υπήρξε μια σχετική στασιμότητα. Από τις αρχές της δεκαετίας του '90 παρατηρείται μια σημαντική δραστηριότητα και ανανεωμένο ενδιαφέρον, το οποίο εκφράζεται με τη δημοσίευση σε τακτική βάση άρθρων σε περιοδικά όπως το *IEEE Transactions on*

PAMI, και τη διοργάνωση ειδικών παρουσιάσεων –sessions- σε μεγάλα διεθνή συνέδρια (*IEEE ICIP*, *IEEE ICASSP*, *IEEE Multimedia*). Έναυσμα για την ανανέωση του ενδιαφέροντος και περαιτέρω δραστηριοποίηση έδωσε η εργασία [1]. Έκτοτε μια μεγάλη ποικιλία προσεγγίσεων έχει εμφανιστεί στη διεθνή βιβλιογραφία.

2. ΕΦΑΡΜΟΓΕΣ ΤΗΣ ΑΝΑΓΝΩΡΙΣΗΣ ΠΡΟΣΩΠΩΝ

Εμπορικές και νομικής φύσεως εφαρμογές της αναγνώρισης προσώπου καταγράφονται στον Πίνακα Α2.1. Τα δεδομένα εισόδου των εφαρμογών αυτών ποικίλουν από στατικές φωτογραφίες με ελεγχόμενο φόντο έως ακολουθίες βίντεο με κανένα περιορισμό στη λήψη και την ανάλυση. Δεδομένης της ανομοιογένειας των πηγών εισόδου είναι φυσικό επακόλουθο η ύπαρξη πολλών μεθόδων που έχουν σχέση με πεδία όπως η επεξεργασία, ανάλυση και κατανόηση εικόνας και η αναγνώριση προτύπων, που προσπαθούν να αντιμετωπίσουν επιμέρους ζητήματα.

Χονδρικά μπορεί κάποιος να ταξινομήσει τις τεχνικές και τα επιμέρους προβλήματα σε δύο κατηγορίες: στατικής ταύτισης -δεν υπάρχει δυνατότητα παρακιολούθησης κίνησης- και δυναμικής ταύτισης -στην οποία υπάρχει δυνατότητα παρακιολούθησης της χρονικής μεταβολής των δεδομένων. Ακόμη και στο πλαίσιο των ανωτέρω γενικών κατηγοριών υπάρχουν σημαντικές διαφοροποιήσεις ανάλογα με τη συγκεκριμένη εφαρμογή. Οι διαφορές που αφορούν στην ποιότητα της εικόνας -χρωική ανάλυση, στην πολυπλοκότητα του φόντου η οποία επηρεάζει αποφασιστικά την ευκολία με την οποία μπορεί να εντοπιστεί το πρόσωπο μέσα στην εικόνα, στην ύπαρξη ενός καλά ορισμένου κριτηρίου ταύτισης καθώς και στο ποσοστό, στο είδος και στην υφή της παρέμβασης από τον άνθρωπο -όπως στις εφαρμογές 4 και 5 του Πίνακα Α2.1. Σε μερικές εφαρμογές, όπως στον εντοπισμό των μεταβολών που επήλθαν με την γήρανση απαιτείται η εύρεση των μετασχηματισμών εκείνων που μετά την εφαρμογή τους οι νέες εικόνες θα είναι παρόμοιες με τις αναμενόμενες.

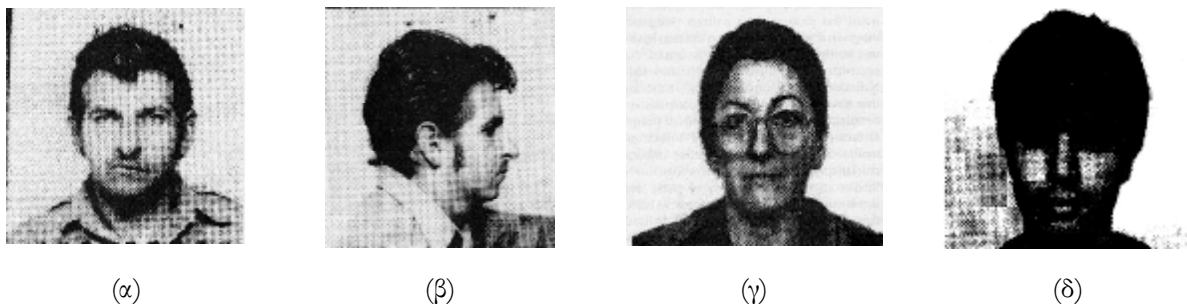
Με βάση τα προηγούμενα προκύπτουν τρία διαφορετικά προβλήματα που αφορούν στις εφαρμογές του Πίνακα Α2.1. Συγκεκριμένα η ταύτιση, η ανίχνευση ομοιότητας και η εφαρμογή μετασχηματισμών. Οι εφαρμογές 1, 2, 3, 10 και 11 θέτουν κυρίως προβλήματα ταύτισης μιας εικόνας προσώπου με κάποια άλλη. Οι εφαρμογές 4-7 αναφέρονται στη δημιουργία μιας εικόνας προσώπου όμοιας με αυτήν που θυμούνται ή περιγράφουν αυτόπτες μάρτυρες. Τέλος οι εφαρμογές 8 και 9 εμπλέκουν την εφαρμογή μετασχηματισμών και τροποποιήσεων σε μια εικόνα με βάση στοιχεία που είναι χρήσιμα σε άλλες εφαρμογές. Κάθε μια από τις εφαρμογές που αναφέρονται στον Πίνακα Α2.1 θέτει διαφορετικές απαιτήσεις και περιορισμούς στη διαδικασία της αναγνώρισης προσώπου. Η ταύτιση απαιτεί όπως η προς ταυτοποίηση εικόνα ανήκει σε ένα σύνολο από εικόνες οι οποίες έχουν επιλεγεί από το υπολογιστικό σύστημα. Η ανίχνευση ομοιότητας εκτός από την ταύτιση απαιτεί επιπλέον οι επιλεγείσες εικόνες να ομοιάζουν με αυτές που αυτόπτες μάρτυρες ανακαλούν στη μνήμη τους. Η τελευταία απαίτηση θέτει ένα σοβαρό περιορισμό στο κριτήριο ταύτισης το οποίο πρέπει να είναι αντίστοιχο με αυτό που χρησιμοποιούν οι άνθρωποι. Οι εφαρμογές μετασχηματισμού από την άλλη πλευρά πρέπει να προσομοιάζουν τον τρόπο με τον οποίο λειτουργεί η γήρανση -ή τουλάχιστον αυτό που αναμένουν οι άνθρωποι να συμβεί ως συνέπεια της παρόδου του χρόνου.

A/A	Εφαρμογές	Πλεονεκτήματα	Μειονεκτήματα
1(α)	Πιστωτικές κάρτες, ΑΤΜ μηχανήματα, δίπλωμα οδήγησης, διαβατήριο, προσωπική ταυτοποίηση.	Καλή ποιότητα εικόνων, ελεγχόμενες συνθήκες λήψης, εύκολος εντοπισμός του προσώπου	Μη υπάρχουσες βάσεις, μεγάλο μέγεθος βάσεων, σπάνια χρησιμοποιούμενη αναζήτηση
1(β)	Mug shots ταύτιση	Ελεγχόμενες συνθήκες λήψης, περισσότερες από μια φωτογραφίες για κάθε πρόσωπο	
2	Ασφάλεια τραπεζών / καταστημάτων	Υψηλή ζήτηση, γεωγραφικά και τοπικά περιορισμένη αναζήτηση.	Μη ελεγχόμενο φόντο, μικρός αριθμός φωτογραφιών
3	Επισκόπηση - παρακολούθηση πλήθους	Υψηλή ζήτηση, μικρό μέγεθος αρχείων, εμμετάλλευση κίνησης	Μη ελεγχόμενο φόντο, χαμηλή ποιότητα εικόνων, εκτέλεση σε πραγματικό χρόνο
4	Ταυτοποίηση από ειδικούς	Υψηλή ζήτηση, δυνατότητα αναβάθμισης των φωτογραφιών	Χαμηλή ποιότητα εικόνων, απαραίτητη η άδεια από τις αστυνομικές αρχές
5	Ανακατασκευή προσώπου με βάση μαρτυρίες	Όρια που τίθενται από τους αυτόπτες μάρτυρες	Αμφίβολη ομοιότητα
6	Ηλεκτρονικό βιβλίο από mug shots	Όρια που τίθενται από την περιγραφή	Κούραση παρατηρητή
7	Ηλεκτρονική παράταξη υπόπτων	Όρια που τίθενται από την περιγραφή	Κούραση παρατηρητή
8	Ανακατασκευή προσώπου από υπολειπόμενα στοιχεία	Υψηλή ζήτηση	Απαιτείται η χρήση δεδομένων από τη φυσιολογία
9	Υπολογισμός αλλαγών που επήλθαν με την πάροδο του χρόνου (aging)	Υψηλή ζήτηση	Απαιτείται η ύπαρξη συγκεκριμένου παραδείγματος
10	Αναζήτηση με βάση το περιεχόμενο	Χρήση χρώματος	Ετερογενείς βάσεις πολυμέσων
11	Υποδιαίρεση ειδήσεων σε θεματικές ενότητες.	Χρήση χρώματος	Ανομοιογένεια συνθηρών studio και εξωτερικών λήψεων

Πίνακας Α2.1: Συνθήκες εφαρμογές της αναγνώρισης προσώπων

3. ΠΡΟΒΛΗΜΑΤΑ ΣΤΑΤΙΚΗΣ ΤΑΥΤΙΣΗΣ

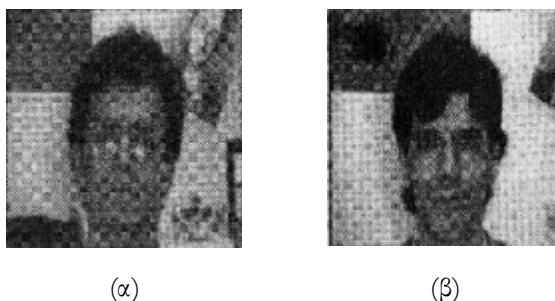
Η πιο συνηθισμένη εφαρμογή της κατηγορίας αυτής είναι η ταύτιση «mug shot» φωτογραφιών. Τυπικά στις φωτογραφίες αυτού του είδους, ο φωτισμός είναι ελεγχόμενος και λαμβάνονται δύο φωτογραφίες μια εμπρόσθιας όψης –ανφάς- και μία πλάγιας όψης -προφίλ. Παρόλο που γενικά οι προδιαγραφές λήψης των φωτογραφιών μπορούν να τεθούν από τις αστυνομικές αρχές, δεν υπάρχει κάποιο συγκεκριμένο πρότυπο και σαν αποτέλεσμα υπάρχει μια ποικιλία ως προς τη μορφή τους ανάλογα με την περιοχή. Οι προδιαγραφές που μπορούν να τεθούν αφορούν στο φόντο, στις συνθήκες φωτισμού, στην χωρική ανάλυση και στην απόσταση του εικονιζόμενου από την κάμερα. Θέτοντας περιορισμούς όπως οι προηγούμενοι, τόσο η διαδικασία εντοπισμού, όσο και η διαδικασία ταύτισης απλοποιούνται σε μεγάλο βαθμό. Παραδείγματα mug shots φωτογραφιών δίνονται στις Εικόνες Α2.1(α) και Α2.1(β)



Εικόνα Α2.1: (α) Mug shot φωτογραφία (front view) (β) Mug shot φωτογραφία (profile) (γ) (δ) Φωτογραφίες σε διπλώματα οδήγησης, διαβατήρια κλπ.

Παραλλαγές της ταύτισης mug shot φωτογραφιών είναι η αναγνώριση προσώπων σε διπλώματα οδήγησης, πιστωτικές κάρτες, διαβατήρια και ταυτότητες. Παραδείγματα τέτοιων φωτογραφιών δίνονται στις Εικόνες Α2.1(γ) και Α2.1(δ). Οι συνθήκες λήψης των φωτογραφιών σε τέτοιου είδους έγγραφα είναι σαφώς περισσότερο ελεγχόμενες από τις τυπικές mug shots φωτογραφίες.

Τυπικά οι φωτογραφίες σε mug shot εφαρμογές είναι καλής ποιότητας και σύμφωνες με τα πρότυπα που τίθενται από το νόμο. Δεδομένων και των σχετικά ελεγχόμενων συνθηκών σύλληψης, ο εντοπισμός και η κατάτμηση των προσώπων στις φωτογραφίες είναι σχετικά εύκολος. Οι μεγαλύτερες δυσκολίες έγκεινται στο μεγάλο μέγεθος των βάσεων από φωτογραφίες που υπάρχουν σε τέτοιου είδους εφαρμογές και φυσικά στη διαδικασία ταύτισης η οποία επηρεάζεται από μεταβολές στα πρόσωπα οι οποίες προέρχονται από γήρανση, απώλεια ή αλλαγή μαλλιών κλπ.



Εικόνα Α2.2: (α) (β) Φωτογραφίες με ανομοιόμορφο φόντο

Η εφαρμογή 2 είναι περισσότερο σύνθετη από την εφαρμογή 1 κυρίως εξαιτίας της μη ελεγχόμενης διαδικασίας λήψης των φωτογραφιών. Το φόντο δεν είναι υποχρεωτικά ομοιόμορφο και κατά συνέπεια η διαδικασία εντοπισμού και κατάτμησης καθίσταται δυσκολότερη. Επιπλέον η ποιότητα των εικόνων τείνει να είναι χαμηλή. Μια προσέγγιση φωτογραφιών της εφαρμογής αυτής δίνεται στις Εικόνες Α2.2(α) και (β). Πρέπει να τονιστεί ότι ουσιαστικά η εφαρμογή 2 βρίσκεται ανάμεσα στη στατική και τη δυναμική ταύτιση δεδομένου ότι σε πολλές περιπτώσεις εκτός από φωτογραφίες από συνήθεις φωτογραφικές μηχανές λαμβάνονται και εικόνες από βιντεοκάμερα. Όπως και στην εφαρμογή 1, μεταβολές στα πρόσωπα οι οποίες προέρχονται από γήρανση, μεταμπίωση και απώλεια ή αλλαγή μαλλιών πρέπει να λαμβάνονται

υπόψη στη διαδικασία ταύτισης και εξαγωγής χαρακτηριστικών. Στις εφαρμογές 1 και 2 το κριτήριο ταύτισης μπορεί να είναι ποσοτικό επιτρέποντας κατάταξη μεριών από τις επιλογές του συστήματος σύμφωνα με τη σειρά επιτυχίας.

Οι εφαρμογές 4-7 αφορούν την εύρεση ή δημιουργία ενός προσώπου όμοιου με αυτό που κάποιος αυτόπτης μάρτυρας είδε ή περιέγραψε. Στην εφαρμογή 4 ένας ειδικός καλείται να βεβαιώσει ότι το πρόσωπο στη δοθείσα εικόνα αντιστοιχεί στα αναζητούμενο άτομο. Είναι πιθανό το πρόσωπο της φωτογραφίας να είναι μεταμφιεσμένο ή να καλύπτεται μερικώς από άλλα πρόσωπα ή αντικείμενα. Τυπικά, της εφαρμογής αυτής προηγείται κάποια διαδικασία κατάταξης, η οποία δημιουργεί ένα σύνολο πιθανών προσώπων τα οποία έχουν σημαντικές ομοιότητες και ο ειδικός καλείται με προσεκτική έρευνα να βρει ποια από αυτές αντιστοιχεί στο αναζητούμενο άτομο. Στην εφαρμογή 5 ο αυτόπτης μάρτυρας καλείται να συνθέσει μια προσωπογραφία του αναζητούμενου ατόμου χρησιμοποιώντας μια βιβλιοθήκη από χαρακτηριστικά προσώπου όπως μάτια, μύτες, χείλη κλπ. Για παράδειγμα η βιβλιοθήκη μπορεί να περιέχει μύτες μακριές, κοντές, καμπυλωτές επίπεδες κοι, κάποια από τις οποίες θα ταιριάζει περισσότερο με αυτήν που ο μάρτυρας έχει στη μνήμη του. Η εφαρμογή 6 αναφέρεται στην ηλεκτρονική πλοήγηση σε συλλογή από φωτογραφίες. Στην εφαρμογή 7 ο μάρτυρας καλείται να αναγνωρίσει τον ύποπτο από ένα σύνολο φωτογραφιών μερικές από τις οποίες περιέχουν και λανθασμένες επιλογές. Τυπικά στις εφαρμογές 4-7 η ποιότητα των εικόνων είναι σχετικά χαμηλή και πέρα από την ταύτιση ζητείται και η εύρεση προσώπων που ομοιάζουν με αυτό που ο μάρτυρας έχει στη μνήμη του. Το κριτήριο ταύτισης είναι δύσκολο να καθοριστεί ποσοτικά, δεδομένου ότι πρέπει να βρεθούν μετρικές οι οποίες προσομοιάζουν την διαδικασία ταύτισης που πραγματοποιεί ο άνθρωπος. Αξίζει να σημειωθεί ότι όταν ο άνθρωπος καλείται να αναγνωρίσει κάποιο πρόσωπο από μια συλλογή από φωτογραφίες, τα λάθη αναγνώρισης αυξάνονται όσο αυξάνεται ο αριθμός των φωτογραφιών. Συμπερασματικά οι εφαρμογές 4-7 απαιτούν ισχυρή αλληλεπίδραση μεταξύ των αλγορίθμων που χρησιμοποιούνται και αποτελεσμάτων από τους τομείς της Ψυχοφυσικής και της Νευροφυσιολογίας.

Οι εφαρμογές 8 και 9 εμπλέκουν μετασχηματισμό της τρέχουσας εικόνας έτσι ώστε το εμφανιζόμενο πρόσωπο να μοιάζει με αυτό που αναμένεται μετά από την πάροδο κάποιου χρονικού διαστήματος ή με αυτό που θα έπρεπε να είναι. Οι εφαρμογές αυτές είναι ακόμη δυσκολότερες από τις εφαρμογές 4-6 δεδομένου ότι απαιτούν την ενσωμάτωση μηχανισμών εξομάλυνσης και πρόβλεψης στους αλγορίθμους.

4. ΠΡΟΒΛΗΜΑΤΑ ΔΥΝΑΜΙΚΗΣ ΤΑΥΤΙΣΗΣ

Η εφαρμογή 3 καθώς και περιπτώσεις της εφαρμογής 2 υπάγονται στη δυναμική ταύτιση, δεδομένου ότι είναι διαθέσιμη μια ακολουθία βίντεο και όχι μεμονωμένες φωτογραφίες. Εικόνες προερχόμενες από βιντεοκάμερες τείνουν να έχουν χαμηλή ποιότητα. Επιπλέον σε εφαρμογές επισκόπησης πλήθους το φόντο είναι εξαιρετικά ανομοιογενές, πράγμα που καθιστά τη διαδικασία εντοπισμού και κατάτμησης του προσώπου πολύ δύσκολη. Από την άλλη πλευρά, η ύπαρξη κίνησης αποτελεί ένα πολύ χρήσιμο στοιχείο για την κατάτμηση προσώπων από κινούμενα άτομα. Ακόμη, με αλγόριθμους ανακατασκευής 3-Δ αντικείμενων, με βάση τη κίνηση, μπορεί να επιτευχθεί μερική ανακατασκευή του προσώπου και να αντιμετωπιστούν θέματα μεταμφίεσης σχετικά καλύτερα από ότι στη στατική ταύτιση. Ένας από τους

ισχυρότερους περιορισμούς στις εφαρμογές δυναμικής ταύτισης είναι η απαίτηση για εκτέλεση σε πραγματικό χρόνο.

Πρέπει να τονιστεί ότι οι ευρέως διαφοροποιούμενοι περιορισμοί που τίθενται στις επιμέρους εφαρμογές καθιστούν αναγκαία τη χρήση διαφορετικών μεθόδων βαθμολόγησης για την αξιολόγηση των υπαρχόντων αλγορίθμων και συστημάτων.

5. ΤΑ ΒΑΣΙΚΑ ΥΠΟΠΡΟΒΛΗΜΑΤΑ ΣΤΗΝ ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΣΩΠΩΝ

Πριν την ταυτοποίηση ενός προσώπου πρέπει πρώτα αυτό να εντοπιστεί. Η ανίχνευση και ο εντοπισμός προσώπων είναι το πρώτο στάδιο της αναγνώρισης προσώπων. Το επόμενο στάδιο είναι η ανάπτυξη μηχανισμών ταυτοποίησης του προσώπου με βάση γνωστές πληροφορίες. Ένα σχετιζόμενο πρόβλημα είναι η αποτελεσματική αναπαράσταση του προσώπου, π.χ. μέσω ενός διδιάστατου πίνακα τιμών ή ενός συνόλου από χαρακτηριστικά. Προβλήματα αντίστοιχα με την αναγνώριση προσώπων είναι και αυτά της αναγνώρισης του φύλου ή της φυλής.

Αναπαράσταση Προσώπων

Η αναπαράσταση των προσώπων είναι ιδιαίτερα σημαντική στην αναγνώριση. Όλα τα γνωστά πρόσωπα πρέπει να αναπαριστώνται σε κάποια μορφή. Τα νεοεντοπισμένα πρόσωπα πρέπει επίσης να αναπαριστώνται σε κατάλληλη μορφή ώστε η διαδικασία ταύτισης να είναι αφενός εφικτή και αφετέρου αποτελεσματική. Ένα πρόσωπο είναι δύσκολο να αναπαρασταθεί πλήρως από τα μεμονωμένα χαρακτηριστικά του, π.χ. από το αν έχει μακριά μύτη, σκούρα μαλλιά και οβάλ σχήμα. Αυτός είναι και ο λόγος για τον οποίο οι αστυνομικές αρχές δημιουργούν σκίτσα με βάση τις περιγραφές των αυτόπτων μαρτύρων που περιγράφουν το πρόσωπο με βάση επιμέρους χαρακτηριστικά. Η αναπαράσταση των προσώπων πρέπει να είναι συμπαγής, αλλά χωρίς την απώλεια σημαντικής πληροφορίας. Για παράδειγμα, αναπαράσταση σε ανάλυση 512x512, με βάθος χρώματος 6 bpp, μπορεί να είναι επαρκής συγκριζόμενη με 8 bpp. Ομοίως φωτογραφίες με ανάλυση μικρότερη από 512x512 μπορεί να είναι ικανοποιητικές τόσο για τον εντοπισμό όσο και για την ταυτοποίηση. Ο τρόπος αναπαράστασης των προσώπων επηρεάζεται αλλά και επηρεάζει το σχήμα ταύτισης το οποίο χρησιμοποιείται για την αναγνώριση.

Ανίχνευση και Εντοπισμός Προσώπων

Σε πολλές περιπτώσεις οι συνθήκες κάτω από τις οποίες έχει ληφθεί η φωτογραφία είναι ελεγχόμενες, όπως π.χ. σε φωτογραφίες που πραγματοποιούνται από τις αστυνομικές αρχές. Ο εντοπισμός της ακριβούς θέσης του προσώπου στις φωτογραφίες αυτές είναι σχετικά εύκολος και σε αρκετές περιπτώσεις δεν είναι καν απαραίτητος. Στις περισσότερες περιπτώσεις εντούτοις, το εάν σε μια φωτογραφία υπάρχουν πρόσωπο, ή πρόσωπα, δεν είναι εκ των προτέρων γνωστό –πολύ περισσότερο δε, δεν είναι γνωστή η ακριβής θέση του. Διάφοροι παράγοντες, όπως η παρουσία μαλλιών που καλύπτουν μέρος του προσώπου, η ύπαρξη γενειάδας, το βήσιμο, η μερική επικάλυψη από άλλα αντικείμενα και, επηρεάζουν τη διαδικασία εντοπισμού του προσώπου σε μια εικόνα γιατί καλύπτουν μερικά από τα βασικά χαρακτηριστικά του. Ένας άλλος ουσιαστικός παράγοντας είναι η κλίμακα και η γωνία κλίσης του προσώπου στις φωτογραφίες. Το πρόβλημα της κλίμακας και του προσανατολισμού αποτρέπει τον εντοπισμό μέσω ενός απλού προτύπου

προσώπου. Τυπικά οι φωτογραφίες προσώπων λαμβάνονται, είτε κοιτάζοντας την κάμερα –εμπρόσθια όψη– είτε σε γωνία 90 μοιρών από αυτή –πλάγια όψη (προφίλ).

Ταυτοποίηση Προσώπων

Το επόμενο στάδιο μετά τον εντοπισμό είναι η ταυτοποίηση. Τα γνωστά πρόσωπα και ενδεχομένως τα βασικά χαρακτηριστικά τους πρέπει να βρίσκονται αποθηκευμένα σε κάποια βάση δεδομένων. Ο στόχος είναι η ταύτιση του προσώπου της φωτογραφίας με κάποιο από τα αποθηκευμένα στη βάση δεδομένων. Δύο γενικές κατηγορίες ταύτισης υπάρχουν: η ολιστική και η βασισμένη στα χαρακτηριστικά. Σημειώνεται ότι τεχνικές που βασίζονται σε φωτογραφίες προφίλ χρησιμοποιούν εντελώς διαφορετικά χαρακτηριστικά από τις αντίστοιχες που βασίζονται σε εμπρόσθια όψη. Παρόλα αυτά, σε κάθε περίπτωση το κεντρικό πρόβλημα είναι η ταύτιση. Μερικές επιπλέον δυσκολίες πηγάζουν από την αλλαγή των προσώπων και των χαρακτηριστικών τους με την πάροδο του χρόνου (γήρανση).

Ταξινόμηση με βάση τα φυσικά χαρακτηριστικά

Άλλη μια διαδικασία την οποία οι άνθρωποι πραγματοποιούν χωρίς ιδιαίτερη προσπάθεια είναι ταξινομήσεις με βάση το φύλο, την ηλικία, τη φυλή και το αναμενόμενο επάγγελμα. Το πρόβλημα αυτό είναι ιδιαίτερα πολυσύνθετο και βασίζεται κατά μεγάλο μέρος στην εμπειρία. Η ταξινόμηση φύλου και φυλής έχει μελετηθεί από διάφορους ερευνητές και αποδεικνύεται ότι υπάρχουν χαρακτηριστικά του προσώπου με βάση τα οποία είναι εφικτή μια τέτοια κατηγοριοποίηση.

Η ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΣΩΠΟΥ ΑΠΟ ΤΟΝ ΑΝΘΡΩΠΟ

Το ανθρώπινο σύστημα αναγνώρισης χρησιμοποιεί ένα ευρύ φάσμα από διεγέρσεις προερχόμενες από όλες σχεδόν τις αισθήσεις. Ένα σύστημα το οποίο μιμείται κάτι τέτοιο πρέπει να θεωρείται ουτοπία -τουλάχιστον σήμερα- επομένως αυτό που επιδιώκεται είναι η κατά το δυνατόν πλησιέστερη προσέγγιση του. Ενδιαφέρον παρουσιάζει και η διαδικασία με την οποία ο άνθρωπος αναγνωρίζει πρόσωπα σε εικόνες ή ακολουθίες βίντεο. Στη τελευταία περίπτωση ένα υπολογιστικό σύστημα μπορεί να πλεονεκτεί σε σχέση με το ανθρώπινο, υπό την έννοια ότι ο άνθρωπος μπορεί να αναγνωρίσει περιορισμένο αριθμό ατόμων και επιπλέον κουράζεται στην προσπάθεια αναγνώρισης προσώπων από μια βάση.

1. ΑΠΟΤΕΛΕΣΜΑΤΑ ΑΠΟ ΤΟΝ ΤΟΜΕΑ ΤΗΣ ΨΥΧΟΦΥΣΙΚΗΣ ΚΑΙ ΝΕΥΡΟΦΥΣΙΟΛΟΓΙΑΣ

Η αναγνώριση προσώπου είναι μια τυποποιημένη διαδικασία.

Στην άποψη αυτή συνηγορούν τα εξής: (α) Τα πρόσωπα είναι πιο εύκολα αναγνωρίσιμα, όταν παρουσιάζονται ανάποδα, σε σχέση με άλλα αντικείμενα, (β) ασθενείς που πάσχουν από Προσωπαγνωσία, δεν έχουν καμιά άλλη μορφή αμνησίας και μπορούν να αναγνωρίσουν ανθρώπους από τις φωνές, το ντύσιμο και το χρώμα μαλλιών, παρόλο που αντιλαμβάνονται τα μάτια, τη μύτη, το στόμα και, δεν μπορούν να τα συνδυάσουν για σκοπούς αναγνώρισης, (γ) Τα νεογνά φαίνεται να έλκονται περισσότερο από πρόσωπα παρά από άλλα αντικείμενα.

Η αντίληψη του προσώπου γίνεται ολιστικά ή με βάση μεμονωμένα χαρακτηριστικά;

Οι μελέτες δείχνουν ότι αρχικά γίνεται μια χονδρική εκτίμηση με βάση όλο το πρόσωπο και στη συνέχεια λεπτομερέστερη βασισμένη σε επιμέρους χαρακτηριστικά. Όταν υπάρχουν προεξέχοντα χαρακτηριστικά (μεγάλα ή πεταχτά αυτιά, μύτη με παράξενο σχήμα κλπ.) τότε η ανθρώπινη προσοχή επικεντρώνεται πολύ γρήγορα σε αυτά και η χονδρική εκτίμηση παραλείπεται.

Η σημαντικότητα των επιμέρους χαρακτηριστικών.

Τα μαλλιά, τα μάτια, το περίγραμμα του προσώπου και το στόμα -όχι απαραίτητα με την παραπάνω σειρά- θεωρούνται τα πιο σημαντικά χαρακτηριστικά για σκοπούς αντίληψης και αναγνώρισης ενός προσώπου. Η μύτη παίζει σημαντικό ρόλο μόνο σε εικόνες προφίλ. Γενικότερα το άνω μέρος του προσώπου είναι περισσότερο σημαντικό από το κάτω. Τέλος τα ελκυστικά πρόσωπα παρουσιάζουν μεγαλύτερο βαθμό αναγνώρισης, στη συνέχεια ακολουθούν τα άσχημα, ενώ δυσκολότερα αναγνωρίζονται τα πρόσωπα με κοινά χαρακτηριστικά.

Πρωτοτυπία

Παράξενα πρόσωπα αναγνωρίζονται ευκολότερα, αν και σε μια απόφαση κατά πόσο ένα αντικείμενο είναι πρόσωπο ή όχι, ένα τυπικό πρόσωπο αναγνωρίζεται γρηγορότερα από κάποιο παράξενο.

Ο ρόλος των χωρικών συχνότητων

Αρχικά είχε προταθεί ότι οι χαμηλές χωρικές συχνότητες παίζουν ένα πολύ σημαντικό ρόλο στην αναγνώριση προσώπου. Στην πραγματικότητα κάθε περιοχή συχνότητων έχει ξεχωριστή σημασία, π.χ. για διαχωρισμό φύλου αρχούν οι χαμηλές συχνότητες, ενώ αντίθετα για σκοπούς αναγνώρισης απαιτούνται και οι υψηλές συχνότητες.

Ανάπτυξη του ανθρώπινου συστήματος αναγνώρισης κατά τη διάρκεια της παιδικής ηλικίας

Παιδιά κάτω των δέκα χρόνων κωδικοποιούν πρόσωπα με βάση μεμονωμένα χαρακτηριστικά. Η αναγνώριση στηρίζεται σε τέτοια χαρακτηριστικά καθώς και σε αντικείμενα που φέρουν μαζί τους οι άνθρωποι όπως γυαλιά, καπέλο κλπ. Μετά την ηλικία των δέκα χρόνων η ανάλυση μεταβάλλεται από μεμονωμένα χαρακτηριστικά σε ολιστική.

Ο ρόλος του φύλου-φυλής

Άνθρωποι αναγνωρίζουν άτομα που ανήκουν στη δική τους φυλή ευκολότερα από αυτά που ανήκουν σε άλλη. Αυτό μπορεί να οφείλεται στη δημιουργία και κωδικοποίηση ενός «μέσου» προσώπου με «μέσες» ιδιότητες με βάση την καθημερινή εμπειρία. Αποδεικνύεται επίσης ότι τα πρόσωπα γυναικών είναι περισσότερο ετερογενή από τα αντίστοιχα των ανδρών.

Συμπεράσματα

Για τους μηχανικούς, οι οποίοι ενδιαφέρονται να σχεδιάσουν και να υλοποιήσουν αλγορίθμους για αναγνώριση προσώπων, πληθώρα εργασιών από τους τομείς της Ψυχοφυσικής και Νευροφυσιολογίας μπορούν να χρησιμεύσουν ως οδηγοί. Ένα κλασσικό παράδειγμα το οποίο πρέπει να ληφθεί υπόψη είναι η χρήση τόσο ολιστικών χαρακτηριστικών όσο και χαρακτηριστικών προσώπου για την αναγνώριση. Όσον αφορά στα χαρακτηριστικά προσώπου, μερικά από αυτά όπως τα μαλλιά, τα μάτια και το στόμα είναι σημαντικότερα από άλλα όπως η μύτη. Αυτό βέβαια είναι αληθές μόνο για εικόνες εμπρόσθιας όψης δεδομένου ότι για εικόνες προφίλ η μύτη αποτελεί ένα πολύ σημαντικό χαρακτηριστικό. Επίσης εργασίες σχετικές με την ιδιαιτερότητα των προσώπων με παράξενα χαρακτηριστικά καθώς και αντίστοιχες που αφορούν καρικατούρες μπορούν να φανούν χρήσιμες για την προσθήκη χαρακτηριστικών τα οποία υποβοηθούν τη διαδικασία της αναγνώρισης. Η σημασία της χωρικής ανάλυσης θέτει ένα καλό υπόβαθρο για τη χρήση πολυδιακριτικών μεθόδων (multiresolution) για την αντιμετώπιση επιμέρους προβλημάτων που αφορούν την αναγνώριση προσώπων. Ζητήματα όπως ο ρόλος του γένους και της φυλής, καθώς και πώς τα νήπια αναγνωρίζουν τα πρόσωπα, είναι ιδιαίτερα σημαντικά σε επιμέρους εφαρμογές, όπως η ανακατασκευή προσώπου με βάση μαρτυρίες, η ταυτοποίηση από ειδικούς και η επισκόπηση ηλεκτρονικού βιβλίου από mug shot φωτογραφίες. Τέλος η οργάνωση της μνήμης για την αποθήκευση προσώπων μπορεί να χρησιμεύσει στην αποδοτική αποθήκευση mug shot φωτογραφιών σε βάσεις δεδομένων.

Μεγάλο ενδιαφέρον από τους σχεδιαστές συστημάτων αναγνώρισης προσώπων έχει συγκεντρώσει η γνώση της λειτουργίας του ανθρώπινου οπτικού συστήματος και η μεταφορά των μηχανισμών του σε πρακτικά συστήματα. Πρέπει να σημειωθεί πάντως ότι η χρήση των εργασιών από τους τομείς της Ψυχοφυσικής και

Νευροφυσιολογίας πρέπει να γίνεται μέχρι τον βαθμό εκείνο που αυτές είναι πρακτικά εφαρμόσιμες, διαφορετικά είναι δυνατό να μας οδηγήσουν σε αποπροσανατολισμό.

2. ΟΙ ΔΥΝΑΤΟΤΗΤΕΣ ΤΟΥ ΟΠΤΙΚΟΥ ΣΥΣΤΗΜΑΤΟΣ ΤΩΝ ΑΝΘΡΩΠΩΝ

Οι διεργασίες που πραγματοποιεί το ανθρώπινο οπτικό σύστημα καθώς και οι δυνατότητές του, όσον αφορά στην αναγνώριση προσώπων, καλό είναι να μελετηθούν ώστε να βοηθήσουν, όπου αυτό είναι εφικτό, στη σχεδίαση ενός αυτοματοποιημένου συστήματος. Ούτως ή άλλως το ανθρώπινο οπτικό σύστημα αποτελεί το καλύτερο σύστημα αναγνώρισης προσώπων που γνωρίζουμε. Μπορεί να μην είναι εφικτό να υλοποιηθούν όλες οι δυνατότητές του, και σε μερικές περιπτώσεις μπορεί αυτό να μην είναι καν επιθυμητό, αλλά ως πηγή αναφοράς η αξία του είναι αναμφισβήτητη.

Αναπαράσταση

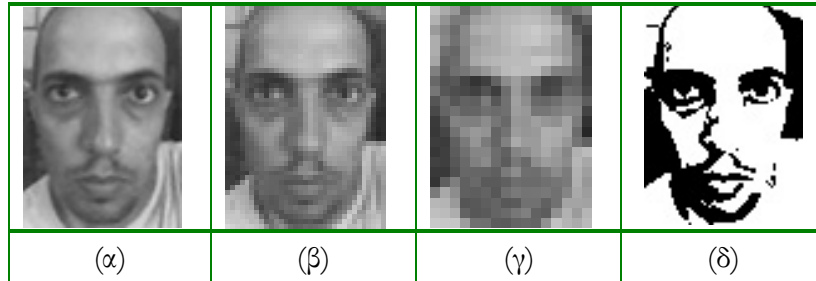
Ο τρόπος αναπαράστασης και κωδικοποίησης των προσώπων από τον άνθρωπο δεν είναι αποσαφηνισμένος. Υπάρχουν κάποιες ενδείξεις ότι η κωδικοποίηση γίνεται είτε με τη μορφή πρωτογενούς σκαριφήματος, είτε με πλήρη απομνημόνευση. Αυτό που μοιάζει περισσότερο σίγουρο, είναι ότι η εσωτερική αναπαράσταση οικείων προσώπων γίνεται με διαφορετικό τρόπο από την αντίστοιχη αγνώστων. Δεν υπάρχει καμία τεκμηριωμένη άποψη για την υφή της εσωτερικής αναπαράστασης ούτε για το μέγεθος που αυτή καταλαμβάνει.

Εντοπισμός και κατάτμηση προσώπου

Οι άνθρωποι ανιχνεύουν, ακόμη και μη οικεία πρόσωπα σε μια σκηνή με τυπική παρατήρηση και χωρίς κάποια ιδιαίτερη προσπάθεια. Το γεγονός αυτό υποδηλώνει, είτε ότι υπάρχουν ειδικές μονάδες στον εγκέφαλο για το σκοπό αυτό, πράγμα για το οποίο πολύ λίγες ενδείξεις υπάρχουν, είτε ότι η ανίχνευση προσώπων πραγματοποιείται πολύ αποτελεσματικά μέσω μαζικής παράλληλης επεξεργασίας. Επίσης το πρόσωπο γίνεται αντιληπτό ως ολότητα και όχι ως ένα σύνολο από χαρακτηριστικά. Για παράδειγμα, οι άνθρωποι αναγνωρίζουν χωρίς δυσκολία μερικώς επικλυπτόμενα πρόσωπα -το σύστημα αντίληψης μας συμπληρώνει το υπολειπόμενο τμήμα. Επιπλέον το ανθρώπινο οπτικό σύστημα είναι πολύ εύρωστο όσον αφορά στην ανίχνευση προσώπων. Τα πρόσωπα εντοπίζονται με μηδενική προσπάθεια κάτω από ποικίλες συνθήκες, όπως κακός φωτισμός, μεγάλη απόσταση και μερική επικάλυψη. Το ανθρώπινο σύστημα αναγνώρισης προσώπων έχει επίσης τη τάση να αναγνωρίζει αντικείμενα ως πρόσωπα, ενώ αν κάποιο αντικείμενο χαρακτηριστεί ως πρόσωπο, είναι πολύ δύσκολο σε κάποια άλλη στιγμή να θεωρηθεί ως κάτι άλλο.

Πολλές φορές η ανίχνευση προσώπων σε μια σκηνή είναι ευκολότερη όταν κοιτάμε από κάποια απόσταση, ή από πλάγια. Η διαδικασία αυτή στην ουσία δημιουργεί θόλωμα στη σκηνή -φιλτράρισμα των υψηλών χωρικών συχνοτήτων- και κάποια βασικά χαρακτηριστικά του προσώπου, όπως οι κόρες των ματιών και οι βλεφαρίδες, γίνονται λιγότερο έντονα. Στον εντοπισμό του προσώπου η παρουσία των βασικών χαρακτηριστικών και η χωροταξική τους διάταξη είναι σημαντικότερη από την λεπτομερή αναπαράστασή τους. Μια βασική παράμετρος στην ανίχνευση προσώπων είναι η ελάχιστη δυνατή ανάλυση που πρέπει να έχει κάποιο αντικείμενο ώστε να αναγνωριστεί ως πρόσωπο. Στην Εικόνα Α3.1 φαίνεται ένα πρόσωπο σε

αναλύσεις (α) 64x88 (β) 32x44 (γ) 16x22. Το πρόσωπο είναι εύκολα ανιχνεύσιμο σε ανάλυση 32x44 ενώ στη περίπτωση της ανάλυσης 16x22 είναι ακόμη διακριτό, όταν το κοιτάζουμε από απόσταση. Ανάλυση 16x16 πιστεύεται ότι είναι η ελάχιστη για ανίχνευση προσώπων από το ανθρώπινο οπτικό σύστημα. Σημειώνετε ότι η ανάλυση αυτή αφορά στην ανίχνευση και όχι στην αναγνώριση.



Εικόνα A3.1: Ανθρώπινο πρόσωπο σε ανάλυση (α) 64x88 (β) 32x44 (γ) 16x22 (δ) 64x88 και κβαντισμό χρώματος 1 brrp

Ένα αντίστοιχο θέμα είναι ο απαιτούμενος κβαντισμός χρώματος σε μονόχρωμες εικόνες. Όπως φαίνεται στην Εικόνα A3.1(δ), ακόμη και με κβαντισμό 1brrp, το πρόσωπο είναι άμεσα ανιχνεύσιμο υπό την προϋπόθεση ότι η χωρική ανάλυση είναι επαρκής. Πειραματικά αποτελέσματα δείχνουν ότι χωρική ανάλυση 32x32 με κβαντισμό σε 4 brrp είναι ικανοποιητική για την ανίχνευση προσώπου από τον άνθρωπο.

Ταυτοποίηση

Στην καθημερινή μας ζωή ταυτοποίηση ονομάζουμε την αντιστοίχιση ενός προσώπου με κάποιο όνομα. Στο κόσμο των υπολογιστών σωστή ταυτοποίηση σημαίνει ότι διαφορετικές εικόνες του ίδιου προσώπου πρέπει να αναγνωρίζονται ως μία. Υπολογίζεται ότι κάθε άνθρωπος γνωρίζει προσωπικά 700 περίπου πρόσωπα και μερικές χιλιάδες ακόμη εξ όψεως. Κατά τη διάρκεια της ζωής μας ερχόμαστε σε οπτική επαφή με μερικές δεκάδες χιλιάδες πρόσωπα.

Η ταυτοποίηση οικείων και διασήμων προσώπων από τον άνθρωπο πραγματοποιείται σε μηδενικό χρόνο. Αντίθετα η ταυτοποίηση μη οικείων προσώπων ή προσώπων τα οποία παρουσιάζονται σε ασυνήθιστες οριοθετήσεις -π.χ. ανάποδα- παίρνει πολύ μεγαλύτερο χρόνο. Το ανθρώπινο σύστημα ταυτοποίησης προσώπων είναι επίσης πολύ εύρωστο. Η διαδικασία ταυτοποίησης πραγματοποιείται αποτελεσματικά κάτω από ποικίλες συνθήκες φωτισμού, μετασχηματισμών και επικάλυψης του προσώπου. Υπάρχουν ενδείξεις ότι για κάθε οικείο πρόσωπο υπάρχει μια μονάδα στην εγκέφαλο η οποία το εγγράφει κάτω από όλες τις συνθήκες. Περισσότερο οικεία πρόσωπα απαιτούν λιγότερη πληροφορία για ταυτοποίηση συγκρινόμενα με λιγότερο οικεία. Κατά συνέπεια οικεία πρόσωπα δυσκολότερα ταυτοποιούνται λανθασμένα. Ο Bartlett [2] θεωρεί ότι όλα τα άγνωστα πρόσωπα έχουν τον ίδιο βαθμό οικειότητας (μηδέν) αλλά νέα πρόσωπα με τυπικά χαρακτηριστικά δίνουν μεγαλύτερη αίσθηση οικειότητας από παράξενα πρόσωπα. Παρόλα αυτά μετά από την πρώτη επαφή με παράξενα πρόσωπα, ο βαθμός αύξησης της οικειότητας για αυτά είναι μεγαλύτερος από τα αντίστοιχα με τυπικά χαρακτηριστικά.

Σε αντίθεση με την ικανότητα μας να ανιχνεύουμε και να ταυτοποιούμε πρόσωπα με μεγάλη ευκολία, τα περιγράψουμε πολύ δυσκολότερα. Η συνηθέστερη μέθοδος περιγραφής προσώπων είναι με απαρίθμηση

των επιμέρους χαρακτηριστικών: καστανά μαλλιά, πράσινα μάτια, επίπεδη μύτη κλπ. Τα επιμέρους χαρακτηριστικά του προσώπου, καθώς και η χωροταξική τοποθέτηση τους, είναι πολύ σημαντικά στη διαδικασία της ταυτοποίησης. Τα σημαντικότερα από αυτά είναι τα μάτια, η μύτη, τα χείλη, τα αυτιά και τα μαλλιά. Σε καμία πάντως περίπτωση το πρόσωπο δεν ταυτοποιείται ως ένα σύνολο από χαρακτηριστικά.

Έχει προταθεί [3] ότι η σημασία των χαρακτηριστικών του προσώπου στη διαδικασία της ταυτοποίησης μειώνεται από πάνω προς τα κάτω. Θα μπορούσε επομένως κάποιος να υποθέσει ότι τα μαλλιά και τα μάτια είναι πιο σημαντικά από το πηγούνι και τα χείλη. Παρόλα αυτά πρόσωπα με παράξενο πηγούνι ή χείλη ταυτοποιούνται πάντοτε ευκολότερα. Δεδομένου ότι τα χαρακτηριστικά είναι πολύ σημαντικά στη διαδικασία της ταυτοποίησης το ερώτημα είναι «Πόσα χαρακτηριστικά απαιτούνται για το διαχωρισμό διαφορετικών προσώπων και την αποτελεσματική ταυτοποίησή τους». Έχει εκτιμηθεί [3] ότι ο αριθμός των απαιτούμενων χαρακτηριστικών αυξάνει λογαριθμικά με τον αριθμό των διαφορετικών προσώπων. Ένας άνθρωπος που μπορεί να ταυτοποιήσει 1000 διαφορετικά πρόσωπα χρειάζεται περίπου 10 χαρακτηριστικά, αριθμός ιδιαίτερα μικρός. Βέβαια, δεδομένου ότι κανείς δεν μπορεί να προσδιορίσει επακριβώς ποια είναι αυτά τα χαρακτηριστικά, ένα αυτόματο σύστημα αναγνώρισης προσώπων θα χρειαζόταν σίγουρα πολύ περισσότερα.

Ταξινόμηση με βάση τα φυσικά χαρακτηριστικά

Τυπικά όταν βλέπουμε κάποιο πρόσωπο το συσχετίζουμε με ένα γενικό χαρακτηρισμό όπως νέος, άνδρας, παιδί. Οι άνθρωποι επικεντρώνονται σε τρεις κατηγορίες χαρακτηρισμών οι οποίες σχετίζονται με το φύλο: άρρεν ή θήλυ, την ηλικία: παιδί, νέος, μεσήλικας, ηλικιωμένος και τη φυλή: ινδοευρωπαίος, ασιάτης, νέγρος. Από τις τρεις προηγούμενες ταξινομήσεις πραγματοποιούμε την πρώτη με χαρακτηριστική ευκολία ενώ αντίθετα οι άλλες είναι περισσότερο επιρρεπείς σε σφάλματα δεδομένου ότι απαιτούν μεγαλύτερη εμπειρία και γνώση. Η διαδικασία της ταξινόμησης με βάση τα φυσικά χαρακτηριστικά πιστεύεται ότι είναι μια ανωτέρου επιπέδου διαδικασία και πραγματοποιείται μετά τη ταυτοποίηση του προσώπου. Η φύση της διαδικασίας αυτής δεν είναι μέχρι τώρα πλήρως κατανοητή.

ΤΕΧΝΙΚΕΣ ΑΥΤΟΜΑΤΗΣ ΑΝΑΓΝΩΡΙΣΗΣ ΠΡΟΣΩΠΩΝ

Η διαδικασία της αναγνώρισης μπορεί να διαιρεθεί σε τρία διαδοχικά στάδια. Σε μια δοσμένη εικόνα ερευνάται, αν αυτή απεικονίζει κάποιο πρόσωπο, το οποίο και εξάγεται από τον περιβάλλοντα χώρο. Στη συνέχεια έχουμε αναπαράσταση του προσώπου με την εξαγωγή διαφόρων χαρακτηριστικών ώστε να μειωθεί η διάσταση του προβλήματος. Τέλος ακολουθεί η διαδικασία της ταυτοποίησης του προσώπου με κάποιο υπάρχον στη βάση ή η προσθήκη του σε αυτήν.

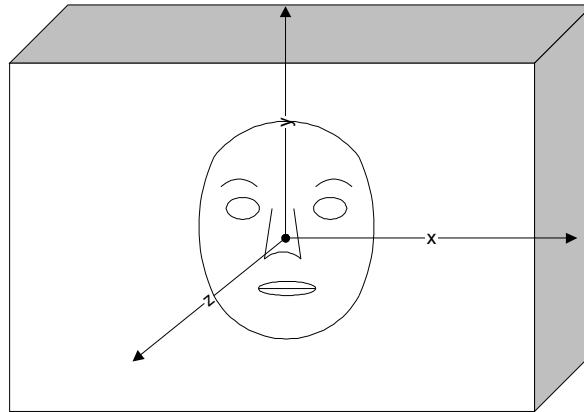
Στο κεφάλαιο αυτό πραγματοποιείται μια ανασκόπηση των τεχνικών που έχουν εμφανιστεί στη βιβλιογραφία σχετικά με τα τρία επιμέρους προβλήματα της αναγνώρισης προσώπων. Σε κάποιες περιπτώσεις δίνονται και μερικές προσομοιώσεις, οι οποίες πραγματοποιήθηκαν από τον συγγραφέα της διατριβής, για βελτίωση της κατανόησης των αλγορίθμων αλλά και για να δοθεί η ευκαιρία να τονιστούν και να σχολιαστούν κάποια κρίσιμα ζητήματα.

1. ΕΙΣΑΓΩΓΗ

Η έρευνα στην αναγνώριση προσώπων ξεκίνησε από τα μέσα του περασμένου αιώνα από τον Francis Galton. Πρέπει να τονιστεί ότι στη γενική περίπτωση κανένα από τα επιμέρους προβλήματα –βλέπε Κεφάλαιο Α2- που σχετίζονται με την αναγνώριση προσώπων δεν έχει λυθεί και πρακτικά μόνο τα προβλήματα της ανίχνευσης και της ταυτοποίησης έχουν προσελκύσει ουσιαστικό ενδιαφέρον από τους ερευνητές. Στη περίπτωση της ταυτοποίησης η θέση του προσώπου είναι εκ των προτέρων γνωστή ή μπορεί να εκτιμηθεί με ευριστικό τρόπο. Σε πολλές περιπτώσεις ισχυρές παραδοχές τίθενται για να διευκολύνουν το πρόβλημα της ταυτοποίησης:

- (1). Το απεικονιζόμενο πρόσωπο είναι σε εμπρόσθια όψη ή σε πλάγια όψη (προφίλ). Με την παραδοχή αυτή είναι ευκολότερο να εντοπιστούν τα βασικά χαρακτηριστικά του.
- (2). Το πρόσωπο έχει μηδενική ή πολύ μικρή κλίση (περί τον άξονα z που περνά από το κέντρο της φωτογραφίας - βλ. Σχήμα Α4.1).
- (3). Δεν υπάρχει επικάλυψη τμήματος του προσώπου.
- (4). Τα μαλλιά δεν επικαλύπτουν τμήμα του προσώπου, το πρόσωπο δεν φορά σκούρα γυαλιά και δεν υπάρχουν ουλές σε αυτό.

Πρέπει να αναφερθεί ότι οι περισσότερες από τις βάσεις δεδομένων που έχουν χρησιμοποιηθεί αποτελούνται από λευκούς άνδρες και περιλαμβάνουν σχετικά μικρό αριθμό φωτογραφιών.



Σχήμα A4.1: Σύστημα συντεταγμένων με βάση τη φωτογραφία

Αναφορά	Πλήθος Φωτογραφιών	Ανάλυση	Κβαντισμός Χρώματος	Σχόλια
Baron	150	512x480	6 bpp	Η ανάλυση μειώθηκε σε 128x120
Buhr	100	512x512	8 bpp	-
Campell	50	256x256	6 bpp	-
Craw et al	1000	128x128	-	Ο έλεγχος έγινε σε μόνο 20 πρόσωπα
Goldstein et al	255	-	-	Προφίλ τα οποία σχεδιάστηκαν από ζωγράφους
Harmon et al	124	-	-	97 άνδρες +27 γυναίκες
Harmon et al	130	-	-	81 άνδρες +49 γυναίκες
Kaufman	120	240x356	1 bpp	-
Nixon	6	128x128	8 bpp	-
Petajan	-	244x248	-	Για εντοπισμό χειλιών μόνο
Riccia	506	-	-	Τα δεδομένα εισάγονται από τον χρήστη
Wong	6	256x256	8 bpp	-
Wu and Huang	18	-	-	-
Olivetti R.L.	400	112x92	8 bpp	40 πρόσωπα από 10 φωτογραφίες το καθένα

Πίνακας A4.1: Δεδομένα που χρησιμοποιήθηκαν σε διάφορες μελέτες αναγνώρισης προσώπων

Στον Πίνακα A4.1 δίνονται κάποια στοιχεία για τις βάσεις εικόνων και τη μορφή των δεδομένων που χρησιμοποιήθηκαν, σε διάφορες εργασίες. Όπου λείπουν στοιχεία αυτά δεν αναφέρονται στη συγκεκριμένη

εργασία. Σε πολλές περιπτώσεις τα δεδομένα εισόδου στο σύστημα δημιουργήθηκαν χειρωνακτικά (manually).

2. ΑΝΑΠΑΡΑΣΤΑΣΗ

Δύο είδη αναπαράστασεων χρησιμοποιούνται στη αναγνώριση και ταυτοποίηση προσώπων: Δισδιάστατοι πίνακες τιμών και διανύσματα χαρακτηριστικών (τα χαρακτηριστικά εδώ δεν αναφέρονται στα χαρακτηριστικά του προσώπου αλλά σε κάποιες τιμές οι οποίες περιγράφουν με αποτελεσματικό τρόπο τη δομή του).

2.1 Αναπαράσταση προσώπων - Εξαγωγή χαρακτηριστικών

Αποδοτική αναπαράσταση ενός προσώπου θεωρείται κάθε αναπαράσταση η οποία μειώνει τη διάσταση της εικόνας διατηρώντας όσο το δυνατό μεγαλύτερο όγκο πληροφορίας. Σε πολλές από τις αρχικές μελέτες η αναπαράσταση των προσώπων γινόταν με χρήση γεωμετρικών χαρακτηριστικών, όπως αποστάσεων, γωνιών και λόγων αποστάσεων μεταξύ βασικών σημείων του προσώπου. Τέτοια σημεία έπρεπε να μπορούν να εντοπίζονται εύκολα και με ακρίβεια. Μια τέτοια αντιμετώπιση δε λαμβάνει υπόψη πληροφορία από την υφή του προσώπου, γεγονός που οδήγησε στην επέκταση του συνόλου των χρησιμοποιούμενων χαρακτηριστικών συμπεριλαμβάνοντας σε αυτά και πληροφορία υφής, είτε με άμεση μορφή (πίνακες φωτεινότητας διαφόρων περιοχών), είτε με χρήση στατιστικών μεγεθών (moment invariants), είτε με τη βοήθεια κάποιων μετασχηματισμών (Fourier descriptors).

Αξίζει να σημειωθεί ότι τελικά με τον όρο χαρακτηριστικά αναφερόμαστε σε κάθε αναπαράσταση του προσώπου σε μικρότερη διάσταση, είτε αυτή γίνεται με γεωμετρικά μεγέθη, είτε με πίνακες, είτε με στατιστικά μεγέθη. Με αυτό υπόψη μπορούμε να διακρίνουμε τέσσερις κατηγορίες [7]:

(a) Οπτικά χαρακτηριστικά.

Στη κατηγορία αυτή περιλαμβάνονται τα γεωμετρικά μεγέθη, οι πίνακες περιγραφής της υφής περιοχών, οι ακμές κ.ο.κ. Η εξαγωγή τέτοιων χαρακτηριστικών βασίζεται σε παραδοσιακές τεχνικές της επεξεργασίας εικόνας όπως το χωρικό φιλτράρισμα, η ταύτιση με πρότυπο, τα ενεργά περιγράμματα –active contours– κ.ο.κ. Περισσότερα στοιχεία για το θέμα αυτό παρουσιάζονται στο Κεφάλαιο Β1. Στον Πίνακα Α4.2 παρουσιάζονται τέτοια χαρακτηριστικά, τα περισσότερα από τα οποία έχουν χρησιμοποιηθεί κατά κόρον σε παλαιότερες μελέτες. Τονίζεται πάντως ότι πολλά από αυτά χρησιμοποιήθηκαν σε μελέτες στις οποίες η ταύτιση πραγματοποιείται από ανθρώπους και όχι από υπολογιστές.

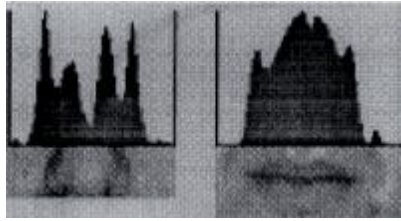
Κατηγορία	Χαρακτηριστικά
Συνολικά	Ύψος, περίγραμμα, εμβαδόν, κέντρο βάρους
Μάτια	Σχήμα, χρώμα, εμβαδόν, άνοιγμα, κέντρο, φωτεινότητα γύρω από την ίριδα, D(άνω βλέφαρο, κάτω βλέφαρο), D(εσωτερική γωνία ματιού, εξωτερική γωνία ματιού), D(κέντρο αριστερού ματιού, κέντρο δεξιού ματιού), D(εξωτερική γωνία αριστερού ματιού, εξωτερική γωνία δεξιού ματιού)
Βλέφαρα	Εμβαδόν, απόσταση από την καμπύλη του ματιού
Φρύδια	Πάχος, D(μάτι, φρύδι)
Χείλη	Πάχος, σχήμα, πρότυπο, πλάτος
Στόμα	Πρότυπο, πλάτος, μήκος, εμβαδόν, άνοιγμα
Μύτη	Πρότυπο, σχήμα, πλάτος, μήκος, εμβαδόν ρουθουνιών
Αυτιά	Μήκος, σχήμα, εμβαδόν
Μαλλιά	Χρώμα, υφή, περιοχή κάλυψης προσώπου, μήκος
Μάγουλα	Φωτεινότητα, υφή
Αποστάσεις	Πλάτος προσώπου στο ύψος των ματιών, D(πηγούνι, γραμμή ματιών), D(πηγούνι, κέντρο χελιών), D(μάτια, γραμμή μαλλιών), D(μάτια, κέντρο μύτης), D(κορυφή πηγουνιού, κέντρο προσώπου), D(αριστερή ακμή προσώπου, κέντρο μύτης), D(δεξιά ακμή προσώπου, κέντρο μύτης), D(εσωτερική κόχη ματιού, κέντρο προσώπου), D(κέντρο ματιού, κέντρο φρυδιού), D(κέντρο προσώπου, κέντρο φρυδιών), D(εσωτερική κόχη ματιού, κόχη στόματος), D(κέντρου ματιού, κέντρο στόματος), D(κορυφή μύτης, κέντρο στόματος)
Εμβαδά	A(εσωτερικές κόχες ματιών, κέντρο στόματος), A(κέντρα φρυδιών, κέντρο στόματος), A(κέντρα ματιών, κέντρο προσώπου)
Λόγοι	D(κέντρο προσώπου, άνοιγμα στόματος)/ D(πηγούνι, άνοιγμα στόματος) D(πηγούνι, άνοιγμα στόματος)/ D(κορυφή μύτης, άνοιγμα στόματος)
Άλλα	Περίγραμμα πηγουνιού, περίγραμμα μετώπου, μέγεθος μετώπου

Πίνακας Α4.2: Χαρακτηριστικά που χρησιμοποιήθηκαν για τη διαδικασία σε μελέτες αναγνώρισης προσώπων. $D(x,y)$ = Ευκλείδεια απόσταση των χαρακτηριστικών x και y

(β) Στατιστικά Χαρακτηριστικά

Τέτοια χαρακτηριστικά είναι το ιστόγραμμα ολόκληρης της εικόνας ή περιοχών αυτής και οι ολοκληρωματικές προβολές εικόνων ακμών στον κάθετο και οριζόντιο άξονα. Έστω για παράδειγμα η υποεικόνα $I(x,y)$ ορισμένη στη περιοχή $[x_1, x_2] \times [y_1, y_2]$. Οι ολοκληρωματικές προβολές κατά τον κάθετο και οριζόντιο άξονα δίνονται από τις σχέσεις:

$$V(x) = \sum_{y=y_1}^{y_2} I(x, y), \quad H(y) = \sum_{x=x_1}^{x_2} I(x, y) \quad (A4.1)$$



Σχήμα A4.2: Ολοκληρωματικές προβολές

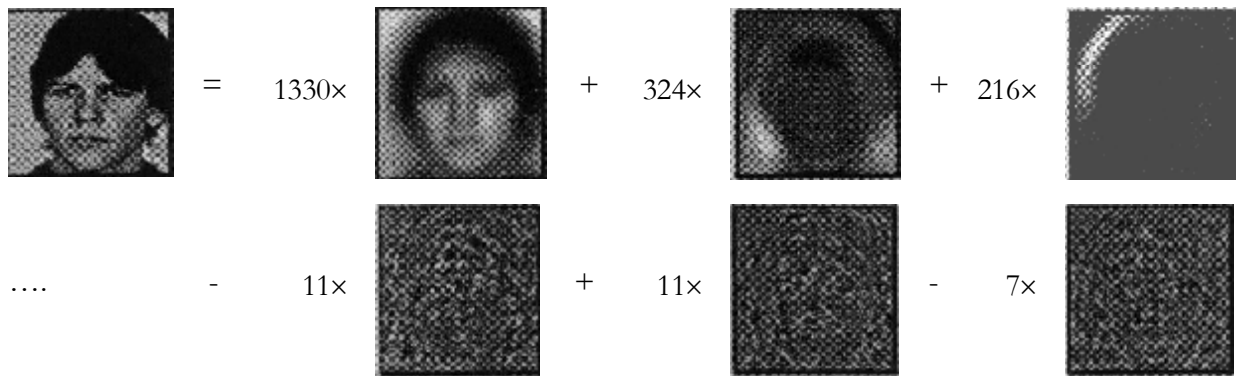
Εξαγωγή στατιστικών χαρακτηριστικών από περιοχές του προσώπου γίνεται επίσης με χρήση κεντρικών ροπών (central moments) και moment invariants. Τα moment invariants είναι γραμμικοί συνδυασμοί κανονικοποιημένων κεντρικών ροπών και έχουν το σημαντικό χαρακτηριστικό της αναισθησίας ως προς τη θέση, προσανατολισμό και αλλαγή κλίμακας [8].

(γ) Χαρακτηριστικά συντελεστών μετασχηματισμού

Η χρήση μετασχηματισμών όπως Fourier, Hadamard για περιγραφή ορίων περιοχών είναι αρκετά διαδεδομένη. Οι περιγραφείς Fourier και οι υπογραφές (signatures) είναι επίσης δύο τεχνικές που χρησιμοποιούνται ευρέως. Οι συντελεστές των μετασχηματισμών λαμβάνονται ως χαρακτηριστικά περιγραφής των επιμέρους περιοχών της εικόνας.

(δ) Αλγεβρικά Χαρακτηριστικά

Τα αλγεβρικά χαρακτηριστικά αναπαριστούν εσωτερικές ιδιότητες της εικόνας η οποία θεωρείται διδιάστατος πίνακας. Η τεχνική εξαγωγής τέτοιων χαρακτηριστικών βασίζεται στην αποσύνθεση πινάκων. Οι πιο διαδεδομένες τεχνικές είναι ο μετασχηματισμός Karhunen Loeve -KL και η SVD -Singular Value Decomposition. Στο μετασχηματισμό KL κάθε εικόνα αναπτύσσεται με βάση τα ιδιοδιανύσματα του πίνακα συμμεταβλητότητας ο οποίος εκτιμάται πάνω σε ένα σύνολο από εικόνες προσώπων –βλέπε Σχήμα A4.3. Οι συντελεστές της ανάπτυξης αυτής αποτελούν τα αλγεβρικά χαρακτηριστικά με βάση τα οποία αναπαρίσταται η εικόνα. Περισσότερα για τον μετασχηματισμό KL δίνονται στο Κεφάλαιο Γ1. Η τεχνική SVD αναλύεται στα Κεφάλαια Γ4 και Γ5.



Σχήμα Α4.3: Αναπαράσταση προσώπου ως άθροισμα ιδιοδιανυσμάτων

2.2 Αναπαράσταση προσώπων με χρήση των τιμών της κλίμακας του γκρι

Η απλούστερη αναπαράσταση της φωτογραφίας ενός προσώπου είναι ο πίνακας με τις τιμές των στοιχείων της. Προφανώς μια τέτοια αναπαράσταση δεν είναι ιδιαίτερα συμπαγής αλλά είναι επιθυμητή όταν ο βασικός στόχος είναι η ευρωστία του συστήματος. Σε συστήματα τα οποία χρησιμοποιούν άλλα χαρακτηριστικά για τη περιγραφή των προσώπων, αποθηκεύεται επίσης ο πίνακας με τις gray scale τιμές, έστω και αν δεν χρησιμοποιείται. Σε μεγάλες βάσεις δεδομένων με πολλές εικόνες, η αναπαράσταση με gray scale τιμές δεν είναι αποδοτική. Παρόλα αυτά, με δεδομένο ότι αναπαράσταση σε ανάλυση 32x32 και κβαντισμό 4 bpp είναι ικανοποιητική τόσο για την ανίχνευση όσο και την ταυτοποίηση, πολλές φορές οι πίνακες με τις gray scale τιμές υποδειγματοληπτούνται στη παραπάνω ανάλυση, επιτρέποντας με τον τρόπο αυτό αποδοτικότερη εφαρμογή των αλγορίθμων.

Ο ρόλος της χωρικής ανάλυσης σε διαδικασίες ταύτισης με βάση τις gray scale τιμές

Ο Πίνακας Α4.3 δείχνει πειραματικά αποτελέσματα για τη διαδικασία αναγνώρισης με χρήση διαφόρων αναλύσεων και πέντε διαφορετικών μετρικών. Συγκεκριμένα έστω $R1$ και $R2$ δύο διδιάστατοι πίνακες διάστασης $k \times l$ που αντιστοιχούν στις gray scale τιμές δύο εικόνων. Ορίζουμε της παρακάτω μετρικές:

$$M1 = \max \left\{ SVD \left\{ \left| \frac{R1}{\text{mean}(R1)} - \frac{R2}{\text{mean}(R2)} \right| \right\} \right\} \Rightarrow \text{μέγιστη ιδιοτιμή του πίνακα των απόλυτων διαφορών μεταξύ } R1 \text{ και } R2$$

$$M2 = \sqrt{\sum \text{diag} \left\{ \left| \frac{R1}{\text{mean}(R1)} - \frac{R2}{\text{mean}(R2)} \right|^T \cdot \left| \frac{R1}{\text{mean}(R1)} - \frac{R2}{\text{mean}(R2)} \right| \right\}} \Rightarrow \text{η Frobenius νόρμα του πίνακα των απόλυτων διαφορών μεταξύ } R1 \text{ και } R2 \text{ οι οποίοι έχουν διαιρεθεί με τη μέση τιμή τους.}$$

$$M3 = \frac{1}{k \cdot l} \sum \left| \frac{R1}{\text{mean}(R1)} - \frac{R2}{\text{mean}(R2)} \right| \Rightarrow \text{άθροισμα των απόλυτων διαφορών -pixel προς pixel- των πινάκων } R1 \text{ και } R2$$

$M4 = \sqrt{\sum \text{diag}\{|R1-R2|^T \cdot |R1-R2|\}} \Rightarrow$ η Frobenius νόρμα του πίνακα των απόλυτων διαφορών μεταξύ R1 και R2

$M5 = \text{corr}(R1, R2) \Rightarrow$ η συσχέτιση των πινάκων R1 και R2

Η βάση δεδομένων που χρησιμοποιήθηκε είναι της Olivetti Research Laboratory (ORL) η οποία είναι μια από τις πλέον χρησιμοποιούμενες και συχνά χρησιμοποιείται ως βάση σύγκρισης διαφορετικών αλγορίθμων. Οι συγκρίσεις πραγματοποιήθηκαν με βάση τους πίνακες των gray scale τιμών και σε όλες τις περιπτώσεις ο κβαντισμός ήταν 8 bpp. Επιτυχής θεωρείται η αναγνώριση αν η προς έλεγχο φωτογραφία και η πλησιέστερη προς αυτήν αποθηκευμένη ανήκουν στο ίδιο πρόσωπο. Από τα αποτελέσματα προκύπτουν μερικά σημαντικά συμπεράσματα:

- (1) Ακόμα και σε ανάλυση 28x23 η αναγνώριση είναι αποτελεσματική. Σημειώνεται πολύ μικρή μείωση στο βαθμό επιτυχίας του συστήματος -σε σχέση με τη μέγιστη ανάλυση.
- (2) Η μέγιστη ανάλυση δεν δίνει υποχρεωτικά και τα καλύτερα αποτελέσματα σε όλες τις μετริกές. Αυτό είναι σύμφωνο και προς την ολοκληρωτική -και όχι διαφορική- συμπεριφορά του ματιού.
- (3) Ο Μέσος Χρόνος Αναγνώρισης (MXA) είναι 1.51, 3.81 και 19.82 sec αντίστοιχα για τις τρεις διαφορετικές αναλύσεις και για τις πέντε μετริกές συνολικά και αναφέρεται σε βάση αποτελούμενη από 200 φωτογραφίες.
- (4) Σχολιασμός της αποδοτικότητας των μετρικών γίνεται σε επόμενη παράγραφο.

Μετริกή	M1	M2	M3	M4	M5
Αποτυχημένες Αναγνώρισεις	6	7	5	7	7
MXA (sec)	1.51				
Ανάλυση	28x23				
Μετρική	M1	M2	M3	M4	M5
Αποτυχημένες Αναγνώρισεις	4	5	3	5	7
MXA (sec)	3.81				
Ανάλυση	56x46				
Μετρική	M1	M2	M3	M4	M5
Αποτυχημένες Αναγνώρισεις	5	6	3	5	5
MXA (sec)	19.82				
Ανάλυση	112x92				

Πίνακας Α4.3: Αναγνώριση προσώπων σε διάφορες αναλύσεις και με χρήση 5 διαφορετικών μετρικών

Μετρική	M1	M2	M3	M4	M5
Αποτυχημένες Αναγνώρισεις	5	6	3	4	7
Ανάλυση	28x23				
Μετρική	M1	M2	M3	M4	M5
Αποτυχημένες Αναγνώρισεις	3	3	2	5	5
Ανάλυση	56x46				
Μετρική	M1	M2	M3	M4	M5
Αποτυχημένες Αναγνώρισεις	3	3	1	4	5
Ανάλυση	112x92				

Πίνακας Α4.4: Αναγνώριση προσώπων σε διάφορες αναλύσεις και με χρήση 5 διαφορετικών μετρητών. Επιτυχής θεωρείται η αναγνώριση αν η προς έλεγχο φωτογραφία και κάποια από τις τρεις πλησιέστερες προς αυτήν αποθηκευμένες ανήκουν στο ίδιο πρόσωπο.

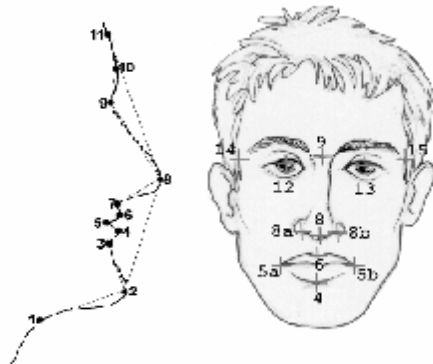
Μετρική	M1	M2	M3	M4	M5
Αποτυχημένες Αναγνώρισεις	3	4	1	3	3
Ανάλυση	28x23				
Μετρική	M1	M2	M3	M4	M5
Αποτυχημένες Αναγνώρισεις	1	3	1	3	2
Ανάλυση	56x46				
Μετρική	M1	M2	M3	M4	M5
Αποτυχημένες Αναγνώρισεις	1	2	1	2	2
Ανάλυση	112x92				

Πίνακας Α4.5: Αναγνώριση προσώπων σε διάφορες αναλύσεις και με χρήση 5 διαφορετικών μετρητών. Επιτυχής θεωρείται η αναγνώριση αν η προς έλεγχο φωτογραφία και κάποια από τις πέντε πλησιέστερες προς αυτήν αποθηκευμένες ανήκουν στο ίδιο πρόσωπο.

Ο Πίνακας Α4.4 δείχνει αποτελέσματα από το προηγούμενο πείραμα, μόνο που τώρα επιτυχημένη θεωρείται η αναγνώριση, αν η προς έλεγχο φωτογραφία και κάποια από τις τρεις πλησιέστερες προς αυτήν αποθηκευμένες ανήκουν στο ίδιο πρόσωπο. Αντίστοιχα, ο Πίνακας Α4.5 δείχνει αποτελέσματα στα οποία επιτυχημένη θεωρείται η αναγνώριση, αν η προς έλεγχο φωτογραφία και κάποια από τις πέντε πλησιέστερες προς αυτήν αποθηκευμένες, ανήκουν στο ίδιο πρόσωπο. Τα αποτελέσματα των Πινάκων Α4.4 και Α4.5 δείχνουν την ευρωστία της αναγνώρισης για τις διάφορες μετρήσιμες

2.3 Γεωμετρικά χαρακτηριστικά από εικόνες εμπρόσθιας όψης και προφίλ

Η αναπαράσταση προσώπων με κάποιο διάνυσμα γεωμετρικών χαρακτηριστικών είναι η πιο διαδεδομένη πρακτική όταν χρησιμοποιούνται εικόνες προφίλ. Αντίθετα σε εικόνες εμπρόσθιας όψης η αναπαράσταση με τις τιμές της κλίμακας του γκρι είναι πιο διαδεδομένη. Πάντως η χρήση γεωμετρικών χαρακτηριστικών για την αναπαράσταση προσώπων έχει ανακτήσει νέο ενδιαφέρον με την ανάπτυξη του προτύπου MPEG-4. Σε εικόνες προφίλ οι τιμές των χαρακτηριστικών παράγονται με βάση κάποια σημεία στη οριογραμμή του, όπως η εγκοπή ανάμεσα στα φρύδια και τη μύτη, η κορυφή της μύτης, η εγκοπή ανάμεσα στη μύτη και το άνω χείλος και η κορυφή του πηγουνιού. Τα χαρακτηριστικά είναι συνήθως αποστάσεις και γωνίες ανάμεσα στα χαρακτηριστικά αυτά σημεία. Ο Πίνακας Α4.6 παρουσιάζει ένα σύνολο από συχνά χρησιμοποιούμενα γεωμετρικά χαρακτηριστικά και για τις δύο κατηγορίες εικόνων.



Εικόνα Α4.1: Σημεία σε προφίλ και frontal views.

Χαρακτηριστικά σε εικόνες profile	Περιγραφή	Χαρακτηριστικά σε εικόνες frontal view	Περιγραφή
Γωνία 1-2-3	Μέτρηση πηγουνιού	Λόγος $d(14,15) / d(4,9)$	Μέτρηση πλάτους προσώπου
Γωνία 7-8-9	Μέτρηση μύτης	Λόγος $d(12,13) / d(14,15)$	Απόσταση ματιών
Λόγος $d(8,10) / d(2,8)$	Μακρύ – κοντό πηγούνι	Λόγος $d(5a, 5b) / d(14,15)$	Μέγεθος στόματος

Πίνακας Α4.6: Χαρακτηριστικά για την περιγραφή του προσώπου (βλέπε Εικόνα Α4.1). $d(i,j)$ είναι η Ευκλείδεια απόσταση ανάμεσα στα σημεία i και j .

2.4 Υβριδική αναπαράσταση

Συχνά για την αναπαράσταση προσώπων χρησιμοποιείται συνδυασμός των gray scale τιμών και αλλά χαρακτηριστικών όπως αυτά των Πινάκων Α4.2 και Α4.6. Για παράδειγμα ο Campell [4] χρησιμοποίησε χαρακτηριστικά όπως η τιμή φωτεινότητας των μαλλιών και των παρειών σε συνδυασμό με ένα πίνακα 32x32 με τις gray scale τιμές της περιοχής των ματιών.

Στις πρώτες προσπάθειες για αναγνώριση προσώπων, οι οποίες κατά βάση στηρίζονταν σε εικόνες προφίλ από πορτρέτα, χρησιμοποιήθηκαν ευρέως τεχνικές τμηματικής προσέγγισης καμπυλών. Συγκεκριμένα ο Galton [5][6] δημιούργησε ένα σύστημα, στο οποίο πορτρέτα προφίλ περιγράφονται με χρήση ενός μαθηματικού τύπου. Κάθε τύπος αποτελείται από τέσσερις ομάδες εικόνων οι οποίες περιέχουν πέντε εικόνες η καθεμιά. Τα τμήματα των προφίλ κοντά στα χαρακτηριστικά σημεία, ή οι συνδέσεις των σημείων αυτών, περιγράφονται με τη χρήση απλών αριθμών. Οι αριθμοί αυτοί υπολογίζονται από κάποιο πίνακα στον οποίο διαφορετικά σχήματα καμπυλών περιγράφονται από συγκεκριμένους αριθμούς (δείκτες).

3. ΕΝΤΟΠΙΣΜΟΣ - ΚΑΤΑΤΜΗΣΗ ΠΡΟΣΩΠΟΥ

Το πρώτο βήμα στη διαδικασία αναγνώρισης είναι ο εντοπισμός της επακριβούς θέσης του προσώπου στη φωτογραφία. Σε πολλές, παρόλα αυτά εργασίες, που πραγματεύονται την ταυτοποίηση προσώπων η θέση του προσώπου θεωρείται εκ των προτέρων γνωστή, ή μπορεί με ελάχιστη προσπάθεια να εντοπιστεί με βάση τη διαδικασία σύλληψης των φωτογραφιών. Τέτοια παραδείγματα εικόνων αποτελούν οι φωτογραφίες οι οποίες λαμβάνονται από τις αστυνομικές αρχές – mug shots.

Ο εντοπισμός προσώπων έχει προσεγγιστεί με δύο διαφορετικές τακτικές. Στη πρώτη προσέγγιση το πρόσωπο θεωρείται ως μια αυτόνομη μονάδα η οποία μοντελοποιείται με τεχνικές υπολογιστικής όρασης. Στις εργασίες των Govindaraju et al. [7][8], για ανίχνευση προσώπων σε φωτογραφίες από εφημερίδες, το πρόσωπο μοντελοποιείται από δύο ευθείες, που αντιπροσωπεύουν τις πλευρές του προσώπου, και δύο τόξα, για το πηγούνι και το άνω μέρος του κεφαλιού. Για κάθε ένα από τα τέσσερα αυτά τμήματα ορίζονται τέσσερις παράμετροι: το μήκος του τμήματος, η χορδή που σχηματίζεται από τα άκρα του, η επιφάνεια ανάμεσα στη χορδή και το καμπύλο τμήμα, και το κέντρο βάρους της επιφάνειας αυτής. Το μέγεθος των προσώπων δεν είναι γνωστό αλλά εκτιμάται ευριστικά με πληροφορία που λαμβάνεται και από τη λεζάντα της φωτογραφίας. Για παράδειγμα αν η λεζάντα αναφέρει ότι υπάρχουν τρία πρόσωπα στην εικόνα, περιορισμοί μπορούν να τεθούν όσον αφορά το μέγιστο μέγεθος του προσώπου, δεδομένου του μεγέθους της φωτογραφίας. Η πληροφορία της λεζάντας αναπαρίσταται με ένα σημασιολογικό (semantic) δίκτυο στο πλαίσιο μιας συλλογιστικής διαδικασίας. Η ανίχνευση των γραμμών και των τόξων πραγματοποιείται με μια παραλλαγή του μετασχηματισμού Hough. Για την διερεύνηση του κατά πόσο υπάρχει πρόσωπο στην εικόνα το μοντέλο (οι δύο γραμμές και τα δύο τόξα και η χωροταξική συσχέτιση τους) συγκρίνεται με τις γραμμές και τα τόξα της προς εξέταση εικόνας. Το κέντρο βάρους το τεσσάρων τμημάτων λαμβάνεται ως το κέντρο του προσώπου. Οι περιορισμοί που τίθενται είναι: (α) το πρόσωπο κοιτάζει προς τη κάμερα (β) δεν υπάρχει ούτε κλίση –βλέπε Σχήμα A4.1, στροφή περί τον άξονα z - ούτε περιστροφή - στροφή περί τον άξονα y . Μειονεκτήματα της μεθόδου είναι η αδυναμία εντοπισμού προσώπων με μικρό μέγεθος, λόγω αποτυχίας αποτελεσματικής ανίχνευσης των ακμών της εικόνας. Η εφαρμογή της τεχνικής πραγματοποιήθηκε σε 10 εικόνες και παρόλο που δεν απέτυχε σε καμία να βρει το πρόσωπο έδωσε αρκετά false alarms (εντόπισε πρόσωπα εκεί που δεν υπήρχαν).

Μια από τις πρώτες εργασίες οι οποίες ασχολήθηκαν με τη διερεύνηση αν σε κάποια φωτογραφία υπάρχει πρόσωπο είναι η [9]. Στην εργασία αυτή υπολογίζεται η εικόνα ακμών από την αρχική φωτογραφία και στη συνέχεια συγκρίνεται με ένα πρότυπο οβάλ αντικείμενο το οποίο μπορεί να μεταβάλλεται σε μέγεθος και

θέση. Σε θέσεις στις οποίες ανιχνεύεται πιθανή ύπαρξη προσώπου αυτή επιβεβαιώνεται με ανίχνευση στις προσδοκώμενες θέσεις ακμών που είναι γνωστό ότι αντιστοιχούν σε χαρακτηριστικά του προσώπου όπως μάτια, στόμα κλπ.

Στην εργασία [10] οι αρχικές φωτογραφίες φιλτράρονται και στη συνέχεια δημιουργούνται οι εικόνες ακμών. Σε αυτές ανιχνεύονται οι ακμές, οι οποίες ενδεχομένως αντιστοιχούν στο περιγράμμα του προσώπου, και στη συνέχεια προβάλλονται ξανά στην αρχική εικόνα όπου και πραγματοποιείται λεπτομερέστερη επεξεργασία για τον επακριβή εντοπισμό τους. Διάφορα ευριστικά τεχνάσματα χρησιμοποιούνται για τη συνένωση των ακμών. Μετά τον πλήρη εντοπισμό του περιγράμματος του προσώπου πραγματοποιείται ο εντοπισμός των επιμέρους χαρακτηριστικών στις αναμενόμενες θέσεις με την εφαρμογή και πάλι ευριστικών τεχνικών.

Στην εργασία του Craw [11] η ανίχνευση του προσώπου πραγματοποιείται με την τεχνική ταύτισης με πρότυπο -template matching. Ο εντοπισμός του προσώπου με την τεχνική αυτή έχει αποδειχτεί πειραματικά πολύ αποτελεσματικός, ιδιαίτερα αν το πρόσωπο κοιτάζει κατευθείαν στη κάμερα και έχει μικρές γωνίες κλίσης και περιστροφής. Αν δεν υπάρχει κάποια πληροφόρηση σχετικά με το μέγεθος του προσώπου, η ανίχνευση πρέπει να πραγματοποιηθεί χρησιμοποιώντας πρότυπα σε διαφορετικές αναλύσεις, κάτι που επιβαρύνει σε μεγάλο βαθμό την υπολογιστική πολυπλοκότητα του αλγορίθμου. Η τεχνική ταύτισης με πρότυπο είναι επιρρεπής σε σφάλματα στην ανίχνευση προσώπων που φοράνε γυαλιά ή έχουν γενειάδα. Για την αποφυγή τέτοιων προβλημάτων συχνά χρειάζεται η χρήση τροποποιημένων προτύπων που να λαμβάνουν υπόψη τις παραμέτρους αυτές. Τέλος ένα πολύ σημαντικό θέμα είναι η απόφαση κατά πόσο σε μια φωτογραφία υπάρχει πρόσωπο ή όχι. Στη τεχνική ταύτισης με πρότυπο, η θέση του προσώπου επιλέγεται ως η θέση με το καλύτερο ταιρίασμα με το πρότυπο. Το γεγονός αυτό δεν εξασφαλίζει και το ότι στη συγκεκριμένη θέση υπάρχει πρόσωπο. Η εφαρμογή κάποιου κατωφλίου είναι αυθαίρετη και δεν μπορεί να στηριχθεί σε τεκμηριωμένα στοιχεία. Τέλος στη τεχνική ταιριάσματος με πρότυπο σημαντικό στοιχείο είναι η επιλογή του μέτρου ταύτισης. Πειραματικά αποδεικνύεται ότι μια παραλλαγή της μετρικής I_1 δίνει τα καλύτερα αποτελέσματα.

Σε μια νεότερη εργασία του, ο Craw [12] περιγράφει ένα σύστημα για την αναγνώριση και μέτρηση των χαρακτηριστικών του προσώπου. Η εργασία αυτή είχε σχεδιαστεί για δεικτοδότηση (indexing) βάσεων από mug shot φωτογραφίες που χρησιμοποιούσαν οι αστυνομικές αρχές. Ο στόχος ήταν ο εντοπισμός 40 χαρακτηριστικών σημείων σε μια φωτογραφία εμπρόσθιας όψης. Η επιλογή των σημείων αυτών έγινε με βάση την εργασία του Shepherd [13] η οποία και χρησιμοποιήθηκε για την αξιολόγηση των αποτελεσμάτων. Η ανίχνευση του προσώπου πραγματοποιείται με τη χρήση ενός πολυγωνικού προτύπου (polygon template) σε μια διαδικασία coarse to fine. Η πολυγωνική μορφή του προτύπου επιτρέπει το μετασχηματισμό του όσον αφορά τη θέση, κλίμακα, κλίση και περιστροφή. Η θέση του προσώπου βρίσκεται με τεχνικές simulated annealing και διαδοχικούς μετασχηματισμούς του προτύπου. Η γεωμετρία του ανθρωπίνου προσώπου μπορεί να χρησιμοποιηθεί για να τεθούν περιορισμοί στους πιθανούς μετασχηματισμούς. Μετά από ένα χονδρικό εντοπισμό του προσώπου εφαρμόζονται μετασχηματισμοί σε μεμονωμένα πολύγωνα για τον επακριβή εντοπισμό του περιγράμματος. Ο εντοπισμός των χαρακτηριστικών σημείων πραγματοποιείται με βάση μοντέλα των χαρακτηριστικών τα οποία οι

συγγραφείς ονομάζουν feature experts. Σύμφωνα με τα αποτελέσματα της εργασίας το περίγραμμα του προσώπου ανιχνεύθηκε και στις 50 φωτογραφίες, σε 43 πλήρως, ενώ στις υπόλοιπες 7 υπήρξε απώλεια του τμήματος που αντιστοιχεί στο πηγούνι εξαιτίας της ύπαρξης μουστακιού και γενειάδας. Δεν εντοπίστηκαν κυρίως τα χαρακτηριστικά σημεία που αντιστοιχούν στα φρύδια, δεδομένου ότι για αυτά οι συγγραφείς δεν είχαν προβλέψει κάποιο συγκεκριμένο μοντέλο. Με δεδομένη τη χρήση των πολυγωνικών μετασχηματισμών και της simulated annealing βελτιστοποίησης το σύστημα είναι αρκετά πολύπλοκο υπολογιστικά.

Η εργασία [14] είναι από τις λίγες που ασχολήθηκαν με την κατάτμηση του προσώπου σε μη ομοιογενές φόντο. Μετά από επεξεργασία της αρχικής εικόνας για την δημιουργία της εικόνας ακμών, η τεχνική που προτείνεται ομαδοποιεί εκείνες τις ακμές οι οποίες ανήκουν σε μια έλλειψη με την οποία έχει μοντελοποιηθεί το πρόσωπο. Οι παράμετροι της έλλειψης είναι το κέντρο (x_0, y_0) και οι ημιάξονες a, b . Όπως φαίνεται και από το σύνολο των παραμέτρων δεν λαμβάνεται πρόνοια για ανίχνευση προσώπων με κλίση -στροφή ως προς τον z άξονα- ενώ αντίθετα μπορούν ανιχνευθούν πρόσωπα με στροφή περί τον άξονα y εφόσον έχει γίνει αποτελεσματική ανίχνευση των ακμών.

Μια τεχνική κατάτμησης προσώπου παρόμοια με την ταύτιση με πρότυπο αλλά με καλύτερο θεωρητικό υπόβαθρο στηρίζεται στο μετασχηματισμό Karhunen-Loeve. Κάθε υποεικόνα προβάλλεται στο χώρο των ιδιοδιανυσμάτων του ζητούμενου χαρακτηριστικού και οι συντελεστές της προβολής συγκρίνονται με τις ιδιοτιμές του χαρακτηριστικού. Η διαφορά αυτή ορίζεται ως DFFS (Distance-From-Feature-Space) και προφανώς σε κάθε pixel της εικόνας αντιστοιχεί μια τέτοια απόσταση. Το pixel με τη μικρότερη απόσταση προσδιορίζει τη θέση του χαρακτηριστικού (εφόσον βέβαια η απόσταση αυτή είναι μικρότερη από κάποιο σιαλοπάτι, ώστε να είμαστε βέβαιοι, ότι το ζητούμενο χαρακτηριστικό υπάρχει στην υπό εξέταση εικόνα). Συγκεκριμένα έστω ένα σύνολο από N πρότυπα χαρακτηριστικών προσώπου –μάτια, μύτη στόμα κοκ- $\{I_1, I_2, \dots, I_N\}$ όπου $I_i \in R^{n \times m}$. Με λεξικογραφική διάταξη των προτύπων I_i σχηματίζουμε το σύνολο των διανυσμάτων $\{v_1, v_2, \dots, v_N\}$ με $v_i \in R^L$ και $L = n \cdot m$. Οι συναρτήσεις βάσης του μετασχηματισμού KL για το παραπάνω σύνολο των διανυσμάτων προκύπτουν από την επίλυση του προβλήματος ιδιοτιμών:

$$\Lambda = \Phi^T \Sigma \Phi \quad (A4.2)$$

όπου Σ είναι ο πίνακας συμμεταβλητότητας, Φ είναι ο πίνακας των ιδιοδιανυσμάτων του Σ , και Λ είναι ο αντίστοιχος διαγώνιος πίνακας των ιδιοτιμών.

Ένα διάνυσμα περιγραφής y_i κάθε υποεικόνας x_i -απεικονιζόμενης μέσω του αντίστοιχου διανύσματος λεξικογραφικής διάταξης- προκύπτει από τη μερική προβολή της στους άξονες του μετασχηματισμού KL:

$$y_i = \Phi_k^T \cdot \hat{x}_i \quad (A4.3)$$

όπου $\hat{x}_i = x_i - \mu$, $\mu = \sum_{j=1}^N v_j$ και Φ_k είναι ο πίνακας με τα ιδιοδιανύσματα στήλες που αντιστοιχούν στις k μεγαλύτερες ιδιοτιμές του Σ .

Ορίζοντας το σφάλμα ανακατασκευής:

$$e_{KLT}(k, i) = (\hat{x}_i - \bar{x}_i)^T \cdot (\hat{x}_i - \bar{x}_i) \quad (A4.4)$$

$$\text{όπου } \bar{x}_i = \Phi_k \cdot y_i, \quad y_i = \Phi_k^T \cdot \hat{x}_i$$

η υποεικόνα με την ελάχιστη DFPS δίνεται από τη σχέση:

$$x_{opt} = \arg \min_i [e_{KLT}(k, i)] \quad (A4.5)$$

Τεχνικές ανίχνευσης του προσώπου ως αυτόνομης μονάδας με χρήση νευρωνικών δικτύων έχουν επίσης αναπτυχθεί [16]. Η ανίχνευση προσώπων με χρήση νευρωνικών δικτύων μοιάζει ίσως περισσότερο από κάθε άλλη τεχνική με τη λειτουργία του ανθρώπινου εγκεφάλου. Δυστυχώς όμως παρουσιάζει όλα τα μειονεκτήματα των νευρωνικών δικτύων όπως η δυσκολία μάθησης και γενίκευσης. Επιπλέον υπάρχει σχετική δυσκολία υλοποίησης των αλγορίθμων που περιγράφονται στις διάφορες εργασίες.

Η δεύτερη προσέγγιση στην ανίχνευση προσώπου απαιτεί την ανίχνευση πρώτα κάποιων επιμέρους χαρακτηριστικών του προσώπου και στη συνέχεια με βάση τη χωροταξική τους διάταξη τον εντοπισμό όλου του προσώπου. Παρόλο που οποιοδήποτε από τα χαρακτηριστικά του προσώπου μπορεί να εντοπιστεί πρώτα τυπική επιλογή είναι τα μάτια.

Στην εργασία του Yulie [17] δίνεται μια τεχνική εντοπισμού των ματιών με χρήση deformable templates (παραμορφώσιμων προτύπων). Μια συνάρτηση ενέργειας με παραμέτρους τις ακμές, τις κορυφές και τις κοιλάδες της gray scale εικόνας χρησιμοποιείται για την ταύτιση. Το πρότυπο εφαρμόζεται στα επιμέρους τμήματα της εικόνας και οι παράμετροί του μεταβάλλονται για να επιτύχουν την ελαχιστοποίηση της συνάρτησης ενέργειας, δηλαδή τη βέλτιστη ταύτιση. Η ελαστικότητα του προτύπου επιτρέπει αλλαγές στην κλίμακα και στην οριοθέτηση των ματιών. Με τον τρόπο αυτό μπορούν να εντοπιστούν τα μάτια ανεξάρτητα από μεταβολές στη κλίμακα, κλίση και στις συνθήκες φωτεινότητας. Εκτός από τα μάτια, με την ίδια τεχνική εντοπίζονται και τα χείλη. Το πρότυπο για τα μάτια αποτελείται από ένα κύκλιο ο οποίος αντιστοιχεί στη κόρη, δύο παραβολικά τμήματα, που αντιστοιχούν στις οριογραμμές ματιών και βλεφάρων, και δύο σημεία τα οποία αντιστοιχούν στα κέντρα των δύο λευκών περιοχών του ματιού. Η τεχνική αυτή είναι αρκετά αποτελεσματική αλλά αντιμετωπίζει δύο βασικά προβλήματα: οι συντελεστές της συνάρτησης ενέργειας υπολογίζονται ευριστικά, με αποτέλεσμα το πρότυπο να αντιμετωπίζει προβλήματα γενίκευσης και ο υπολογιστικός φόρτος είναι ιδιαίτερα υψηλός. Συγκεκριμένα η ανίχνευση των ματιών απαιτεί χρόνους της τάξης των πέντε λεπτών.

Ο Nixon στην εργασία του [18] προτείνει μια τεχνική στην οποία ανιχνεύει τα μάτια χρησιμοποιώντας το μετασχηματισμό Hough. Η ίριδα μοντελοποιείται ως κύκλος και η οριογραμμή ματιών και βλεφάρων προσεγγίζεται με τη χρήση μιας τροποποιημένης εκθετικής συνάρτησης. Στην εργασία του Craw [11] τα κέντρα των ματιών εντοπίζονται ως ένα ζεύγος από ελάχιστα της φωτεινότητας της εικόνας, με αναζήτηση τους κάτω από τα φρύδια. Είναι προφανές ότι της ανίχνευσης των ματιών προηγείται η ανίχνευση των φρυδιών.

Ως τελικό συμπέρασμα αναφέρεται ότι η ανίχνευση του προσώπου με χρήση πρώτα των επιμέρους χαρακτηριστικών μπορεί να εφαρμοστεί με επιτυχία μόνο σε περιπτώσεις στις οποίες το πρόσωπο στη φωτογραφία βρίσκεται σε αρκετά μεγάλη ανάλυση -ώστε να υπάρχει επαρκής πληροφορία για την μοντελοποίηση των επιμέρους χαρακτηριστικών- και τα χαρακτηριστικά δεν έχουν επικαλύψεις. Από την άλλη πλευρά, στις περιπτώσεις όπου ο εξαιρετικά ακριβής εντοπισμός του προσώπου είναι πολύ σημαντικός, η χρήση των επιμέρους χαρακτηριστικών ίσως είναι αναπόφευκτη. Ένα επιπλέον στοιχείο είναι η εκ των προτέρων γνώση που έχουμε για τις θέσεις των χαρακτηριστικών στο πρόσωπο. Αν για κάποιο λόγο η ανίχνευση ενός χαρακτηριστικού αποτύχει, αυτό μπορεί να γίνει αντιληπτό μετά την ανίχνευση κάποιου άλλου, εφόσον η γεωμετρική διάταξη των δύο χαρακτηριστικών δεν συμφωνεί με την ανατομία του προσώπου.

4. ΤΑΥΤΟΠΟΙΗΣΗ ΠΡΟΣΩΠΩΝ

Η ταυτοποίηση είναι το επόμενο και πιο σημαντικό στάδιο στην αναγνώριση προσώπου. Υπάρχουν δύο απόψεις σχετικές με την ταυτοποίηση: (α) Το πρόσωπο πρέπει να συσχετιστεί με κάποιο όνομα και (β) πολλές εμφανίσεις του ιδίου προσώπου πιθανόν με διαφορετικές αναλύσεις, συνθήκες φωτισμού και οριοθέτησης πρέπει να αναγνωρίζονται ως μία. Η πρώτη άποψη αναφέρεται περισσότερο στην λογική των ανθρώπων και για την υλοποίηση της σε περιβάλλον υπολογιστή χρειάζεται πρώτα η υιοθέτηση της δεύτερης άποψης. Ο πρώτος που ασχολήθηκε εκτεταμένα με τη διαδικασία ταυτοποίησης προσώπων ήταν ο Galton [5][6] στο τέλος του 19^{ου} αιώνα. Με βάση τα δεδομένα εισόδου -διαθέσιμες φωτογραφίες- η ταυτοποίηση γίνεται με χρήση εικόνων προφίλ ή με φωτογραφίες εμπρόσθιας όψης.

Στα πρώτα χρόνια της ερευνητικής δραστηριότητας για την ταυτοποίηση προσώπων οι περισσότερες τεχνικές χρησιμοποιούσαν εικόνες προφίλ. Αντίθετα στην σύγχρονη βιβλιογραφία αναφέρονται τεχνικές που χρησιμοποιούν εικόνες εμπρόσθιας όψης. Παρόλο που η αφετηρία των δύο προσεγγίσεων διαφέρει, οι αρχές στις οποίες στηρίζονται είναι παρόμοιες. Μια τυπική ακολουθία βημάτων που ακολουθούνται στην ταυτοποίηση προσώπων και στις δύο προσεγγίσεις δίνεται στη συνέχεια:

- (1) Προσδιόρισε ένα σύνολο από ανεξάρτητα χαρακτηριστικά για την αναπαράσταση του προσώπου.
- (2) Αναπαράστησε όλα τα γνωστά πρόσωπα με βάση τα χαρακτηριστικά του βήματος 1 και αποθήκευσέ τα σε μία βάση δεδομένων
- (3) Υπολόγισε τα χαρακτηριστικά του προς ταυτοποίηση προσώπου
- (4) Χρησιμοποίησε μια διαδικασία ταύτισης συνδυασμένη με κατάλληλη μετρική για να βρεις την βέλτιστη ταύτιση με τα γνωστά πρόσωπα.

Η διαφοροποίηση των τεχνικών που έχουν προταθεί για την ταυτοποίηση προσώπου έγκειται στην επιλογή των χαρακτηριστικών και της διαδικασίας ταύτισης. Στις περισσότερες περιπτώσεις ο αριθμός των γνωστών προσώπων στη βάση δεδομένων είναι σχετικά μικρός -λιγότερος από 100 πρόσωπα (βλέπε Πίνακα Α4.1). Ο αριθμός των χαρακτηριστικών ποικίλει από 10-40.

Η πρώτη εργασία στην ταυτοποίηση προσώπων από τον Galton [5][6] στηρίχθηκε σε εικόνες προφίλ. Ένα σύνολο από πέντε πρωτεύοντα σημεία που αντιστοιχούν στην εγκοπή ανάμεσα στη μύτη και τα φρύδια - σημείο 9 στην Εικόνα A4.1, στη κορυφή της μύτης, στην εγκοπή ανάμεσα στη μύτη και το άνω χείλος, στο τμήμα διαχωρισμού των χειλιών, και στη κορυφή του πηγουνιού ορίζονται στην οριογραμμή του προφίλ. Με βάση τα σημεία αυτά παράγονται τα χαρακτηριστικά τα οποία χρησιμοποιούνται για την ταύτιση των προσώπων. Η διαδικασία ταύτισης που ακολούθησε ο Galton ήταν η εξής:

Έστω X τα σύνολα των χαρακτηριστικών που αντιστοιχούν στα αποθηκευμένα στη βάση γνωστά πρόσωπα. Αν Y είναι το σύνολο των χαρακτηριστικών του προς ταυτοποίηση πρόσωπου τότε αναζητείται το σύνολο των χαρακτηριστικών $Z \in X$ το οποίο προσομοιάζει το σύνολο Y με ένα προκαθορισμένο βαθμό ομοιότητας. Η ιδιαιτερότητα στη διαδικασία ταύτισης έγκειται στο γεγονός ότι δύο αντίστοιχα χαρακτηριστικά $y_i \in Y$ και $z_i \in Z$ δεν μπορούν να διαφέρουν περισσότερο από κάποιο προκαθορισμένο όριο. Η τεχνική αυτή ταύτισης είναι συνηθισμένη σε προσεγγίσεις που χρησιμοποιούν γεωμετρικά χαρακτηριστικά τα οποία είναι λιγα στον αριθμό.

Ο Harmon [20] χρησιμοποίησε μια παρόμοια με τον Galton τεχνική για την ταυτοποίηση προσώπων από προφίλ. Χρησιμοποίησε εννέα θεμελιώδη σημεία διατηρώντας τα πέντε σημεία του Galton και άλλα τέσσερα που αντιστοιχούν στο μέτωπο, στις κορυφές των δύο χειλιών και το λαιμό - Εικόνα A4.1. Με βάση τα θεμελιώδη σημεία δημιουργείται το σύνολο των χαρακτηριστικών που περιλαμβάνει αποστάσεις και γωνίες μεταξύ των σημείων, καθώς και το εμβαδόν κάποιων τριγώνων που αυτά σχηματίζουν. Ο βαθμός ομοιότητας μεταξύ των συνόλων από χαρακτηριστικά Y και Z υπολογίζεται με μια παραλλαγή της Ευκλείδειας απόστασης:

$$d^2(Y, Z) = \sum_{i=1}^n \frac{y_i^2 - z_i^2}{\sigma_i^2} \quad (A4.6)$$

όπου n είναι ο αριθμός των χαρακτηριστικών και σ^2 είναι η διασπορά τους. Τα προφίλ από 256 πρόσωπα εισήχθησαν στον υπολογιστή, στη συνέχεια δε τα θεμελιώδη σημεία εντοπίζονται αυτόματα. Για παράδειγμα για τον εντοπισμό των σημείων που αντιστοιχούν στις κορυφές των χειλιών και την εγκοπή του στόματος ακολουθείται η παρακάτω διαδικασία:

- (1) Ξεκίνα από το σημείο εγκοπής της μύτης και του άνω χείλους και προχώρα προς τα κάτω
- (2) Υπολόγισε τη καμπυλότητα κάθε σημείου που ανήκει στην οριογραμμή του προφίλ και
- (3) Επέλεξε τα τρία σημεία που αντιστοιχούν στις μέγιστες καμπυλότητες που έχουν τα κοίλα προς τα μέσα, προς τα έξω και πάλι προς τα μέσα. Τα σημεία αυτά αντιστοιχούν στα ζητούμενα θεμελιώδη σημεία.

Σε μια νεότερη εργασία του [21] προσέθεσε περισσότερα πρόσωπα, μερικά από τα οποία ανήκαν σε γυναίκες. Χρησιμοποιώντας τα ίδια θεμελιώδη σημεία κατασκεύασε ένα 11-Δ διάνυσμα χαρακτηριστικών, αποτελούμενο από την προεξοχή της μύτης, το εμβαδόν και την διακύμανση του προφίλ καθώς και αρκετές αποστάσεις όπως οι D_{12} , D_{14} , D_{15} , D_{16} , D_{18} και D_{79} όπου D_j είναι η απόσταση μεταξύ των θεμελιωδών

σημείων i και j . Για τη διαδικασία ταύτισης χρησιμοποιήθηκε μια τεχνική διαμέρισης συνόλων η οποία στηρίζεται στο γεγονός ότι, αν έστω και ένα χαρακτηριστικό από δύο συγκρινόμενα προφίλ διαφέρει σε μεγάλο βαθμό, τότε τα προφίλ ανήκουν σε διαφορετικά πρόσωπα. Μετά τη διαδικασία διαμέρισης, η ταύτιση επιτυγχάνεται με βάση την μικρότερη Ευκλείδεια απόσταση ανάμεσα στα διανύσματα χαρακτηριστικών που ανήκουν στην ίδια διαμέριση (υποσύνολο). Οι συγγραφείς της εργασίας κατέληξαν στο συμπέρασμα ότι ένα 10-Δ διάνυσμα χαρακτηριστικών είναι επαρκές για εύρωστη και αποτελεσματική ταυτοποίηση. Στη εργασία [22] τα θεμελιώδη σημεία αυξήθηκαν σε έντεκα και προστέθηκαν κάποια επιπλέον χαρακτηριστικά. Επίσης για πρώτη φορά η σύλληψη των εικόνων προφίλ έγινε αυτόματα με χρήση κάμερας και με ομογενές και με μεγάλη αντίθεση (contrast) φόντο.

Οι Kaufman και Breeding [23] σχεδίασαν επίσης ένα σύστημα ταυτοποίησης προσώπων με χρήση προφίλ. Τα χαρακτηριστικά τα οποία χρησιμοποίησαν ήταν συντελεστές υπολογισμένοι από την πολική μορφή της συνάρτησης αυτοσυσχέτισης. Χρησιμοποίησαν επίσης συναρτήσεις moment invariants.

Διάφορες άλλες τεχνικές έχουν επίσης χρησιμοποιηθεί για ταυτοποίηση με βάση τα προφίλ. Μια σημαντική παράμετρος είναι ο τρόπος εισαγωγής των προφίλ στον υπολογιστή. Πολλές προσεγγίσεις χρησιμοποίησαν για το σκοπό αυτό συστήματα τηλεμετρίας - range finders.

Οι σύγχρονες τεχνικές ταυτοποίησης χρησιμοποιούν κυρίως εικόνες εμπρόσθιας όψης, οι οποίες μεταξύ άλλων εισάγονται ευκολότερα στον υπολογιστή. Μια από τις πρώτες τεχνικές της κατηγορίας αυτής προτάθηκε από τον Goldstein [24]. Τα πρόσωπα σχεδιάζονταν από ζωγράφους, και οι τιμές των χαρακτηριστικών υπολογίζονταν από διάφορους κριτές. Αρχικά χρησιμοποιήθηκαν 34 χαρακτηριστικά τα οποία όμως περιορίστηκαν τελικά στα 22, δεδομένου ότι δεν ήταν όλα μεταξύ τους ανεξάρτητα (π.χ. το πάχος του άνω και κάτω χείλους). Κάθε χαρακτηριστικό μπορούσε να πάρει μια τιμή στη κλίμακα 1-5 (χαμηλό – ενδιάμεσες τιμές - υψηλό). Το σύνολο μάθησης αποτελείτο από 255 πρόσωπα. Τα χαρακτηριστικά περιελάμβαναν διάφορες μετρήσεις χαρακτηριστικών προσώπου όπως την απόχρωση και υφή των μαλλιών, τα φρύδια, τα μάτια, τη μύτη, το στόμα, τα αυτιά, τα μάγουλα και το μέτωπο. Μετά την εισαγωγή των δεδομένων το σύστημα υπολόγιζε την Ευκλείδεια απόσταση ανάμεσα στο προς ταυτοποίηση πρόσωπο και όλα τα πρόσωπα της βάσης δεδομένων. Το αποθηκευμένο πρόσωπο με τη μικρότερη απόσταση θεωρείτο το ζητούμενο. Από τα συμπεράσματα των συγγραφέων προέκυψε ότι μόνο 6-7 χαρακτηριστικά είχαν ουσιαστική συμμετοχή στη διαδικασία ταύτισης. Οι Kaya και Kobayashi [25] πρότειναν ένα αντίστοιχο σύστημα, αλλά χρησιμοποίησαν τις αποστάσεις ανάμεσα στα χαρακτηριστικά του προσώπου για την κατασκευή του διανύσματος χαρακτηριστικών.

Ο Buhr [19] χρησιμοποίησε 33 πρωτεύοντα και 12 δευτερεύοντα χαρακτηριστικά για τη διαδικασία της ταυτοποίησης. Ανάμεσα στα πρωτεύοντα χαρακτηριστικά ήταν 21 αποστάσεις -π.χ. αποστάσεις ανάμεσα στα κέντρα των ματιών και του στόματος- 4 διαφορές τεταγμένων, 4 εμβαδά τριγώνων, 2 λόγοι αποστάσεων και δύο ειδικά χαρακτηριστικά που αφορούσαν το εμβαδόν της περιοχής των ματιών. Το σχήμα ταύτισης αποτελείτο από ένα γραμμικό δέντρο αποφάσεων. Ο Wong [26] χρησιμοποίησε μια παρόμοια τεχνική με σχήμα ταύτισης ένα συνδυασμό Ευκλείδειας απόστασης και της διαφοράς του αθροίσματος των

χαρακτηριστικών. Διάφορες άλλες τεχνικές χρησιμοποιούν ταύτιση με βάση τα ελάχιστα τετράγωνα ή την ομαδοποίηση (clustering).

Ο Baron [27] χρησιμοποίησε τις gray scale τιμές της εικόνας ως χαρακτηριστικά. Συγκεκριμένα μετά από την ανίχνευση των ματιών, με χρήση ενός 20x23 πίνακα ταύτισης και εφαρμογή της συσχέτισης, οι εικόνες κανονικοποιούνται, διατηρώντας την απόσταση ανάμεσα στα μάτια σταθερή. Στη συνέχεια, μετά από υποδειγματοληψία, η ανάλυση των εικόνων μειώνεται σε 15x16 pixels. Για τη διαδικασία ταύτισης, η κανονικοποιημένη προς ταυτοποίηση εικόνα συγκρίνεται με κάθε εικόνα της βάσης δεδομένων και οι εικόνες για τις οποίες η τιμή συσχέτισης είναι υψηλή περνάνε στο δεύτερο στάδιο ελέγχου. Στο στάδιο αυτό ελέγχεται η συσχέτιση των επιμέρους χαρακτηριστικών του προσώπου τα οποία είναι αποθηκευμένα σε ξεχωριστούς πίνακες ταύτισης. Ταύτιση θεωρείται ότι επιτυγχάνεται όταν τα τρία τέταρτα των χαρακτηριστικών των δύο προσώπων παρουσιάζουν υψηλό βαθμό συσχέτισης.

Υπάρχουν προσεγγίσεις της ταυτοποίησης προσώπων οι οποίες χρησιμοποιούν εξειδικευμένο λογισμικό. Συγκεκριμένα οι Preston [28] και Taylor [29] χρησιμοποίησαν ένα οπτικό σύστημα αναγνώρισης και ένα δίκτυο φωτοπολλαπλασιαστών αντίστοιχα για το σκοπό αυτό. Βέβαια, ο αριθμός των προσώπων που χρησιμοποιήθηκαν ήταν εξαιρετικά περιορισμένος -6 και 10 πρόσωπα αντίστοιχα.

Συμπερασματικά για τη διαδικασία ταύτισης πρέπει να έχουμε υπόψη μας μερικά βασικά σημεία:

4.1 Επίλογή των χαρακτηριστικών

Συστήματα τα οποία χρησιμοποιούν χαρακτηριστικά και όχι τις gray scale τιμές της εικόνας θα πρέπει να επιλέξουν ένα κατάλληλο και αποτελεσματικό σύνολο από χαρακτηριστικά για την ταυτοποίηση. Έχουν προταθεί πολλά χαρακτηριστικά στις διάφορες ερευνητικές εργασίες. Στους Πίνακες A4.2 και A4.6 παρουσιάζονται αρκετά από αυτά τα οποία διαφέρουν ανάλογα με το αν χρησιμοποιούνται εικόνες εμπρόσθιας όψης ή προφίλ.

4.2 Σχήματα ταύτισης

Πολλές τεχνικές έχουν χρησιμοποιηθεί για τη διαδικασία ταύτισης. Στην πλειοψηφία τους είναι κλασικές τεχνικές ή παραλλαγές τους. Συνοπτικά οι πιο δημοφιλείς φαίνεται ότι είναι οι:

- (1) *Ευκλείδεια απόσταση*. Αποτελεί την πρώτη επιλογή στη διαδικασία ταύτισης. Ο στόχος είναι να ευρεθεί το πρόσωπο εκείνο στη βάση δεδομένων του οποίου το διάνυσμα χαρακτηριστικών παρουσιάζει την ελάχιστη Ευκλείδεια απόσταση με το αντίστοιχο διάνυσμα του προς ταυτοποίηση πρόσωπου. Η χρήση της Ευκλείδειας απόστασης είναι αποτελεσματική ανάλογα με τα χαρακτηριστικά που χρησιμοποιούνται.
- (2) *Ομαδοποίηση*. Τεχνικές ομαδοποίησης χρησιμοποιούνται σχεδόν αποκλειστικά όταν τα πρόσωπα αναπαριστώνται από διανύσματα χαρακτηριστικών. Διακρίνονται για την ευστάθειά τους αλλά όχι για τη τέλεια διαχωριστική τους ικανότητα. Τυπικά παραδείγματα αποτελούν οι αλγόριθμοι k-means και τα νευρωνικά δίκτυα SOM.

- (3) *Διαμερισμός συνόλων.* Εφαρμόστηκε σε πολλές περιπτώσεις σε τεχνικές που χρησιμοποιούν διανύσματα γεωμετρικών χαρακτηριστικών όπως αποστάσεις, γωνίες κλπ. Η διαδικασία στηρίζεται στην σταδιακή απόρριψη διανυσμάτων τα οποία διαφέρουν σημαντικά έστω και σε ένα επιμέρους χαρακτηριστικό. Η τελική επιλογή πραγματοποιείται από ένα πολύ μικρό υποσύνολο των προσώπων της βάσης δεδομένων και συχνά απαιτεί τη χρήση μια άλλης μετρικής όπως η Ευκλείδεια απόσταση.
- (4) *Συσχέτιση.* Δημοφιλής διαδικασία ταύτισης για τεχνικές που χρησιμοποιούν τις gray scale τιμές της εικόνας ως χαρακτηριστικά. Πάντως πειραματικά αποδεικνύεται ότι δεν είναι η βέλτιστη επιλογή για την ταύτιση.
- (5) *Συνδυασμός των ανωτέρω.* Ο συνδυασμός των ανωτέρω τεχνικών ταύτισης τις περισσότερες φορές είναι επιβεβλημένος από την ίδια τη μορφή τους. Δεν αποτελεί έκπληξη επομένως το γεγονός ότι πολλοί ερευνητές ακολούθησαν την τακτική αυτή. Άλλες προσεγγίσεις όπως τα γραμμικά δέντρα απόφασης και τα ελάχιστα τετράγωνα έχουν επίσης χρησιμοποιηθεί αλλά δεν είναι τόσο διαδεδομένες.

Στους Πίνακες Α4.3 - Α4.5 πραγματοποιήσαμε μια σύγκριση πέντε διαφορετικών μετρικών οι οποίες χρησιμοποιούνται για ταύτιση πινάκων. Από τα αποτελέσματα της σύγκριση προκύπτει ότι οι μετρικές *M1* και *M3* όχι μόνο επιτυγχάνουν υψηλότερο βαθμό επιτυχούς ταυτοποίησης αλλά είναι λιγότερο επιρρεπείς στη χωρική ανάλυση των εικόνων. Επίσης παρουσιάζουν σημαντική ευρωστία, δεδομένου ότι ακόμα και οι αποτυχημένες ταυτοποιήσεις δεν είναι ιδιαίτερα απομακρυσμένες από τα σωστά πρότυπα που βρίσκονται αποθηκευμένα στη βάση δεδομένων. Σημειώνεται ότι οι εικόνες που χρησιμοποιήθηκαν στο πείραμα δεν είχαν κανονικοποιηθεί ως προς την απόσταση των ματιών, μια τακτική η οποία είναι σχεδόν επιβεβλημένη στις τεχνικές που χρησιμοποιούν για την ταυτοποίηση τους πίνακες των gray scale τιμών.

ΤΕΧΝΙΚΕΣ ΑΝΑΛΥΣΗΣ ΕΚΦΡΑΣΕΩΝ ΜΕ ΒΑΣΗ ΤΗΝ ΜΟΡΦΟΠΟΙΗΣΗ ΤΟΥ ΠΡΟΣΩΠΟΥ

Οι προσεγγίσεις σχετικά με την ανάλυση εκφράσεων διακρίνονται χονδρικά σε τρεις κατηγορίες ανάλογα με την πηγή της πληροφορίας που χρησιμοποιείται: (α) στατικές – χρήση εικόνων προσώπων που απεικονίζουν κάποια έκφραση, (β) ημιστατικές – χρήση δύο εικόνων μια με το πρόσωπο σε ουδέτερη κατάσταση και μία με το πρόσωπο στη κορύφωση της έκφρασης και (γ) δυναμικές – χρήση ακολουθίας βίντεο στην οποία απεικονίζεται η εξέλιξη της έκφρασης.

Στο κεφάλαιο αυτό πραγματοποιείται μια ανασκόπηση των τεχνικών που έχουν εμφανιστεί στη βιβλιογραφία σχετικά με την ανάλυση εκφράσεων και την εξαγωγή συμπεράσματος σχετικά με τα συναισθήματα με βάση τις διαμορφώσεις των ιστών του προσώπου –οπτική πληροφορία. Σε κάποιες περιπτώσεις δίνονται και μερικές προσομοιώσεις, οι οποίες πραγματοποιήθηκαν από τον συγγραφέα της διατριβής, για βελτίωση της κατανόησης των αλγορίθμων αλλά και για να δοθεί η ευκαιρία να τοιστούν και να σχολιαστούν κάποια σημαντικά ζητήματα.

1. ΕΙΣΑΓΩΓΗ

Η έρευνα στην ανάλυση εκφράσεων έχει έλθει ξανά στο προσκήνιο τα τελευταία χρόνια. Ο Shlosberg [30] σε μια από τις πρώτες προσεγγίσεις χρησιμοποίησε τρεις άξονες για την περιγραφή των συναισθημάτων: αποδοχής-απόρριψης A-R (attention-rejection), ευαρέσκειας-δυσαρέσκειας P-U (pleasantness-unpleasantness), και βαθμού ενεργοποίησης (level of activation). Για παράδειγμα συναισθήματα όπως η περιφρόνηση και η απέχθεια χαρακτηρίζονται από υψηλή τιμή απόρριψης ενώ συναισθήματα όπως η οργή χαρακτηρίζονται από υψηλή τιμή απαρέσκειας. Οι Ekman και Friesen [31] θεωρούν έξι πρωτεύουσες εκφράσεις οι οποίες συνδέονται με αντίστοιχα συναισθήματα και οι οποίες μπορούν να αναγνωριστούν από τις μορφοποιήσεις του προσώπου. Οι εκφράσεις αυτές είναι χαρά, λύπη, έκπληξη, απέχθεια, οργή και φόβος. Πέρα από τις πρωτεύουσες εκφράσεις κατέγραψαν πολλές άλλες οι οποίες όμως δεν είναι τόσο εύκολα ανιχνεύσιμες. Στο πεδίο των γραφικών για υπολογιστές και ιδιαίτερα στο animation έχουν εμφανιστεί πολλές εργασίες οι οποίες μοντελοποιούν τις εκφράσεις με βάση την κίνηση των μυών του προσώπου. Πρωτοποριακή στον τομέα αυτό θεωρείται η εργασία των Terzopoulos και Waters [43] οι οποίοι μεταξύ άλλων πρότειναν και ένα μοντέλο για τους ιστούς και το δέρμα του προσώπου. Πέρα από την μοντελοποίηση χρησιμοποίησαν την αντίστροφη διαδικασία για την ανάλυση των εκφράσεων (analysis by synthesis).

Στην ανάλυση εκφράσεων τα χαρακτηριστικά του προσώπου είναι ιδιαίτερα σημαντικά και η αποτελεσματική ανίχνευση τους καθίσταται επιτακτική. Για το σκοπό αυτό οι εικόνες των προσώπων πρέπει να αριετὰ υψηλής ανάλυσης και τα χαρακτηριστικά ευδιάκριτα.

2. ΠΡΟΣΩΠΑ ΚΑΙ ΣΥΝΑΙΣΘΗΜΑΤΙΚΕΣ ΚΑΤΑΣΤΑΣΕΙΣ

Η έρευνα σχετικά με την αναγνώριση συναισθημάτων μέσα από τις εκφράσεις του προσώπου [31], εμπνεύστηκε από την πρωτοποριακή εργασία του Darwin [32] και από τις μελέτες σχετικά με την αναγνώριση προσώπων που εμφανίστηκαν στην διεθνή βιβλιογραφία τελευταίες δεκαετίες [33]-[35]. Συμπληρωματικές, ως προς τις παραδοσιακές υπολογιστικές τεχνικές, είναι οι μελέτες οι οποίες προσεγγίζουν τον τρόπο λειτουργίας των νευρολογικών μηχανισμών οι οποίοι εμπλέκονται στη διαμόρφωση της διαδικασίας αναγνώρισης συναισθημάτων από τον άνθρωπο [36].

2.1 Ένα πλαίσιο περιγραφής της απεικόνισης των συναισθημάτων στα πρόσωπα

Η παραδοσιακή τακτική αναγνώρισης των συναισθημάτων μέσα από την έκφρασή τους στο πρόσωπο είναι όμοια με αυτήν της αναγνώρισης τους μέσω της ομιλίας: Τα δεδομένα εισόδου είναι στατικά και επιδεικνύουν απλά την κορύφωση του συναισθήματος – ‘apex’. Στην περίπτωση της οπτικής πληροφορίας η απεικόνιση είναι απλά μια φωτογραφία στην οποία το υποκείμενο εικονίζεται στο υψηλότερο εκφραστικό επίπεδο. Οι σύγχρονες τεχνικές πάντως τείνουν να εμπλέκουν πληροφορία δυναμικής υψής για την κατανόηση των συναισθημάτων μέσα από τις εκφράσεις του προσώπου.

Η αλήθεια είναι ότι η αναγνώριση συναισθημάτων από την οπτική πληροφορία έχει ένα ισχυρότερο υπόβαθρο, προερχόμενο από μελέτες στο τομέα της Νευροφυσιολογίας, από ότι η αναγνώριση τους από ηχητικά δεδομένα. Αυτό οφείλεται σε μεγάλο βαθμό στο γεγονός ότι πολλές μελέτες σχετικές με την αναγνώριση προσώπων διερεύνησαν και τον τρόπο με τον οποίο οι άνθρωποι ερμηνεύουν τα συναισθήματα των συνανθρώπων τους με βάση τις αντιλαμβανόμενες εκφράσεις [36][37]. Από την άλλη πλευρά η συντριπτική πλειοψηφία των μελετών σχετικά με την κατανόηση συναισθημάτων μέσω των εκφράσεων επικεντρώθηκαν απλά στην ταξινόμηση τους σε έξι πρωτεύουσες κατηγορίες. Η τάση αυτή θα πρέπει να αποδοθεί στον Ekman και τους συνεργάτες του οι οποίοι αποφάνθηκαν ότι πρακτικά μόνο τα έξι πρωτεύοντα συναισθήματα –ή καλύτερα κατηγορίες συναισθημάτων- μπορούν να απεικονιστούν μέσα από συγκεκριμένες μορφοποιήσεις του προσώπου.

Το παράδοξο είναι ότι οι Ekman και Friesen [38] είναι από τους ελάχιστους ερευνητές οι οποίοι εξέτασαν τον τρόπο μορφοποίησης και σύνθεσης και μη πρωτευουσών εκφράσεων –με αλλαγή για παράδειγμα του αριθμού των περιοχών του προσώπου που εμπλέκονται στις εκφράσεις, της χρονικής εξέλιξης τους και της έντασης της δράσης των μυών. Στην ίδια μελέτη εξετάστηκε και πώς η ανάμιξη συναισθημάτων μπορεί να οδηγήσει σε ουδετεροποίηση των εκφράσεων και φυσικά απώλεια της αντίληψής τους. Για παράδειγμα η ανάμιξη της οργής με τη λύπη μπορεί είτε να οδηγήσει σε μια μη αναγνωρίσιμη έκφραση του προσώπου είτε στην επικάλυψη της λιγότερη ισχυρής εκ των δύο από την άλλη. Σχετικά πρόσφατα ξεκίνησε κάποια πρώιμη έρευνα σχετικά με τη σύνθεση ενδιάμεσων εκφράσεων με προβολή των παραμέτρων μορφοποίησης των εκφράσεων –ουσιαστικά δράση μυών- στο χώρο ενεργοποίησης–επαλήθευσης [39]. Η επέκταση πάντως των τεχνικών αυτών και στην ανάλυση μη πρωτευουσών εκφράσεων συναντά σημαντικές δυσκολίες με βασικότερη τη δυσκολία αντίληψης της δράσης των μυών χωρίς τη χρήση ηλεκτροδυναμικών διατάξεων.

Βέβαια παρά την επιέντρωση στην ανάλυση των εκφράσεων που σχετίζονται με τα πρωτεύοντα συναισθήματα, η οποία δείχνει πρακτικά προσέγγιση της αναγνώρισης συναισθημάτων μέσα από ένα στενό πρίσμα, ουδείς μπορεί να αμφισβητήσει την αξία της υλοποίησης ενός τέτοιου συστήματος σε περιβάλλοντα επικοινωνίας ανθρώπου μηχανής. Στη πράξη ακόμα και ένας υπολογιστής που αναγνωρίζει τις πρωτεύουσες εκφράσεις είναι σημαντικά πιο έξυπνος από αυτόν που δεν το καταφέρνει.

3. ΑΠΕΙΚΟΝΙΣΕΙΣ ΠΟΥ ΣΧΕΤΙΖΟΝΤΑΙ ΜΕ ΤΗΝ ΣΥΝΑΙΣΘΗΜΑΤΙΚΗ ΕΚΦΡΑΣΗ

Τα χαρακτηριστικά του προσώπου [38] μπορούν να θεωρηθούν είτε στατικά –όπως για παράδειγμα το χρώμα του δέρματος- είτε αργά μεταβαλλόμενα –όπως υφή η οποία μεταβάλλεται με την ανάπτυξη ρυτίδων- είτε κινούμενα –όπως οι βλεφαρίδες, τα φρύδια κ.ο.κ. Η ανίχνευση της θέσης των χαρακτηριστικών αυτών από χρονικά σταθερές απεικονίσεις –φωτογραφίες- είναι ο στόχος των στατικών προσεγγίσεων της ανάλυσης εκφράσεων. Παρόλα αυτά υπάρχει ισχυρή ένδειξη ότι η αναγνώριση εκφράσεων από τον άνθρωπο στηρίζεται περισσότερο σε πληροφορία δυναμικής υψής παρά σε στατικές απεικονίσεις. Ο Bassili [40] πραγματοποίησε μια μελέτη στην οποία μια ομάδα ανθρώπων κλήθηκε να αναγνωρίσει εκφράσεις σε ακολουθίες βίντεο στις οποίες υπήρχαν φωτεινές κουκκίδες μόνο στις θέσεις των χαρακτηριστικών του προσώπου και οι υπόλοιπες περιοχές ήταν σκοτεινές. Από την συγκεκριμένη μελέτη προέκυψε ότι αναγνώριση πάνω από το επίπεδο τυχαιότητας ήταν εφικτή για όλες τις εκφράσεις, όταν χρησιμοποιούνταν ακολουθίες βίντεο ως πηγή πληροφορίας, ενώ με βάση τις στατικές εικόνες μόνο οι εκφράσεις “χαρά” και “λύπη” αναγνωρίστηκαν σε ποσοστό υψηλότερο από το επίπεδο τυχαιότητας. Δυστυχώς το συμπέρασμα του Bassili έχει ουσιαστική αξία μόνο όσον αφορά την αντίληψη των εκφράσεων από τον άνθρωπο γιατί και η αναγνώριση εκφράσεων από τον υπολογιστή με βάση ακολουθίες βίντεο αντιμετωπίζει πολλά προβλήματα. Η ειδοποιός διαφορά μεταξύ ανθρώπου και υπολογιστή είναι η ακρίβεια εντοπισμού των προσώπων και των χαρακτηριστικών τους στο χώρο. Το ανθρώπινο οπτικό σύστημα είναι εξαιρετικά αποτελεσματικό στον τομέα αυτό. Αντίθετα στους υπολογιστές τα σφάλματα εντοπισμού του προσώπου και των χαρακτηριστικών του, λειτουργούν προσθετικά και σε πολλές περιπτώσεις καλύπτουν την ουσιαστική πληροφορία κίνησης που διατίθεται από τις ακολουθίες. Το τελικό συμπέρασμα είναι ότι οι στατικές προσεγγίσεις οι οποίες είναι λιγότερο επιρρεπείς στον εντοπισμό των χαρακτηριστικών του προσώπου διατηρούν την αξία τους όσον αφορά την αναγνώριση των εκφράσεων. Από την άλλη πλευρά τα στάδια προεπεξεργασίας τα οποία αφορούν στον εντοπισμό του προσώπου, των βασικών χαρακτηριστικών του –όπως τα μάτια, μύτη, το στόμα κ.ο.κ- και σημείων στη περιοχή των χαρακτηριστικών αυτών είναι εξαιρετικά σημαντικά σε όλες τις περιπτώσεις. Οι δυναμικές προσεγγίσεις αντιμετωπίζουν επιπλέον και την πρόκληση της παρακολούθησης της κίνησης των χαρακτηριστικών με τεχνικές εκτίμησης κίνησης και μοντελοποίησης των μυών και των δράσεων τους.

Τη βάση για τα περισσότερα από τα συστήματα αναγνώρισης εκφράσεων αποτέλεσε η εργασία των Ekman και Friesen[31], οι οποίοι δημιούργησαν ένα σύστημα κωδικοποίησης το οποίο περιγράφει όλες τις οπτικά διαχωρίσιμες κινήσεις του προσώπου το οποίο ονόμασαν FACS - Facial Action Coding System. Το FACS είναι ένα σύστημα ανατομικής περιγραφής το οποίο στηρίζεται στον ορισμό των «μονάδων δράσης» –AU action units. Κάθε AU αντιστοιχεί στην ταυτόχρονη δράση μιας ομάδας μυών οι οποίοι διαμορφώνουν

μια συγκεκριμένη δράση στο πρόσωπο. Δεδομένου ότι αρκετοί μύες συμμετέχουν σε περισσότερες από μία AU δεν υπάρχει σαφής αντιστοιχία μύων και AU. Ένα σύνολο από 46 AU καλύπτει πλήρως τον έλεγχο των εκφράσεων ενώ άλλες 12 είναι υπεύθυνες για την θέση και κίνηση της ίριδας των ματιών. Το μοντέλο FACS χρησιμοποιήθηκε αποδοτικά για την σύνθεση εκφράσεων ενώ η χρήση του για ανάλυση εκφράσεων εξακολουθεί να ερευνάται [41]-[47]. Ο Ekman και οι συνεργάτες του δημιούργησαν επίσης και ένα λεξικό το EMFACS στο οποίο δηλώνονται οι AU οι οποίες περιγράφουν τις πρωτεύουσες εκφράσεις. Στη συνέχεια δημιούργησαν τη βάση FACS-AID η οποία χρησιμοποιείται για τον υπολογισμό των συναισθηματικών εκφράσεων με βάση τις μετρήσεις των παραμέτρων του FACS [48].

Το μοντέλο FACS ενέπνευσε και τη δημιουργία των παραμέτρων περιγραφής προσώπου και απόδοσης κίνησης προσώπου στο πλαίσιο του προτύπου ISO MPEG-4 [49]. Το σύνολο παραμέτρων για τον ορισμό προσώπου (Facial Definition Parameter set, FDP) και το σύνολο παραμέτρων απόδοσης κίνησης προσώπου (Facial Animation Parameter set, FAP) έχουν σχεδιαστεί για να επιτρέπουν τον ορισμό του σχήματος και της υφής ενός προσώπου καθώς και την αναπαράσταση εκφράσεων, συναισθημάτων καθώς και προφοράς ομιλίας. Το σύνολο FDP αποτελείται από ένα τρισδιάστατο πλέγμα, ένα σύνολο από σημεία στο τρισδιάστατο χώρο και ένα σύνολο από εναλλακτικά χαρακτηριστικά όπως μαλλιά, γυαλιά, ρυτίδες κ.ο.κ., τα οποία επιτρέπουν τον ορισμό του ακριβούς σχήματος προσώπου καθώς και της υφής του στην φάση του setup. Εάν αυτά τα στοιχεία χρησιμοποιηθούν στην φάση της αρχικής ρύθμισης ενός συνθετικού μοντέλου, είναι πιθανόν να παράγουμε με ακόμα μεγαλύτερη ακρίβεια τις κινήσεις του συγκεκριμένου προσώπου. Το σύνολο παραμέτρων απόδοσης κίνησης προσώπου (FAPs) είναι βασισμένο στην μελέτη των ελάχιστων δράσεων του προσώπου και είναι στενά συνδεδεμένο με τις δράσεις των μύων. Οι δράσεις, όπως για παράδειγμα η σύμπτυξη των φρυδιών και το άνοιγμα του στόματος, επιτρέπουν την αναπαράσταση των πιο φυσικών μορφοποιήσεων του προσώπου. Όλες οι παράμετροι που αφορούν την απόδοση κίνησης προσώπου εκφράζονται σύμφωνα με τις μονάδες παραμέτρων αναπαράστασης προσώπου (Facial Animation Parameters Units). Οι μονάδες αυτές έχουν σχεδιαστεί έτσι ώστε να επιτρέπουν την απεικόνιση των παραμέτρων απόδοσης κίνησης σε κάθε μοντέλο προσώπου με έναν συνεχή τρόπο παράγοντας λογικά αποτελέσματα σε ότι αφορά τις εκφράσεις και την ανθρώπινη προφορά. Ανταποκρίνονται στον τεμαχισμό των αποστάσεων ανάμεσα σε κάποια χαρακτηριστικά κλειδιά του ανθρώπινου προσώπου. Το μέγεθος του τεμαχισμού που εφαρμόζεται επιλέγεται ανάλογα με την επιθυμητή ακρίβεια. Για λεπτομερέστερη ανάλυση ο αναγνώστης παραπέμπεται στο Κεφάλαιο Δ3.

4. ΝΕΥΡΟΦΥΣΙΟΛΟΓΙΑ ΚΑΙ ΣΥΝΑΙΣΘΗΜΑΤΑ

Όπως και στις περιπτώσεις αναγνώρισης προσώπου, υποστηρίζεται ότι υπάρχουν συγκεκριμένες περιοχές του εγκεφάλου οι οποίες είναι υπεύθυνες για την αντίληψη εκφράσεων από εικόνες. Συγκεκριμένα η απώλεια του αμυγδάλου (amygdala) προκαλεί την αδυναμία αναγνώρισης των εκφράσεων «οργή» και «φόβος» σε φωτογραφίες προσώπων [51]. Επίσης εμφανίζεται αυξημένη δραστηριότητα του αμυγδάλου κατά την παρακολούθηση παραλλαγών της έκφρασης «απέχθεια» ακόμα και αν το πρόσωπο στο οποίο απεικονίζεται η έκφραση είναι άγνωστο [52]. Πάντως δεν υπάρχει σαφής ένδειξη για την περιοχή του εγκεφάλου που είναι υπεύθυνη για την κατανόηση ευχάριστων συναισθημάτων όπως η «χαρά», η «ικανοποίηση» κ.ο.κ [53].

Το πιο αξιοσημείωτο συμπέρασμα πάντως είναι ότι η επεξεργασία της πληροφορίας που σχετίζεται με την αναγνώριση συναισθημάτων πραγματοποιείται σε διαφορετικές περιοχές του εγκεφάλου από ότι η αναγνώριση προσώπων και επιπλέον είναι πιθανόν να εμπλέκει και διαφορετικά είδη διεγέρσεων.

5. ΤΕΧΝΙΚΕΣ ΑΝΑΓΝΩΡΙΣΗΣ ΕΚΦΡΑΣΕΩΝ

Στην παράγραφο αυτή παρουσιάζονται μερικές από τις τεχνικές οι οποίες έχουν εμφανιστεί στη βιβλιογραφία και πραγματεύονται τους τρόπους με τους οποίους ο υπολογιστής μπορεί να ανακτήσει πληροφορίες σχετικά με τη συναισθηματική κατάσταση κάποιου ατόμου μέσα από τις εκφράσεις του προσώπου του. Οι προσεγγίσεις του ανωτέρω ζητήματος διακρίνονται σε δύο κατηγορίες: (α) στατικές, στις οποίες η αναγνώριση της έκφρασης βασίζεται σε μια απλή φωτογραφία στην οποία εικονίζεται η έκφραση στην κορύφωση της, και (β) δυναμικές, στις οποίες η αναγνώριση πραγματοποιείται με τη χρήση μιας ακολουθίας από καρέ τα οποία επιδεικνύουν την χρονική εξέλιξη της έκφρασης. Η διάρκεια μιας ακολουθίας που απεικονίζει κάποια έκφραση κυμαίνεται από 0.5 έως 4 δευτερόλεπτα [35]. Μια ενδιάμεση κατηγορία αποτελούν οι τεχνικές οι οποίες χρησιμοποιούν δύο καρέ, ένα με το πρόσωπο σε ουδέτερη κατάσταση και ένα με το πρόσωπο στη κορύφωση της έκφρασης. Παρόλο που τεχνικές αυτές πλησιάζουν την περισσότερη δυναμική προσέγγιση θα μπορούσε καταχρηστικά κάποιος να τις χαρακτηρίσει ημιστατικές.

Παρακολούθηση της κίνησης του προσώπου

Όπως αναφέρθηκε στη προηγούμενη παράγραφο οι άνθρωποι οφείλουν ένα μεγάλο ποσοστό της ικανότητας τους να αναγνωρίζουν εκφράσεις στον εξαιρετικά ακριβή εντοπισμό του προσώπου και των χαρακτηριστικών του. Σε κάθε αλγόριθμο αναγνώρισης εκφράσεων ο εντοπισμός του προσώπου και η παρακολούθηση της κίνησης του είναι ζητήματα πολύ σημαντικά. Στο δεύτερο μέρος της διατριβής προτείνονται διάφορα σχήματα εντοπισμού προσώπων. Στη παράγραφο αυτή συνοψίζονται μερικές από τις τεχνικές που χρησιμοποιούνται για την παρακολούθηση της κίνησης του προσώπου [54]-[59]. Στις εργασίες [54]-[59] το πρόσωπο ανιχνεύεται με βάση την χρωματική κατανομή του δέρματος [60]-[62], σε κάθε καρέ ξεχωριστά, και στη συνέχεια υπολογίζεται η συνολική μετατόπιση του ανάμεσα στα δύο καρέ. Η προσέγγιση αυτή είναι αξιόπιστη όταν εφαρμόζεται σε ακολουθίες βίντεο με ελεγχόμενες συνθήκες φωτισμού και όταν το πρόσωπο κοιτάζει ευθεία στη κάμερα –εφαρμογές τηλεδιάσκεψης. Σε κάθε άλλη περίπτωση η ακρίβεια υπολογισμού της κίνησης είναι χαμηλή και επηρεάζεται από την κλίση και στροφή του προσώπου. Ως τελικό αποτέλεσμα το σφάλμα, όσον αφορά στην εκτίμηση της κίνησης του προσώπου, είναι μεγαλύτερο από την ίδια την κίνηση των χαρακτηριστικών που διαμορφώνουν την έκφραση. Οι τεχνικές παρακολούθησης της κίνησης του προσώπου με χρήση προτύπων, active contours, γράφων [63], wavelets [64] και RBF συναρτήσεων [65], είναι σε γενικές γραμμές αναίσθητες ως προς affine διακυμάνσεις και μετασχηματισμούς αλλά έχουν υψηλή υπολογιστική πολυπλοκότητα η οποία τις καθιστά ακατάλληλες σε εφαρμογές πραγματικού χρόνου. Εκτός από τις τεχνικές παρακολούθησης της κίνησης του προσώπου σημαντικές είναι και οι τεχνικές οι οποίες προσπαθούν να εκτιμήσουν την θέση του στο τρισδιάστατο χώρο -pose estimation- και να αντισταθμίσουν τις αλλαγές που εμφανίζονται στα χαρακτηριστικά του προσώπου εξαιτίας της οριοθέτησης του [66][67].

5.1 Στατικές προσεγγίσεις

Οι περισσότερες από τις μελέτες όσον αφορά την ανάλυση εκφράσεων πραγματοποιήθηκαν σε «mug shot» φωτογραφίες στις οποίες απεικονίζονται οι εκφράσεις στην κορύφωση τους [68]. Οι φωτογραφίες αυτές επιτρέπουν την ανίχνευση κάποιων στατικών στοιχείων όπως οι ρυτίδες στο μέτωπο, στα μάγουλα και ανάμεσα στα φρύδια, με βάση τα οποία πραγματοποιείται η ταξινόμηση τους σε συγκεκριμένες κατηγορίες. Η εξαγωγή των στοιχείων αυτών δεν είναι ούτε εύκολη αλλά ούτε ικανή για το διαχωρισμό εκφράσεων. Η λογική της ταξινόμησης είναι μάλλον «το μη χειρόν βέλτιστο» παρά η ουσιαστική περιγραφή των εκφράσεων. Πολύ λίγες από τις στατικές προσεγγίσεις εμφανίζουν αξιοσημείωτα αποτελέσματα [69]. Μια σχετικά επιτυχημένη τεχνική παρουσιάζεται στην εργασία [70], στην οποία ένα σύνολο από νευρωνικά δίκτυα, τα οποία λειτουργούν παράλληλα, πραγματοποιούν προβολή των μπλοκ της εικόνας στους PCA άξονες των περιοχών ενδιαφέροντος –μάτια και στόμα. Μετά τη διαδικασία εκπαίδευσης το νευρωνικό δίκτυο με την μεγαλύτερη ικανότητα ανάκλησης χρησιμοποιείται ως ταξινομητής. Οι συγγραφείς της εργασίας αναφέρουν ποσοστό γενίκευσης 86% -στα ίδια δεδομένα οι άνθρωποι είχαν ποσοστό επιτυχίας 92%- αλλά για την εφαρμογή της τεχνικής απαιτείται επακριβής κανονικοποίηση των εικόνων εισόδου ως προς την κλίμακα και την διάταξη των χαρακτηριστικών. Επίσης η αυτόματη εξαγωγή των μπλοκ που περιέχουν τα μάτια και το στόμα με αρκετά μεγάλη ακρίβεια είναι αμφίβολη.

Το γεγονός ότι οι στατικές προσεγγίσεις, παρά την αμφίβολη αποτελεσματικότητά τους, είναι αρκετά δημοφιλείς στη βιβλιογραφία μπορεί να αποδοθεί σε δύο παράγοντες: (α) στην ύπαρξη αρκετών βάσεων με στατικές εικόνες σε αντίθεση με τις λίγες, στον αριθμό αλλά και πλήθος ακολουθιών, δυναμικές βάσεις, και (β) στα επίπονα στάδια προεπεξεργασίας που απαιτούνται στις δυναμικές προσεγγίσεις.

5.2 Δυναμικές προσεγγίσεις

(a) Εκτίμηση της κίνησης των σημείων στην επιφάνεια του προσώπου.

Ένα κρίσιμο θέμα στις δυναμικές προσεγγίσεις είναι η εκτίμηση της κίνησης των χαρακτηριστικών και γενικότερα των σημείων του προσώπου από ένα καρέ στο επόμενο. Οι πιο διαδεδομένες προσεγγίσεις στο πρόβλημα αυτό χρησιμοποιούν εκτίμηση των χωρικών και χρονικών παραγώγων των καρέ, φιλτράρισμα ή απλά συσχέτιση στο χώρο των εικόνων. Η εκτίμηση των χωρικών και χρονικών παραγώγων βασίζεται στην προσέγγιση των Horn και Schunck [71], οι οποίοι υποθέτουν ότι το δέρμα του προσώπου είναι τοπικά ομοιογενές και τα βασικά χαρακτηριστικά του διακρίνονται από σχετικά υψηλό βαθμό ανομοιογένειας. Η υπόθεση αυτή είναι ορθή όταν το πρόσωπο έχει απομονωθεί από το φόντο, σε διαφορετική περίπτωση αν το φόντο είναι ανομοιογενές και υψίσυχο, η τεχνική οδηγείται σε αποτυχία [45]. Η εκτίμηση της κίνησης με φιλτράρισμα [72] απαιτεί την εφαρμογή φίλτρων τόσο στο πεδίο του χρόνου όσο και στο χώρο. Με τον τρόπο αυτό εκτιμάται η χωροχρονική συχνοτική κατανομή, ανάμεσα σε μια σειρά από καρέ, η εξέλιξη της οποίας οδηγεί στη κατασκευή του πεδίου των διανυσμάτων κίνησης. Το μειονέκτημα της μεθόδου είναι ότι απαιτεί ένα αρκετά μεγάλο αριθμό από καρέ για τη σωστή εκτίμηση του πεδίου κίνησης. Οι τεχνικές συσχέτισης [73][74] στηρίζονται στην κλασική ιδέα της πρόβλεψης κίνησης η οποία χρησιμοποιείται στο πρότυπο MPEG. Ο υπολογισμός της κίνησης ενός pixel σε δύο διαδοχικά καρέ βασίζεται στην υπόθεση

ότι κίνηση αυτή είναι μικρή και υπολογίζεται με απευθείας σύγκριση της γραμμικά φιλτραρισμένης τιμής του pixel στο τρέχον καρέ με τις αντίστοιχες γειτονικές του pixel θέσεις στο επόμενο καρέ. Ο υπολογισμός της κίνησης των pixels με τον τρόπο αυτό μοιάζει απλοϊκός και επιρρεπής σε σφάλματα, κάτω όμως από ένα μακροσκοπικό πρίσμα μπορούν να εξαχθούν ορθά συμπεράσματα. Για παράδειγμα η συνέχεια η οποία παρουσιάζει το δέρμα του προσώπου υποχρεώνει ομοιόμορφη ομαδική κίνηση των σημείων σε αυτό. Επομένως ακόμα και αν για κάποια από αυτά η κίνηση έχει εκτιμηθεί εσφαλμένα η μέση κίνηση θα έχει εκτιμηθεί ορθά. Από την άλλη πλευρά πάντως, με την ανωτέρω τεχνική, αξιοπιστία μπορεί να αποδοθεί στην εκτίμηση της διεύθυνσης της κίνησης παρά στο μέτρο της.

Σε όλες τις περιπτώσεις η σωστή εκτίμηση της κίνησης των σημείων του προσώπου απαιτεί αποτελεσματική αντιστάθμιση της ολικής κίνησης του προσώπου. Όπως αναφέρθηκε στην προηγούμενη παράγραφο σε πολλές περιπτώσεις το σφάλμα στην εκτίμηση της ολικής κίνησης του προσώπου είναι μεγαλύτερο από την ίδια την κίνηση των σημείων σε αυτό.

Ημιστατικές προσεγγίσεις

Οι ημιστατικές προσεγγίσεις ταξινομούν τις εκφράσεις βασιζόμενες σε χαρακτηριστικά τα οποία πηγάζουν από τον υπολογισμό της κίνησης, είτε των μυών του προσώπου είτε των χαρακτηριστικών του, ανάμεσα σε δύο καρέ τα οποία απεικονίζουν το πρόσωπο στην ουδέτερη κατάσταση και στην κορύφωση της έκφρασης. Είναι προφανές ότι στηρίζονται σε δυναμικά στοιχεία και έχουν δύο πλεονεκτήματα: (α) Η κίνηση των επιμέρους σημείων του προσώπου ανάμεσα στα δύο αυτά καρέ είναι αρκετά μεγάλη, και (β) υπάρχει αρκετό υλικό για τον αποτελεσματικό έλεγχο των αλγορίθμων. Από την άλλη πλευρά ουδείς αμφισβητεί ότι η χρονική εξέλιξη της έκφρασης αποτελεί σημαντικό χαρακτηριστικά της και προφανώς η χρήση δύο μόνο καρέ αδυνατεί να καταγράψει την χρονική αυτή εξέλιξη. Ο Mase [45] προσπάθησε να υπολογίσει την κίνηση κάποιων ομάδων μυών οι οποίοι αντιστοιχούν στις AUs όπως αυτές ορίζονται στο FACS. Η κίνηση των μυών προσεγγίζεται με την εκτίμηση της κίνησης κάποιων ορθογωνίων της επιφάνειας του προσώπου και στη συνέχεια χρησιμοποιείται για τον χαρακτηρισμό των εκφράσεων. Το πρόβλημα στην προσέγγιση αυτή είναι ότι ο αυτόματος εντοπισμός των ορθογωνίων που περιλαμβάνουν τους μύες είναι πολύ δύσκολος, δεδομένου ότι στην πλειοψηφία τους τα ορθογώνια αντιστοιχούν σε ομοιόμορφες περιοχές δέρματος οι οποίες δεν προσδιορίζονται από κάποιο συγκεκριμένο χαρακτηριστικό. Επιπλέον τίθεται και το ερώτημα κατά πόσον η τρισδιάστατη κίνηση των μυών μπορεί να απεικονιστεί από την κίνηση των σημείων της επιφάνειας του προσώπου. Βέβαια ούτε και ο άνθρωπος υπολογίζει την πραγματική κίνηση των μυών, απλά παρατηρεί τις μορφοποιήσεις του προσώπου. Ο ίδιος συγγραφέας σε μια πιο ρεαλιστική προσέγγιση θεωρεί το πρόσωπο ως ένα ψηφιδωτό απαρτιζόμενο από επιμέρους περιοχές για κάθε μια από τις οποίες υπολογίζεται το μέσο διάνυσμα κίνησης και η διασπορά του. Η ταξινόμηση των εκφράσεων πραγματοποιείται με τον κανόνα k -means. Τα αποτελέσματα της προσέγγισης αυτής αποδεικνύουν ότι μοντελοποίηση της φυσικής δομής των μυών, για σκοπούς ανάλυσης, προσθέτει υπολογιστική πολυπλοκότητα χωρίς αντίστοιχο κέρδος όσον αφορά την απόδοση. Στο Κεφάλαιο Δ4 της διατριβής παρουσιάζεται μια επέκταση της τεχνικής του Mase σε πλήρως δυναμικό περιβάλλον και με κατευθυντική ομαδοποίηση των διανυσμάτων κίνησης σε κάθε περιοχή του προσώπου.

Οι Yacoob και Davis [75] επικεντρώθηκαν στις ακμές αντί στους μύες του προσώπου, θεωρώντας ότι οι ακμές και η κίνηση τους υπολογίζονται ευκολότερα και είναι περισσότερο ευσταθή χαρακτηριστικά, όσον αφορά τις συνθήκες φωτισμού και τις αλλαγές της θέσης του προσώπου στον τρισδιάστατο χώρο. Ενοποιώντας τις περιγραφές των Ekman και Friesen [38] και τα πρότυπα κίνησης των επιμέρους εκφράσεων του Bassili [40], κατέληξαν στη δημιουργία μιας γλωσσικής και ενδιάμεσου επιπέδου περιγραφής η οποία μοντελοποιεί την χωροχρονική δραστηριότητα του προσώπου. Η τελική ταξινόμηση πραγματοποιείται με τη χρήση κανόνων απόφασης [38][40]. Η προσέγγιση των Yacoob και Davis, ανεξάρτητα από τα προβλήματα στον υπολογισμό της κίνησης των ακμών και της αμφιβολίας του κατά πόσον η κίνηση αυτή δίνει όλη την απαιτούμενη πληροφορία της μορφοποίησης του προσώπου κατά τη διάρκεια μιας έκφρασης, έχει υψηλό λογικό υπόβαθρο: Η ερμηνεία των εκφράσεων από τις κινήσεις των σημείων της επιφάνειας του προσώπου δεν μπορεί να είναι ούτε ομοιόμορφη σε όλο το πλάτος του ανθρώπινου πληθυσμού αλλά ούτε και ντετερμινιστική. Η χρήση κανόνων για την ταξινόμηση εκφράσεων υιοθετήθηκε και από τον συγγραφέα της διατριβής και αναλύεται στο Κεφάλαιο Δ3.

Η θεώρηση των Li, Roivainen και Forchheimer [44] κλίνει περισσότερο προς τη σύνθεση εκφράσεων παρά προς την ανάλυση. Χρησιμοποίησαν ως βάση το FACS για την ανάλυση εικόνων με εκφράσεις με τελικό σκοπό την συνθετική αναπαράστασή τους. Η αξία της εργασίας τους επικεντρώνεται κυρίως στο διαχωρισμό της κίνησης «δύσκαμπτων» -rigid- και μη περιοχών του προσώπου η οποία υπολογίζεται με τη χρήση τουλάχιστον δύο καρέ. Χρησιμοποίησαν επίσης μόλις 6 AUs για να αναπαραστήσουν τις πιθανές εκφράσεις.

(β) *Η ολική κατανομή των διανυσμάτων κίνησης στην επιφάνεια του προσώπου και οι εκφράσεις*

Σε μια προσπάθεια να διευκρινίσουμε την φιλολογία σχετικά με την αξία των διανυσμάτων κίνησης που σχετίζονται με τα σημεία της επιφάνειας του προσώπου, όσον αφορά τον χαρακτηρισμό των εκφράσεων, πραγματοποιήσαμε το πείραμα που περιγράφεται στη συνέχεια:

Έστω ότι όλες οι εικόνες που απεικονίζουν τις επιμέρους εκφράσεις του ίδιου προσώπου είναι κανονικοποιημένες ως προς την κλίμακα και ευθυγραμμισμένες ως προς τη θέση κάποιων σταθερών σημείων του προσώπου –κόγχες ματιών και βάση της μύτης. Η υπόθεση αυτή εξαλείφει την ολική κίνηση του κεφαλιού και το zoom της κάμερας. Υπολογίζουμε το πεδίο των διανυσμάτων κίνησης απευθείας από τις τιμές των pixels και για εκείνες μόνο τις περιοχές του προσώπου για τις οποίες έχει λάβει χώρα ουσιαστική δράση [76].

Έστω F_k και F_{k+1} δύο καρέ τα οποία απεικονίζουν το ίδιο πρόσωπο στην ουδέτερη κατάσταση και την κορύφωση κάποιας συγκεκριμένης έκφρασης. Κάθε pixel $p_k(x, y)$ του k -στού καρέ περιγράφεται μέσω του $2n \times 2n$ μπλοκ $b_k(x, y)$ που το περιβάλλει και αντιστοιχίζεται με το επόμενο σφάλμα (MAD-Mean Absolute Difference):

$$e_k(x, y) = |b_k(x, y) - b_{k+1}(x, y)| = \sum_{l=-n}^n \sum_{m=-n}^n |p_k(x+l, y+m) - p_{k+1}(x+l, y+m)| \quad (A5.1)$$

Τα διανύσματα μετατόπισης υπολογίζονται μόνο για εκείνα τα μπλοκ του καρέ τα οποία αντιστοιχούν σε μεγάλο σφάλμα $e_k(x, y)$ -βλέπε Κεφάλαιο Δ4.

Το διάνυσμα μετατόπισης $\hat{v}_k(x, y)$ του block $b_k(x, y)$ υπολογίζεται με τη διαδικασία ταύτισης μπλοκ σε μια γειτονιά του μπλοκ $b_{k+1}(x, y)$ σύμφωνα με την εξίσωση:

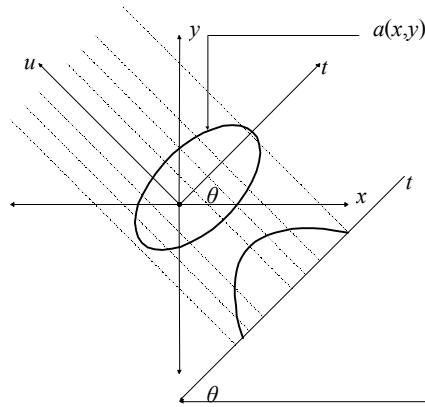
$$\hat{v}_k(x, y) = (\hat{v}_x, \hat{v}_y) = \arg \min_{(v_x, v_y) \in Q} \sum_{l=-n}^n \sum_{m=-n}^n |p_k(x+l, y+m) - p_{k+1}(x+l-v_x, y+m-v_y)| \quad (A5.2)$$

όπου $Q = \{-q, \dots, q\} \times \{-q, \dots, q\}$ ένα τετραγωνικό πλέγμα που αντιπροσωπεύει την περιοχή έρευνας. Για την μείωση του χρόνου εκτέλεσης πραγματοποιείται λογαριθμική αντί για εξαντλητική έρευνα, δηλαδή μόνο ένα υποσύνολο των συνδυασμών $(\hat{v}_x, \hat{v}_y) \in Q$ ελέγχεται στη διαδικασία ταύτισης. «Θορυβώδη» λάθος εκτιμημένα- διανύσματα κίνησης απομακρύνονται με φιλτράρισμα ενδιάμεσης τιμής, αρχικά ως προς τη φάση και στη συνέχεια ως προς το μέτρο.

Το διάνυσμα κίνησης στη θέση (x, y) μπορεί να εκφραστεί ως $\hat{v}_k(x, y) = a_k(x, y)e^{j\phi_k(x, y)}$. Ο διακριτός μετασχηματισμός Radon του μέτρου των διανυσμάτων κίνησης σε γωνία θ , δίνεται από τη σχέση:

$$R(\theta) = \sum_{u=-\infty}^{\infty} a_k(x, y) \Big|_{x=t \cos \theta - u \sin \theta, y=t \sin \theta + u \cos \theta} \quad (A5.3)$$

όπου t και u αντιστοιχούν στους άξονες x και y μετά από στροφή του συστήματος συντεταγμένων κατά γωνία θ —βλέπε Σχήμα A5.1.



Σχήμα A5.1: Ο μετασχηματισμός Radon

Ο υπολογισμός του μετασχηματισμού Radon, όπως ορίστηκε στην παραπάνω σχέση, για τις γωνίες 0^0 και 90^0 δημιουργεί τις λεγόμενες «υπογραφές» οι οποίες μπορούν να χρησιμοποιηθούν για τον χαρακτηρισμό των επιμέρους εκφράσεων. Στην προσομοίωση που πραγματοποιήσαμε, χρησιμοποιήθηκε η βάση εικόνων του Yale και πήραμε σχετικά ικανοποιητικά αποτελέσματα όσον αφορά την ταξινόμηση -82% έως 87.5% ανάλογα με το σχήμα ταύτισης [76]. Δεν πρέπει πάντως να μας διαφεύγει το γεγονός ότι τα καρέ που

αντιστοιχούν στην ουδέτερη κατάσταση και την κορύφωση ουσιαστικά ευθυγραμμίστηκαν χειρονακτικά έτσι ώστε να αντισταθμιστεί αποτελεσματικά η ολική κίνηση του προσώπου.

(γ) Πλήρως δυναμικές προσεγγίσεις

Οι τεχνικές ανάλυσης εκφράσεων από ακολουθίες βίντεο –ακολουθίες σε μορφή τηλεδιάσκεψης- μπορούν να υποδιαιρεθούν σε τρεις κατηγορίες. Ενδιαφέρον παρουσιάζει και η προσέγγιση που ορίζεται στο πρότυπο MPEG-4 για την οποία γίνεται εκτεταμένη αναφορά στο Κεφάλαιο Δ3.

Τεχνικές βασισμένες στο πεδίο των διανυσμάτων κίνησης

Στις τεχνικές αυτές υπολογίζεται το πεδίο των διανυσμάτων κίνησης είτε σε όλο το καρέ είτε σε επιλεγμένες περιοχές του προσώπου. Η ταξινόμηση των εκφράσεων πραγματοποιείται χρησιμοποιώντας παράγωγα χαρακτηριστικά των διανυσμάτων κίνησης και η ταύτιση υλοποιείται με κάποιο δυναμικό σχήμα –Hidden Markov Models (HMM), recurrent neural networks κ.ο.κ. Τυπικά χαρακτηριστικά είναι η ενεργειακή κατανομή των διανυσμάτων κίνησης σε επιμέρους περιοχές του προσώπου [76][77] και ο συνολικός προσανατολισμός τους. Δεδομένου ότι η διαδικασία υπολογισμού του πεδίου των διανυσμάτων είναι υπολογιστικά επίπονη, πολλές προσεγγίσεις υιοθετούν τη στρατηγική coarse-to-fine χρησιμοποιώντας wavelets [78]. Ένα άλλο πρόβλημα με τις τεχνικές αυτές είναι αναπόφευκτη παρουσία θορύβου στα εκτιμώμενα διανύσματα κίνησης η οποία μπορεί να επηρεάσει σημαντικά τη διαδικασία ανάλυσης.

Ο Ohya και οι συνεργάτες του [79]-[81] εφάρμοσαν Hidden Markov Models για τη μοντελοποίηση της χρονικής εξέλιξης των εκφράσεων. Η κατάσταση των μυών σε κάθε χρονική στιγμή προσεγγίζεται από τους εσωτερικούς κόμβους των HMMs –έναν για κάθε έκφραση. Στην εργασία [79] χρησιμοποίησαν wavelets για την εξαγωγή του διανύσματος χαρακτηριστικών –ισχύς του πεδίου κίνησης για διάφορες συχνοτικές περιοχές υπολογισμένες με τη βοήθεια του μετασχηματισμού wavelet- ενώ στη μελέτη [81] χρησιμοποίησαν το μετασχηματισμό Fourier. Στην περίπτωση του μετασχηματισμού Fourier εκτός από τη μέση ισχύ το διάνυσμα χαρακτηριστικών εμπλουτίστηκε με τη μέση γωνία φάσης σε κάθε πεδίο, ενώ και στις δύο περιπτώσεις τα πεδία κίνησης υπολογίστηκαν μόνο στις περιοχές των ματιών και του στόματος. Στην εργασία [80] χρησιμοποιήθηκαν HMM με συνεχή κατανομή πυκνότητας πιθανότητας για να αποφευχθεί το μειονέκτημα του αναποτελεσματικού διανυσματικού κβαντισμού εξαιτίας του μικρού αριθμού δειγμάτων [79].

Παρακολούθηση των χαρακτηριστικών του προσώπου

Η παρακολούθηση της κίνησης των σημαντικότερων χαρακτηριστικών του προσώπου κατά τη διάρκεια των εκφράσεων είναι ίσως η πιο προφανής αντιμετώπιση της ανάλυσης εκφράσεων από ακολουθίες βίντεο. Το πεδίο των διανυσμάτων κίνησης εκτιμάται μόνο στις περιοχές του προσώπου που περιλαμβάνουν τα χαρακτηριστικά αυτά. Η διαδικασία διαιρείται σε δύο στάδια: Για κάθε καρέ της ακολουθίας πραγματοποιείται εντοπισμός των χαρακτηριστικών χαμηλού επιπέδου, όπως ακμών και γωνιών που αυτές σχηματίζουν, ή υψηλότερου επιπέδου όπως τα μάτια, τα φρύδια, το στόμα κ.ο.κ [82],[84]-[87], και στη συνέχεια ελέγχεται η κίνησή τους ανάμεσα σε διαδοχικά καρέ [88]. Πολύ διαδεδομένος αλγόριθμος παρακολούθησης της κίνησης των χαρακτηριστικών είναι ο αλγόριθμος των Lucas-Kanade [83], ο οποίος εμφανίζει σχετικά υψηλή ακρίβεια όσον αφορά στην εκτίμηση της κίνησης. Η παρακολούθηση της κίνησης

χαρακτηριστικών περιορίζει την υπολογιστική πολυπλοκότητα αλλά υποθέτει μεγάλη ακρίβεια στον εντοπισμό τους. Επιπλέον δεν είναι προφανές ότι όλη η πληροφορία σχετικά με τη διαμόρφωση των εκφράσεων εκφράζεται μόνο από τη μεταβολή των βασικών χαρακτηριστικών του προσώπου. Για παράδειγμα σε πολλές εκφράσεις η εμφάνιση ρυτίδων σε ομοιόμορφες περιοχές του προσώπου –μέτωπο, μάγουλα- είναι μια σημαντική πληροφορία.

Ο Yacoob στις εργασίες [89][90], επέκτεινε την προηγούμενη μελέτη του χρησιμοποιώντας ακολουθίες με δειγματοληψία έως και 30 καρέ το δευτερόλεπτο. Υπολόγισε τη κίνηση σε ακολουθίες με το πρόσωπο σε εμπρόσθια όψη και μόνο για τις περιοχές των ακμών –περιοχές των καρέ με υψηλή τιμή στη χωρική παράγωγο [73]. Με δεδομένο ότι οι περιοχές των ματιών, των φρυδιών, του στόματος και της μύτης χαρακτηρίζονται από υψηλή πυκνότητα ακμών, η τεχνική παρακολουθεί την κίνηση των χαρακτηριστικών αυτών σε χαμηλό επίπεδο και χρειάζεται τη βοήθεια κάποιων κανόνων υψηλότερου επιπέδου για την εξαγωγή συμπερασμάτων.

Οι Rosenblum και Yacoob [91] χρησιμοποίησαν τα ίδια χαρακτηριστικά όπως στη μελέτη [69], για την ταξινόμηση εκφράσεων σε ένα πλήρως δυναμικό περιβάλλον. Παρόμοια με την τεχνική τους είναι η τεχνική του Thalmann [92] ο οποίος όμως χρησιμοποίησε μια δομή νευρωνικού δικτύου για την ταξινόμηση των εκφράσεων.

Χρήση τρισδιάστατων μοντέλων για το κεφάλι

Η τρίτη κατηγορία δυναμικών προσεγγίσεων κάνει χρήση της τρισδιάστατης υψής του κεφαλιού και προσαρμόζει ένα τέτοιο μοντέλο με βάση τα δεδομένα που προέρχονται από τα καρέ της ακολουθίας βίντεο. Η χρήση των τρισδιάστατων μοντέλων αποσκοπεί στην καλύτερη εκτίμηση της θέσης του προσώπου στο χώρο και στην εκτίμηση της κίνησης σε τρισδιάστατο επίπεδο και όχι απλά σε επίπεδο μετατόπισης. Οι Essa και Pentland σε μια σειρά μελετών [93]-[95] ασχολούνται με την παρακολούθηση των εκφράσεων στο χρόνο χρησιμοποιώντας μια εκτεταμένη αναπαράσταση η οποία βασίζεται στο FACS. Το πεδίο των διανυσμάτων κίνησης αντισταθμίζεται με τη χρήση του 3-Δ μοντέλου και η εξέλιξη της ενεργειακής κατανομής του χρησιμοποιείται για τον χαρακτηρισμό των εκφράσεων. Η ύπαρξη 3-Δ μοντέλων δεν φαίνεται να συνεισφέρει στην αποτελεσματικότερη ανάλυση εκφράσεων, αλλά για σκοπούς σύνθεσης και αναπαραγωγής εκφράσεων είναι μείζονος σημασίας.

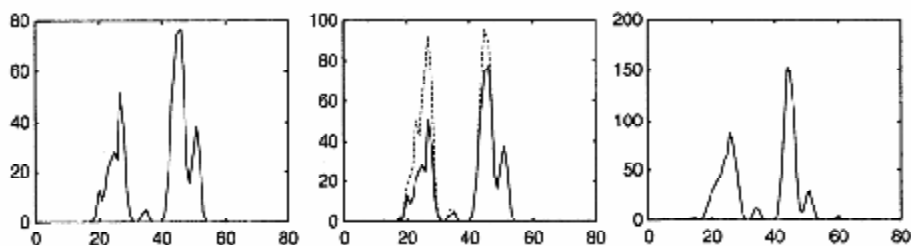
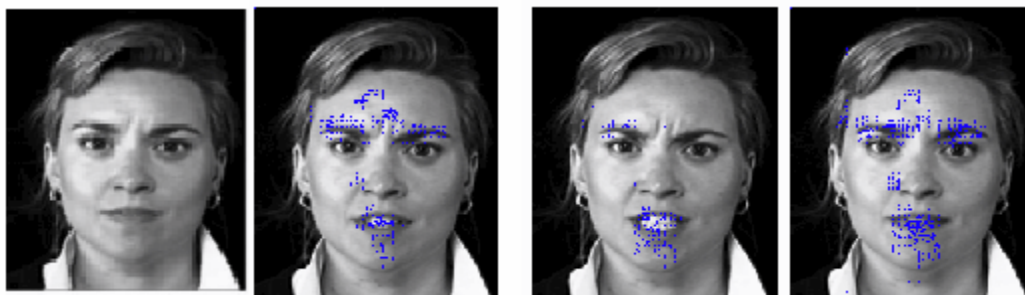
Η εργασία των Terzopoulos και Waters [43] βασίστηκε στη μελέτη των Platt και Badler [96], οι οποίοι μοντελοποίησαν το κεφάλι με ένα πλέγμα –mesh- από ισόπλευρα τρίγωνα, το μέγεθος και η πυκνότητα των οποίων σχετίζονται με την τοπική ανομοιογένεια του κεφαλιού. Η ιδιαιτερότητα της εργασίας [43] έγκειται στην μοντελοποίηση των μυών και των ιστών του δέρματος του προσώπου, η ελαστική υφή του οποίου αναπαρίσταται αρκετά αποτελεσματικά. Οι μετασχηματισμοί και οι παράμετροι ελέγχου της μοντελοποίησης αποτελούν τα χαρακτηριστικά περιγραφής των εκφράσεων. Κάθε έκφραση θεωρείται ότι περιλαμβάνει τρεις φάσεις: *εφαρμογή, απελευθέρωση και χαλάρωση*. Κάθε μια από τις φάσεις αυτές χαρακτηρίζεται από τον τρόπο δράσης των μυών και τις μορφοποιήσεις που προκαλούν στο δέρμα. Από την πλευρά της σύνθεσης εκφράσεων η εργασία των Terzopoulos και Waters θεωρείται πρωτοποριακή και δημιουργεί αρκετά φυσικές αναπαραστάσεις. Η εφαρμογή της στην ανάλυση εκφράσεων απαιτεί τη

διαίρεση της έκφρασης στις τρεις φάσεις που αναφέρθηκαν νωρίτερα και την εξαγωγή επιμέρους χαρακτηριστικών για κάθε φάση. Παρόλο που οι συγγραφείς αναφέρουν ποσοστό επιτυχούς ταξινόμησης εκφράσεων 98% -η επαλήθευση της τεχνικής πραγματοποιήθηκε σε μια μικρή βάση από 52 ακολουθίες- στη πράξη υπάρχουν σοβαρές δυσκολίες στην εφαρμογή της σε δισδιάστατες ακολουθίες. Η εκμετάλλευση της μοντελοποίησης τους θα είναι όμως χρήσιμη σε εφαρμογές 3-Δ βίντεο.

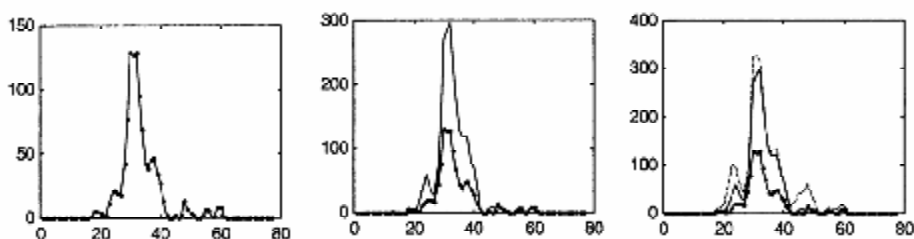
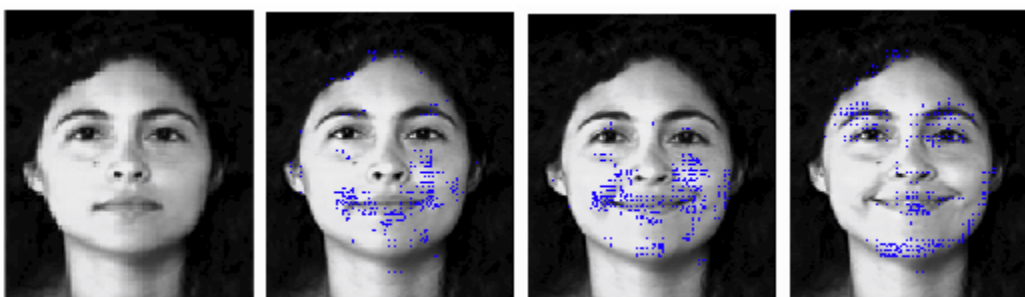
(δ) Η χρονική εξέλιξη του ενεργειακού πεδίου των διανυσμάτων κίνησης και οι εκφράσεις

Όπως και στην Παράγραφο A5.1(β) σε μια προσπάθεια να διερευνήσουμε την αξία της παρακολούθησης της χρονικής εξέλιξης των εκφράσεων υλοποιήσαμε ένα σύστημα που ταξινομεί τις εκφράσεις με βάση την χρονική εξέλιξη του ενεργειακού πεδίου των διανυσμάτων κίνησης που αντιστοιχούν σε περιοχές του προσώπου. Τα διανύσματα κίνησης υπολογίζονται με τον αλγόριθμο που περιγράφηκε στην παράγραφο Παράγραφο A5.1(β) και η ενέργειά τους συσσωρεύεται αθροιστικά με την πάροδο του χρόνου. Στις Εικόνες A5.1(α), A5.2(α) και A5.3(α) επιδεικνύονται τα «σχυρότερα» διανύσματα κίνησης σε τρεις διαδοχικές χρονικές στιγμές και για τις εκφράσεις «οργή», «χαρά» και «έκπληξη» αντίστοιχα. Οι Εικόνες A5.1(β), A5.2(β) και A5.3(β) επιδεικνύουν τις «υπογραφές» που προκύπτουν από τον υπολογισμό του μετασχηματισμού Radon σε γωνία 90° για τις παραπάνω περιπτώσεις. Παρατηρώντας τις Εικόνες A5.1(β), A5.2(β) και A5.3(β) διακρίνουμε κάποια σημαντικά στοιχεία που τεκμηριώνουν την αξία της χρονικής εξέλιξης όσον αφορά τον χαρακτηρισμό των εκφράσεων. Η έκφραση «χαρά» έχει μια σχετικά συμμετρική χρονική εξέλιξη και εμφανίζει δραστηριότητα γύρω από την περιοχή του στόματος. Η έκφραση «έκπληξη» είναι πιο κοφτή, το μεγαλύτερο ποσοστό της δραστηριότητας εκφράζεται στα πρώτα στάδια της έκφρασης. Εμφανίζει επίσης κατανομημένη ως προς το χώρο του προσώπου δραστηριότητα –περιοχές γύρω από τα μάτια, τα φρύδια και το στόμα. Η έκφραση «οργή» εκφράζεται μέσα από δύο ουσιαστικά περιοχές: γύρω από το στόμα και στη περιοχή των φρυδιών και ανάμεσα σε αυτά. Το αξιοσημείωτο είναι ότι η δραστηριότητα στις περιοχές αυτές εμφανίζεται σε διαφορετικές χρονικές στιγμές: Προηγείται ενεργοποίηση των μυών που βρίσκονται στη περιοχή των φρυδιών και ακολουθεί αυτή των μυών που βρίσκονται γύρω από το στόμα. Οι «υπογραφές» σε συνδυασμό με την ενεργειακή κατανομή των διανυσμάτων κίνησης σε οκτώ διαφορετικές κατευθύνσεις –βλέπε Κεφάλαιο Δ4, Σχήμα Δ4.1- τροφοδοτούν ένα ταξινομητή HMM απαριζόμενο από έξι κλάδους –ένα για κάθε πρωτεύουσα έκφραση- ο οποίος πραγματοποιεί την απαιτούμενη χρονική ανάλυση λαμβάνοντας υπόψη και τις μεταβολές στην κλίμακα του χρόνου [97].

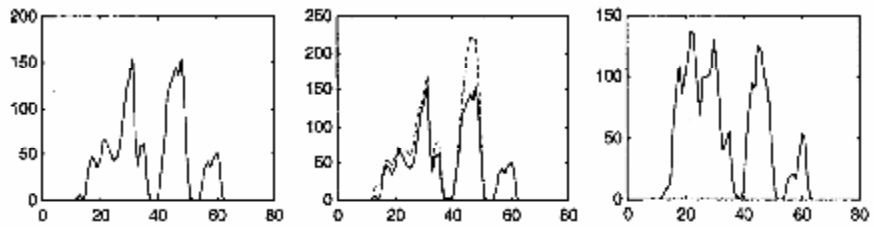
Τα αποτελέσματα της ταξινόμησης που προέκυψαν από το δικτύωμα HMMs προβάλλονται στο χώρο ενεργοποίησης – επαλήθευσης –βλέπε Κεφάλαιο Δ3. Το κρίσιμο ερώτημα είναι κατά πόσο οι μεταβλητές που χρησιμοποιεί ο ταξινομητής HMM σχετίζονται με κάποιο χώρο χαμηλότερης διάστασης, όπως αυτός των αξόνων ενεργοποίησης – επαλήθευσης, παρά με συγκεκριμένες κατηγορίες. Η Εικόνα A5.4 συνηγορεί στη, μέχρι ενός σημείου, καταφατική απάντηση στο προηγούμενο ερώτημα: Η πλειοψηφία των αποτυχημένων ταξινομήσεων παραμένει στο ίδιο τεταρτημόριο με την πραγματική έκφραση –εξαιρέση αποτελούν οι αποτυχημένες ταξινομήσεις της έκφρασης «χαρά». Το συμπέρασμα είναι ότι ακόμη και οι αποτυχημένες ταξινομήσεις διατηρούν μέρος της πληροφορίας σχετικά με την πραγματική συναισθηματική κατάσταση, η οποία άλλωστε με αφαιρετικό τρόπο προσδιορίζεται με μία και μόνη λεκτική έννοια.



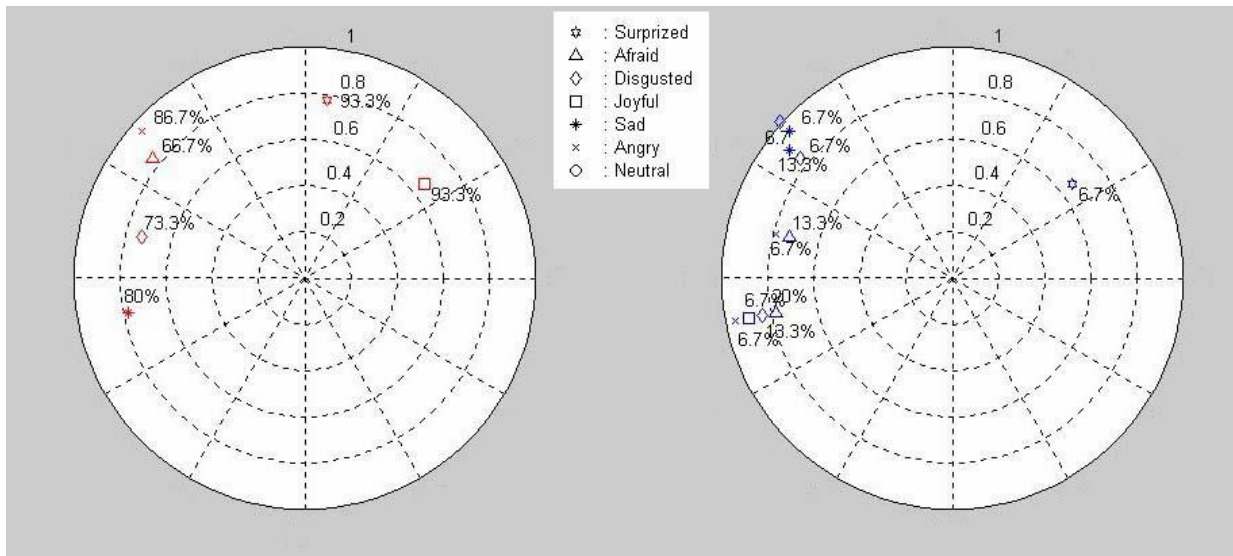
Εικόνα A5.1: Έκφραση “ΟΡΓΗ” (α) Καρέ 01, κίνηση ανάμεσα στα καρέ 01 και 03, 03 και ‘apex’, ‘apex’ και ‘release’, (β) Ο μετασχηματισμός Radon, του μέτρου των διανυσμάτων κίνησης, σε γωνία 90^0 ανάμεσα στα καρέ 01 και 03, 01 και ‘apex’ (dotted line), ‘apex’ και ‘release’.



Εικόνα A5.2: Έκφραση “ΧΑΡΑ” (α) Καρέ 01, κίνηση ανάμεσα στα καρέ 01 και 05, 05 και 09, 09 και ‘apex’, ‘apex’ και ‘release’, (β) Ο μετασχηματισμός Radon, του μέτρου των διανυσμάτων κίνησης, σε γωνία 90^0 ανάμεσα στα καρέ 01 και 05 (dashed line), 01 και 09, 01 και ‘apex’ (dotted line).



Εικόνα Α5.3: Έκφραση “ΕΚΠΛΗΞΗ” (α) Καρέ 01, κίνηση ανάμεσα στα καρέ 01 και 04, 04 και ‘apex’, ‘apex’ και ‘release’, (β) Ο μετασχηματισμός Radon, του μέτρου των διανυσμάτων κίνησης, σε γωνία 90° ανάμεσα στα καρέ 01 και 04, 01 και ‘apex’ (dotted line), ‘apex’ και ‘release’.



Εικόνα Α5.4: (α) Αποτελέσματα ορθής και (β) λανθασμένης ταξινόμησης με χρήση HMM, μετά από προβολή στο χώρο activation - evaluation.

BIBΛΙΟΓΡΑΦΙΑ

- [1]. P. Chellapa, C. Wilson and S. Sirohey, "Human and Machine Recognition of Faces: A Survey," *Proc. IEEE*, vol. 83, no. 5, pp. 705-740, 1995.
- [2]. J.C. Bartlett, S. Hurry and W. Thorley, "Typicality and familiarity of faces," *Memory Cognition* **12**, pp. 219-228, 1984.
- [3]. L.D.Harmon, "The recognition of faces," *Scientific America*. **229**, pp.71-82, Oct. 1973
- [4]. R.A.Campell, S. Cannon, G. Jones and N. Morgan, "Individual face identification by computer vision," *Proc. Conf. Modeling Simulation Microcomputer*, pp. 62-63, 1987.
- [5]. Sir F. Galton, "Personal identification and description - I," *Nature*, pp.173-177, June 1888.
- [6]. Sir F. Galton, "Numeralized profile for classification and recognition," *Nature* **83**, pp.127-130, March 1910.
- [7]. V. Govindaraju, D.B.Sher, R.K.Srihari and S.N.Srihari, "Locating of human faces in newspapers photographs," *Proc. of CVBR*, pp. 549-554, 1989.
- [8]. V. Govindaraju, S.N.Srihari and D.B.Sher, "A computational model for face location," *Proc. of 3rd Int. Conf. on Computer Vision*, pp. 718-721, 1990.
- [9]. T. Sakai, M. Nagao and S. Fujibayshi, "Line extraction and pattern recognition in a photograph," *Pattern Recognition* **1**, pp. 233-248, 1969.
- [10]. M. Kelly, "Visual identification of people by computer," *Technical Report AI-130*, Stanford, CA, 1970
- [11]. I. Craw, H . Ellis and J.R.Lishman, "Automatic extraction of face-features," *Pattern Recognition Letters* **5**, pp.183-187, 1987.
- [12]. I. Craw, D. Tock and A. Benett, "Finding face features," *Proc. of 2nd European Conference on Computer Vision*, pp.92-96, 1992.
- [13]. J. Sepherd, "An interactive computer system for retrieving faces," Ellis (Eds), Dordrecht: Nijhoff, pp. 398-409, 1985.
- [14]. S. Sirohey, "Human face segmentation and identification," *Technical Report CAR-TR-695*, University of Maryland, MD, 1993.
- [15]. Sir F. Galton, "Personal identification and description - II," *Nature*, pp.201-203, June 1888.
- [16]. J.L.Perry and J.M.Carney, "Human face recognition using a multilayer perceptron," *Proc. of Int. Conf. on Neural Networks II*, pp. 413, January 1990.
- [17]. A.L.Yulie, D.S.Cohen and P.W.Hallinan, "Feature extraction of faces using deformable templates," *Proc. of CVBR*, pp. 104-109, 1989.
- [18]. M. Nixon, "Eye spacing measurements for facial recognition," *SPIE Proc.* **575**, *Applications of Digital Image Processing VIII*, pp. 279-285, 1985.
- [19]. R. Buhr, "Analyse und klassifikation von gesichtsbildern," *ntzArchiv* **8**, pp. 245-256, 1986.
- [20]. L.D.Harmon, "Automatic recognition of human face profiles," *Proc. of 3rd Int. Joint Conf. on Pattern Recognition*, pp. 183-188, 1976.
- [21]. L.D.Harmon, S.C.Kuo, P.F.Ramig and U. Raudkivi, "Identification of human face profile by computers," *Pattern Recognition* **10**, pp. 301-312, 1978.
- [22]. L.D.Harmon, M.K.Khan, R. Lasch and P.F.Ramig, "Machine identification of human faces," *Pattern Recognition* **13**, pp. 97-110, 1981.
- [23]. G.J.Kaufman K.J.Breeding, "The automatic recognition of human faces from profile silhouettes," *IEEE Trans. on System Man and Cybernetics* **6**, pp. 113-121, 1976.
- [24]. A.J.Goldstein, L.D.Harmon and A.B.Lesk, "A basic study on human face recognition," *Proc. of IEEE* **59**, pp. 748-760, May 1971.

- [25]. Y. Kaya and K. Kobayashi, "Identification of human faces," *Frontiers of Pattern Recognition*, pp. 265-289, Academic Press, New York 1971.
- [26]. K.H.Wong, H.M.Law and P.W.M.Tsang, "A system for recognizing human faces," *Proc. of ICASSP*, pp. 1638-1642, 1989.
- [27]. R.J.Baron, "Mechanisms of human facial recognition," *Int. Journal of Man Machine Studies* **15**, pp. 137-178, January 1990.
- [28]. K. Preston Jr, "Computing at the speed of light," *Electronics* **38**, pp. 72-83, 1965.
- [29]. W.K.Taylor, "Machine learning and recognition of faces," *Electronic Letters* **3**, pp. 436-437, 1967.
- [30]. H. Shlosberg, "Three dimensions of emotion," *Psychol. Rev* **61**, pp. 81-88, 1954.
- [31]. P. Ekman and W.V.Friesen, "Manual for the Facial Action Coding System," *Consulting Psychologists Press*, Palo Alto, 1978.
- [32]. C. Darwin, *The Expression of Emotions in Man and Animals*, John Murray, 1872, reprinted by University of Chicago Press, 1965.
- [33]. P. Ekman, *Darwin and Facial Expressions*, Academic Press, 1973.
- [34]. M. Davis and H. College, *Recognition of Facial Expressions*, Arno Press, New York, 1975.
- [35]. K. Scherer and P. Ekman, *Approaches to Emotion*, Lawrence Erlbaum Associates, 1984.
- [36]. P. Ekman, T. Huang, T. Sejnowski and J. Hager, NSF Planning Workshop on Facial Expression Understanding, *Technical report, National Science Foundation*, Human Interaction Lab, 1992.
- [37]. W. E. Rinn, 'The Neuropsychology of Facial Expression: A Review of Neurological and Psychological Mechanisms for Producing Facial Expressions', *Psychological Bulletin*, vol. 95, pp. 52-77, 1984.
- [38]. P. Ekman and W. Friesen, *Unmasking the Face*, Prentice-Hall, 1975.
- [39]. S. Shibui, H. Yamada, T. Sato and K. Shigemasu, "Categorical Perception and Semantic Information Processing of Facial Expressions," *Perception*, vol. 28 S, p. 114, 1999.
- [40]. J. N. Bassili, "Emotion recognition: The role of facial movement and the relative importance of upper and lower areas of the face," *Journal of Personality and Social Psychology*, 37, 2049-2059, 1979.
- [41]. M. Bartlett, J. Hager, P. Ekman and T. Sejnowski, "Measuring Facial Expressions by Computer Image Analysis," *Psychophysiology*, vol. 36, pp. 253-263, 1999.
- [42]. I. Essa and A. Pentland, "A Vision System for Observing and Extracting Facial Action Parameters," *Proc. Intern. Conf. on CVPR*, pp. 76-83, 1994.
- [43]. D. Terzopoulos and K. Waters, "Analysis and Synthesis of Facial Image Sequences using Physical and Anatomical Models," *IEEE Trans. on PAMI*, Vol. 15, No. 6, 569-579, 1993.
- [44]. Li, Roivainen and Forcheimer, "3-D motion estimation in model-based facial image coding," *IEEE Trans. on PAMI*, Vol. 15, 545-555, 1993.
- [45]. K. Mase, "Recognition of facial expression from optical flow," *IEICE Trans.*, vol. E74, pp. 3474-3483, 1991.
- [46]. C. Pelachaud, N. Badler and M. Viaud, "Final Report to NSF of the Standards for Facial Animation Workshop," *Technical report*, NSF, University of Pennsylvania, Philadelphia, PA 19104-6389, 1994.
- [47]. P. Ekman, T. Huang, T. Sejnowski and J. Hager (Eds), Final Report to NSF of the Planning Workshop on Facial Expression Understanding, *Technical report*, NSF, Human Interaction Lab., UCSF, CA 94143, 1993.
- [48]. P. Ekman, J. Hager and E. Rosenberg, "ACSAID: A Computer DataBase for Predicting Affective Phenomena from Facial Movement," <http://www.nirc.com/facsaid.html>.
- [49]. *ISO/IEC JTC1/SC29/WG11 MPEG96/N1365*, "MPEG4 SNHC: Face and Body Definition and Animation Parameters," 1996.
- [50]. R. Cabeza and J. Nyburg, *Cognitive Neuroscience* **9**, 1-26, 1997.
- [51]. V. Bruce and A. Young, *In the Eye of the Beholder: The science of face perception*, Oxford Univ. Press, 1998.

- [52]. Phillips et al., *Nature* 389, 495-8, 1997.
- [53]. S. Scalaidhe et al., *Science* 278, 1135-08, 1997.
- [54]. S. McKenna and S. Gong, "Tracking Faces," *Proc. of the 2nd Intl. Conf. on Automatic Face and Gesture Recognition*, pp. 271-276, 1996.
- [55]. J. Crowley and F. Berard, "Multi-Modal Tracking of Faces for Video Communications," *Proc. of IEEE CVPR*, pp. 640-645, Puerto Rico, 17-19 June 1997.
- [56]. H. Graf, E. Cosatto, D. Gibbon, M. Kocheisen, and E. Petajan, "Multi-modal System for Locating Heads & Faces," *Proc. Int. Conf. on Automatic Face & Gesture Recognition*, pp. 88-93, Vermont, USA, Oct. 1996.
- [57]. M. Collobert et al, "Listen: A System for Locating and Tracking Individual Speakers," *Proc. of the Intl. Conf. on Automatic Face and Gesture Recognition*, pp. 283-288, Vermont, October 1996.
- [58]. M. Hunke and A. Waibel, "Face Locating and Tracking for Human Computer Interaction," *IEEE Computers*, pp. 1277-1281, November 1994.
- [59]. S. Basu, I. Essa and A. Pentland, "Motion Regulization for Model-based Head Tracking," *Proc. 13th Intern. Confer. On Pattern Recognition*, August 1996.
- [60]. P. Fieguth and D. Terzopoulos, "Color-based Tracking of Image Regions with Changes in Geometry and Illumination," *Proc. of IEEE CVPR*, pp. 403-410, 1996.
- [61]. J. Terillon, M. David and S. Akamatsu, "Automatic Face Detection in Natural Scene Images Using a Skin Color Model & Moments," *Proc. Int. Conf. Automatic Face & Gesture Recognition*, Nara, Japan, April 1998.
- [62]. Y. Raja, S. Mckenna and S. Gong, "Tracking and Segmenting People in Varying Lighting Conditions Using Color," *Proc. Int. Conf. Automatic Face and Gesture Recognition*, pp. 228-233, Nara, Japan, April 1998.
- [63]. T. Maurer and C.V.D. Malsburg, "Tracking and Learning Graphs on Image Sequences of Faces," *Proc. of the Intl. Conf. on Automatic Face and Gesture Recognition*, pp. 176-181, Vermont, USA, Oct. 1996.
- [64]. V. Kruger and G. Sommer, "Affine Face Tracking Using a Wavelet Network," *Proc. Intl. Workshop on Recognition, Analysis and Tracking of Faces and Gestures in Real-time Systems*, Corfu, Greece, Sept. 1999.
- [65]. N. Arad and D. Reisfeld, "Image Warping using few Anchor Points and Radial Functions," *Computer Graphics Forum*, vol 14 (1), 35-46, 1994.
- [66]. S. Gong, S. McKenna and S. Collins, "An Investigation into Face Pose Distributions," *Proc. of the Intl. Conf. on Automatic Face and Gesture Recognition*, pp. 265-270, Vermont, 1996.
- [67]. A. Zelinsky and J. Heinzmaan, "Real-time Visual Recognition of Facial Gestures for HCI," *Proc. of the Intl. Conf. on Automatic Face and Gesture Recognition*, pp. 351-356, October 1996.
- [68]. A. Young and H. Ellis, *Handbook of Research on Face Processing*, Elsevier Science Publishers, 1989.
- [69]. C. Padgett and G. W. Cottrell, "Representing Face Images for Emotion Classification," *Advances in Neural Information Processing Systems*, Vol. 9, p. 894, The MIT Press, 1997.
- [70]. C. Padgett, G. Cottrell and B. Adolps, "Categorical Perception in Facial Emotion Classification," *Proc. of the Cognitive Science Conference*, vol. 18, pp. 249-253, 1996.
- [71]. B. Horn and B. Schunk, "Determining optical flow," *Artificial Intelligence*, Vol.17, 185-203, 1981.
- [72]. D. Heeger, "Optical flow using spatiotemporal filters," *Intern. Journal of Computer Vision* 1, 279-302, 1988.
- [73]. M. Abdel-Mottaleb, R. Chellappa and A. Rosenfeld, "Binocular motion stereo using MAP estimation," *IEEE Conference on Computer Vision and Pattern Recognition*, 321-327, 1993.
- [74]. P. Anandan, "A computational network and an algorithm for the measurement of visual motion," *International Journal of Computer Vision* 2, 283-310, 1989.
- [75]. Y. Yacoob and L. Davis, "Computing spatio-temporal representations of human faces," *Proc. of the Computer Vision and Pattern Recognition Conference*, 70-75, 1994.
- [76]. N. Tsapatsoulis, I. Avrithis and S. Kollias, "On the use of Radon Transform for Facial Expression Recognition," *Proc. of the 5th Intl. Conf. on Information Systems Analysis and Synthesis*, Orlando, U.S.A, July 1999.
- [77]. M. Black and P. Anandan, "The Robust Estimation of optical flow," *Proc. Int. Conf. on Computer Vision*, Berlin, Germany, 231-236, 1993.

- [78]. Y. T. Wu, T. Kanade, J. Cohn and C-C. Li, "Optical Flow Estimation Using Wavelet Motion Model," 1997.
- [79]. T. Sakaguchi, J. Ohya and F. Kishino, "Facial expression recognition from image sequence using Hidden Markov Model," *VLBV* 95, A-5, 1995.
- [80]. T. Otsuka and J. Ohya, "Recognition of facial expressions using HMM with continuous output probabilities," *Proc. of 5th IEEE International Workshop on Robot and Human Communication RO-MAN*, 323-328, 1996.
- [81]. T. Otsuka and J. Ohya, "Recognizing Multiple Persons' Facial Expressions using HMM based on Automatic Extraction of Καρές from Image Sequences," *Proc. IEEE Int. Conf. on Image Proc.*, vol. 2, 546-549, 1997.
- [82]. J. Lien, T. Kanade, J. Cohn and C-C. Li, 'Subtly Different Facial Expression Recognition and Emotion Expression Intensity Estimation', *Proc. of IEEE CVPR* pp. 853-859, Santa Barbara, CA, 1998.
- [83]. B. Lucas and T. Kanade, "An Iterative Image Registration technique with an Application to Stereo Vision," *Proc. of the 7th Intl. Joint Conf. on AI*, 1981.
- [84]. J. Cohn, A. Zlochow, J. Lien and T. Kanade, "Automated Face Analysis by Feature Point Tracking Has High Concurrent Validity with Manual FACS Coding," *Psychophysiology*, vol. 26, pp. 35-43, 1999.
- [85]. H. Wu et al, 'Face and Facial Feature Extraction from Color Images', *Proc. of the Intl. Conf. on Automatic Face and Gesture Recognition*, pp. 345-350, October 1996.
- [86]. R. Herpes et al, 'Edge and Keypoint Detection in Facial Regions', *Proc. of the Intl. Conf. on Automatic Face and Gesture Recognition*, pp. 212-217, Vermont, USA, October 1996.
- [87]. R. Herpes et al, "An Attentional Processing Strategy to Detect and Analyse the Prominent Facial Regions," *Proc. of the Intl. Conf. on Automatic Face and Gesture Recognition*, Zurich, Switzerland, pp. 214-220, 1995.
- [88]. M. Black, Y. Yacoob, A. Jepson and D. Fleet, "Learning Parameterized Models of Image Motion," *Proc. of IEEE CVPR*, pp. 561-567, 1997.
- [89]. Y. Yacoob and L. Davis, "Recognizing human facial expressions," *The Second Workshop on Visual Form*, Capri, 584-593, 1994.
- [90]. Y. Yacoob and L. Davis, "Recognizing human facial expressions from long image sequences using optical flow," *IEEE Transactions on Pattern Analysis and Machine Intelligence* 18(6), 636-642, 1996.
- [91]. M. Rosenblum, Y. Yacoob and L. Davis, "Human Emotion Recognition from Motion Using a Radial Basis Function Network Architecture," *IEEE Trans. On NNs* vol. 7, no 5, 1996.
- [92]. N. Thalmann, P. Kalra and M. Escher, "Face to Virtual Face," *Proc. of the IEEE*, vol.86, pp.870-883, 1998.
- [93]. I. Essa and A. Pentland, "Coding, Analysis, Interpretation and Recognition of Facial Expressions," *Technical Report No. 325*, M.I.T. Media Laboratory, 1995.
- [94]. I. Essa, T. Darrell and A. Pentland, "Tracking facial motion," *Proceedings of the Workshop on Motion of Nonrigid and Articulated Objects*, 36-42, IEEE Computer Society, 1994.
- [95]. I. Essa, S. Sclaroff and A. Pentland, "Physically-based modeling for graphics and vision," in R. Martin (Ed) *Directions in Geometric Computing*. Information Geometers, U.K., 1993.
- [96]. S. Platt and N. Badler, "Animating facial expression," *Proc. of ACM SIGGRAPH Conference* 15(3):245-252, 1981.
- [97]. N. Tsapatsoulis, M. Leonidou and S. Kollias, "Facial Expression Recognition Using HMM with Observation Dependent Transition Matrix," *Proc. of MMSP'98*, Portofino CA, December 1998.

ΕΝΤΟΠΙΣΜΟΣ ΠΡΟΣΩΠΟΥ ΚΑΙ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ ΣΗΜΕΙΩΝ ΤΟΥ ΣΕ ΜΟΝΟΧΡΩΜΕΣ ΕΙΚΟΝΕΣ

Η παρούσα διατριβή έχει σκοπό να προτείνει λύσεις σε κάποια από τα βασικά θέματα της αναγνώρισης προσώπων και της ανάλυσης εκφράσεων στο πλαίσιο της επικοινωνίας ανθρώπου μηχανής. Όπως θα έχει γίνει αντιληπτό από το πρώτο μέρος της διατριβής ο εντοπισμός του προσώπου και κάποιων χαρακτηριστικών σημείων σε αυτό είναι ένα κυρίαρχο θέμα και στις δύο εφαρμογές. Στο τρέχον κεφάλαιο περιγράφεται μια διαδικασία εντοπισμού προσώπων από μονόχρωμες εικόνες. Οι περιορισμοί που τίθενται συνάδουν με τις συνθήκες υπό τις οποίες πραγματοποιείται η επικοινωνία ανθρώπου μηχανής: το πρόσωπο θεωρείται ότι βρίσκεται σε κοντινή απόσταση από την κάμερα και δεν εμφανίζει υπερβολική στροφή ως προς τον διαμήκη άξονα του σώματος – άξονας y στο Σχήμα A4.1. Πέρα από τον εντοπισμό του προσώπου παρουσιάζεται μια μεθοδολογία που αφορά στον εντοπισμό της ίριδας των ματιών, των ρουθουνιών και του κέντρου του στόματος. Επιπλέον δίνονται κάποιες κατευθύνσεις όσον αφορά τον εντοπισμό και άλλων σημείων όπως οι κόγχες των ματιών και οι γωνίες του στόματος. Τέλος παρουσιάζεται μια διαδικασία για τον εντοπισμό των σημείων στην οριογραμμή εικόνων προφίλ. Παρόλο που η αξία των σημείων του προφίλ όσον αφορά στην επικοινωνία ανθρώπου μηχανής είναι μάλλον περιορισμένη, η εν λόγω τεχνική χρησιμοποιήθηκε για την εξαγωγή πληροφορίας σχετικής με την ανατομία του κεφαλιού η οποία μπορεί να ενταχθεί στη διαδικασία σύνθεσης τρισδιάστατων μοντέλων για το πρόσωπο.

1. ΕΝΤΟΠΙΣΜΟΣ ΠΕΡΙΓΡΑΜΜΑΤΟΣ ΠΡΟΣΩΠΟΥ

Για τον εντοπισμό του προσώπου στην εικόνα κλίμακας του γκρι υιοθετήθηκε η τεχνική της ανίχνευσης του περιγράμματός του. Η επιλογή αυτή εμφανίζει υψηλό βαθμό επιτυχίας και αξιοπιστίας όταν το απεικονιζόμενο πρόσωπο βρίσκεται κοντά στην κάμερα, υπόθεση που δεν είναι ιδιαίτερα περιοριστική σε περιβάλλοντα επικοινωνίας ανθρώπου μηχανής. Εναλλακτικές τεχνικές εντοπισμού του προσώπου, όπως η ταύτιση με πρότυπο θα μπορούσαν επίσης να χρησιμοποιηθούν, αλλά εμφανίζουν μικρότερη ακρίβεια όσον αφορά στον εντοπισμό και θέτουν επιπλέον περιορισμούς. Αντίστοιχα, τεχνικές ενεργών περιγραμμάτων (active contours) έχουν την απαιτούμενη ακρίβεια, αλλά χρειάζονται μια καμπύλη αρχικοποίησης. Υπό αυτό το πρίσμα, η τεχνική που παρουσιάζεται στις επόμενες παραγράφους, θα μπορούσε να χρησιμοποιηθεί για την αρχικοποίηση του ενεργού περιγράμματος το οποίο στη συνέχεια μπορεί να προσαρμοστεί τέλεια στην οριογραμμή του προσώπου.

Έστω I η εικόνα προσώπου αποχρώσεων του γκρι και χωρικής ανάλυσης ixc . Με βάση την εικόνα I υπολογίζουμε τις εικόνες H και L οι οποίες αντιστοιχούν στις εικόνες τοπικών μεγίστων και ελαχίστων της I σε ένα παράθυρο $(2n+1) \times (2n+1)$, όπου n φυσικός αριθμός (τυπική επιλογή $n=1$).

Έστω:

$$N(t_{ij}) = \{t_{i-k, j-l} \mid k, l = -n, \dots, n\} \quad (B1.1)$$

η $(2n+1) \times (2n+1)$ γειτονιά του pixel $t_{ij}=(i,j)$. Η εικόνα H των τοπικών μεγίστων υπολογίζεται από την σχέση:

$$H(i, j) = \max_{p_{ij} \in N} [I(N)] \quad (\text{B1.2})$$

Ομοίως η εικόνα $L(i,j)$ των τοπικών ελαχίστων υπολογίζεται από την σχέση:

$$L(i, j) = \min_{p_{ij} \in N} [I(N)] \quad (\text{B1.3})$$



Εικόνα B1.1: (α) Αρχική Εικόνα (β) Εικόνα τοπικών μεγίστων (γ) Εικόνα τοπικών ελαχίστων (δ) Εικόνα διαφορών (τοπικά μέγιστα – τοπικά ελάχιστα)

Η εικόνα της διαφοράς $D=H-L$ τονίζει τα σημεία εκείνα της εικόνας στα οποία έχουμε απότομες εναλλαγές στις τιμές φωτεινότητας της εικόνας. Οι περιοχές των ματιών, του στόματος, της μύτης και του περιγράμματος αντιστοιχούν σε σημεία της εικόνας D με υψηλή τιμή. Η Εικόνα B1.1 δίνει ένα χαρακτηριστικό παράδειγμα της διαδικασίας.

Υπολογίζοντας τις διαφορές ανάμεσα στα μέγιστα και στα ελάχιστα οι οποίες υπερβαίνουν ένα συγκεκριμένο κατώφλι, μπορούμε να εντοπίσουμε τις ακμές της εικόνας. Στην συνέχεια, από τα pixels που έχουμε εντοπίσει πως ανήκουν σε ακμή, αναζητάμε αυτά που βρίσκονται πάνω σε έλλειψη προσπαθώντας με αυτόν τον τρόπο να διαμορφώσουμε το περίγραμμα του προσώπου (έχοντας κάνει την υπόθεση πως το περίγραμμα του προσώπου ανήκει σε έλλειψη). Στην πράξη η περιοχή στην οποία ψάχνουμε τα pixels δεν είναι απλά η περιφέρεια μιας έλλειψης αλλά ουσιαστικά μια ελλειψοειδής σαμπρέλα. Συγκεκριμένα, έστω B είναι η εικόνα των ακμών η οποία προήλθε από κατωφλίωση της εικόνας D με κάποιο στατιστικό κατώφλι. Τυπική επιλογή κατωφλίου $c=m+\sigma$ με:

$$m = \frac{1}{M \cdot N} \sum_{i=1}^M \sum_{j=1}^N D(i, j) \quad (\text{B1.4})$$

$$\sigma = \frac{1}{M \cdot N} \sum_{i=1}^M \sum_{j=1}^N |D(i, j) - m| \quad (\text{B1.5})$$

Τότε η επιλογή της έλλειψης η οποία προσεγγίζει το περίγραμμα του προσώπου δίνεται από την σχέση:

$$[X_0, Y_0, A, B, \Theta] = \arg \max_{x_0, y_0, a, b, \theta} E[x_0, y_0, a, b, \theta] \quad (\text{B1.6})$$

όπου X_0, Y_0 το κέντρο της έλλειψης, A, B οι ημιάξονές της και Θ η κλίση της –στροφή ως προς τον άξονα \mathbf{z} του Σχήματος A4.1, και

$$E[x_0, y_0, a, b, \theta] = \sum_{p_{ij} \in S} B(t_{ij}) \quad (\text{B1.7})$$

όπου

$$S[x_0, y_0, a, b, \theta] = \left\{ t_{ij} : 0.95 \leq \frac{(i-x_0)^2}{a^2} + \frac{(j-y_0)^2}{b^2} \leq 1.05 \right\} \quad (\text{B1.8})$$

εκφράζει την επιφάνεια της ελλειψοειδούς συμπτρέλας με κέντρο (x_0, y_0) κλίση θ και ημιάξονες a, b .

$$\begin{bmatrix} i \\ j \end{bmatrix} = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} \quad (\text{B1.9})$$

Η ελλειψοειδής συμπτρέλα πρέπει να έχει κάποιο μετρήσιμο εμβαδόν έτσι ώστε να αποφύγουμε την επιλογή μιας έλλειψης με άκρα τα μάτια και το στόμα, περιοχές που και αυτές αντιστοιχούν σε πολύ έντονες ακμές. Εναλλακτικά η ίδια τεχνική μπορεί να εφαρμοστεί άμεσα στην εικόνα D –χωρίς κατωφλίωση- χρησιμοποιώντας τη σχέση:

$$E[x_0, y_0, a, b, \theta] = \sum_{p_{ij} \in S} D(t_{ij}) \quad (\text{B1.10})$$

Το μειονέκτημα στην περίπτωση αυτή είναι ότι πρέπει να εξεταστούν όλα τα pixels της εικόνας και όχι μόνο αυτά που ανήκουν σε ακμές, γεγονός που οδηγεί σε πολλαπλασιασμό της υπολογιστικής πολυπλοκότητας.

Η τεχνική είναι περισσότερο αποτελεσματική όταν οι ακμές έχουν κάποιο πάχος, σημαντικό. Το μειονέκτημα της διαδικασίας εντοπισμού του περιγράμματος όπως δόθηκε ανωτέρω, είναι η μεγάλη υπολογιστική πολυπλοκότητα η οποία προκύπτει από τον 5-Δ χώρο των παραμέτρων. Βέβαια, δεδομένης της ανατομίας του προσώπου μπορούν να τεθούν περιορισμοί όσον αφορά τη σχέση των ημιάξονων. Μια λογική επιλογή είναι η θεώρηση μόνο εκείνων των ελλείψεων με λόγο ημιάξονων a/b που κυμαίνεται στο διάστημα [1.1 1.6]. Επιπλέον εφαρμόζοντας τον διακριτό μετασχηματισμό KL στις συντεταγμένες των pixels που αντιστοιχούν σε ακμές –βλέπε Κεφάλαιο B2- μπορούμε να εκτιμήσουμε τους βασικούς άξονες της κατανομής τους και στη συνέχεια να υπολογίσουμε χονδρικά τη γωνία θ .

Η μοντελοποίηση του περιγράμματος του προσώπου με έλλειψη παρουσιάζει υψηλή ευρωστία ως προς τις διακυμάνσεις στροφής του προσώπου περί τους άξονες \mathbf{x} και \mathbf{y} . Στροφή περί τον άξονα \mathbf{x} μεταβάλλει τις αποστάσεις ανάμεσα στα χαρακτηριστικά του προσώπου αλλά δεν περιστρέφει την έλλειψη, το αποτέλεσμα

είναι απλά μείωση του λόγου a/b . Στροφή περί τον άξονα y αυξάνει το λόγο a/b χωρίς όμως να αλλάζει το σχήμα του προσώπου το οποίο παραμένει ελλειψοειδές. Στην Εικόνα B1.3 επιδεικνύεται η αποτελεσματική κατάτμηση του προσώπου σε μη ομοιογενές φόντο και υπό σημαντική στροφή περί τον άξονα y . Η στροφή ως προς τον άξονα z λαμβάνεται υπόψη μέσω της γωνίας θ . Πρέπει επίσης να τονιστεί ότι μερική επικάλυψη του περιγράμματος δεν οδηγεί σε αποτυχία την τεχνική δεδομένου ότι αρκεί ένα μέρος του για να διαμορφώσει κάποια επιμέρους τόξα που την απαρτίζουν. Το ίδιο ισχύει και για μη οβάλ πρόσωπα όπως αυτό της Εικόνας B1.2. Στην ίδια εικόνα φαίνεται και η αποτυχία σχηματισμού του περιγράμματος στη περιοχή του πηγουνιού. Παρόλα αυτά ο εντοπισμός του προσώπου είναι σχεδόν τέλειος.



(α)



(β)



(γ)

Εικόνα B1.2: Εντοπισμός και κατάτμηση προσώπου με ανίχνευση των pixels που ικανοποιούν την εξίσωση της έλλειψης.



(α)



(β)



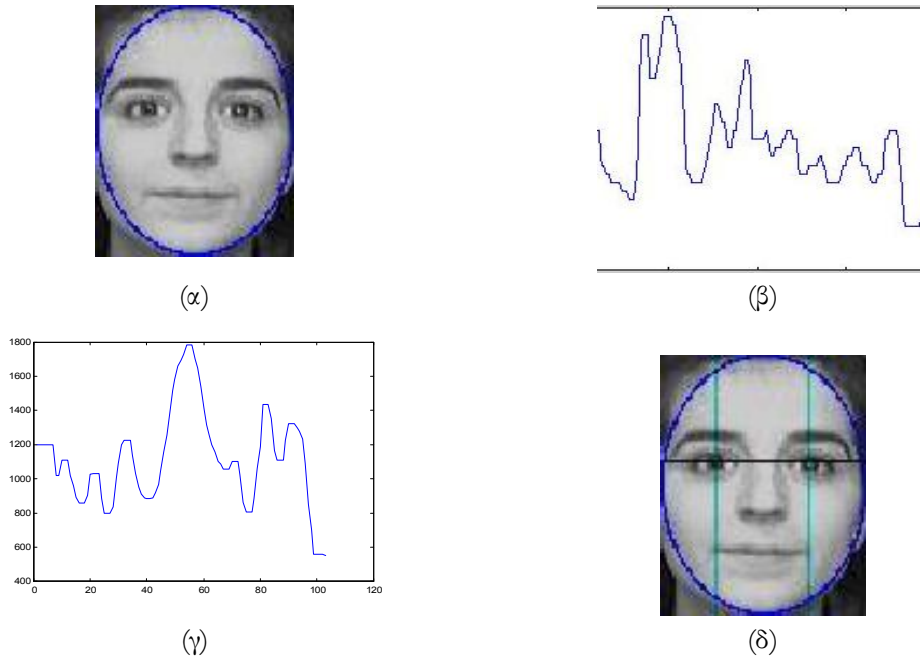
(γ)

Εικόνα B1.3: Κατάτμηση από ανομοιογενές φόντο και υπό περιστροφή

2. ΕΝΤΟΠΙΣΜΟΣ ΣΗΜΕΙΩΝ ΣΕ ΕΙΚΟΝΕΣ ΕΜΠΡΟΣΘΙΑΣ ΟΨΗΣ

Ο εντοπισμός των χαρακτηριστικών σημείων της εμπρόσθιας όψης που περιγράφεται στη συνέχεια εφαρμόζεται αφού έχει προηγηθεί κατάτμηση του προσώπου από το φόντο με τη βοήθεια της τεχνικής που περιγράφηκε στη προηγούμενη παράγραφο. Επίσης με βάση την εκτίμηση της γωνίας στροφής θ της έλλειψης, πραγματοποιείται το σχετικό αντιστάθμισμα ώστε το πρόσωπο να έλθει σε πλήρως όρθια θέση. Η τεχνική μπορεί να εφαρμοστεί χωρίς καμία προεργασία σε περιπτώσεις mug shot φωτογραφιών στις οποίες το φόντο είναι ομοιογενές και τα απεικονιζόμενα πρόσωπα εμφανίζουν μικρή κλίση ως προς τον άξονα z .

Η προτεινόμενη τεχνική στηρίζεται στο γεγονός ότι η ανομοιογένεια στην κατανομή των τιμών της κλίμακας του γκρι στις περιοχές των φρυδιών και ματιών είναι πυκνότερη από κάθε άλλη περιοχή του προσώπου. Με προβολή της εικόνας ακμών –εικόνα B της προηγούμενης παραγράφου- στον κατακόρυφο άξονα δημιουργούνται διάφορες κορυφές που αντιστοιχούν στα φρύδια, τα μάτια, τη μύτη, το στόμα και – βλέπε Εικόνα B1.4(β). Με δεδομένη την ανατομία του προσώπου, η σειρά των κορυφών θα είναι όπως παραπάνω με υψηλότερη αυτήν που αντιστοιχεί στα μάτια ή τα φρύδια.



Εικόνα B1.4: (α) Εντοπισμός του περιγράμματος του προσώπου, (β) προβολή των τιμών της εικόνας ακμών στον κατακόρυφο άξονα, (γ) άθροιση των τιμών των pixels σε μια λωρίδα μικρού πάχους η οποία περιλαμβάνει την ευθεία που διέρχεται από τα μάτια, (δ) εντοπισμός ματιών

Στην πράξη αντί για προβολή της εικόνας ακμών στο κατακόρυφο άξονα προβάλλουμε την εικόνα διαφορών D . Με τον τρόπο αυτό τονίζονται περισσότερο οι περιοχές ανομοιογένειας στο πρόσωπο. Στην εικόνα ακμών όλα τα pixels μιας ακμής έχουν την ίδια βαρύτητα ενώ στην εικόνα διαφορών αντιπροσωπεύονται από κυμαινόμενες τιμές. Η σχέση B1.11 περιγράφει την προβολή της εικόνας $D \in \mathcal{R}^{r \times c}$ στον κατακόρυφο άξονα:

$$v_e(y) = \sum_{i=1}^r D(i, y) \quad (B1.11)$$

Οι κορυφές που προκύπτουν εξομαλύνονται με φίλτρο ενδιάμεσης τιμής ούτως ώστε να αποφευχθεί το ενδεχόμενο να λάβουμε υπόψη κάποια μεμονωμένη κορυφή η οποία δεν αφορά σημαντικό τμήμα του προσώπου.

Με βάση την υπόθεση ότι η περιοχή με την μεγαλύτερη πυκνότητα ανομοιογένειας στο πρόσωπο είναι η περιοχή των ματιών, μια προσέγγιση της οριζόντιας θέσης τους δίνεται από την σχέση:

$$y_0 = \arg \max_i (v_e(i)) \quad (\text{B1.12})$$

Σημειώνεται πως ο παραπάνω αλγόριθμος είναι σχετικά επιρρεπής στην στροφή περί τον άξονα \mathbf{z} που διαταράσσει την ευθυγράμμιση των ματιών. Η γνώση της γωνίας περιστροφής περί τον άξονα \mathbf{z} αυξάνει σημαντικά την ευστάθεια του αλγόριθμου.

Για τον εντοπισμό της κάθετης θέσης των ματιών θεωρούμε την οριζόντια λωρίδα $[y_1, y_2] \times [1, \eta]$ με $y_1 < y_0 < y_2$ που διέρχεται από τα μάτια – με βάση την ευθεία $y=y_0$. Υπολογίζουμε την οριζόντια ολοκληρωματική προβολή της αρχικής εικόνας I στην ανωτέρω λωρίδα σύμφωνα με τη σχέση:

$$h_e(x) = \sum_{i=y_1}^{y_2} I(x, i) \quad (\text{B1.13})$$

Εναλλακτικά αντί της ολοκληρωματικής προβολής της αρχικής εικόνας μπορεί να ληφθεί η ολοκληρωματική προβολή της εικόνας διαφορών:

$$h_e(x) = \sum_{i=y_1}^{y_2} D(x, i) \quad (\text{B1.14})$$

Ανάλογα με το αν χρησιμοποιείται η σχέση B1.13 ή B1.14 ζητείται η εύρεση των δύο ελαχίστων ή των δύο μεγίστων από τις σχέσεις:

$$x_0 = \arg \min_i (h_e(i)) \quad (\text{B1.15})$$

$$x_0 = \arg \max_i (h_e(i)) \quad (\text{B1.16})$$

Όταν χρησιμοποιηθούν οι σχέσεις B1.14 και B1.16 οι κάθετες θέσεις των ματιών δεν εντοπίζονται ακριβώς στις ίριδες των ματιών αλλά σε ελαφρά γειτονικές περιοχές. Αυτό οφείλεται στο γεγονός ότι οι ίριδες είναι τοπικά ομοιογενείς περιοχές. Αντίθετα όταν χρησιμοποιηθούν οι σχέσεις B1.13 και B1.15 οι ίριδες εντοπίζονται επακριβώς. Σημειώνεται ότι για την αποτελεσματική εφαρμογή του αλγορίθμου οι σχέσεις B1.11 έως B1.16 πρέπει να υπολογίζονται εντός του περιγράμματος του προσώπου.

Ο υπολογισμός της θέσης των φρυδιών, των ρουθουνιών και του κέντρου του στόματος γίνεται με εντελώς αντίστοιχη διαδικασία δεδομένου ότι και αυτές οι περιοχές διακρίνονται από υψηλό βαθμό ανομοιογένειας. Η εύρεση των ματιών μπορεί να χρησιμοποιηθεί ως οδηγός για την εύρεση των υπολοίπων περιοχών. Στην περίπτωση των φρυδιών οι σχέσεις B1.13 και B1.15 είναι περισσότερο κατάλληλες ενώ για την εύρεση των ρουθουνιών και του στόματος αποτελεσματικότερη είναι η χρήση των σχέσεων B1.14 και B1.16. Στην Εικόνα B1.5 εμφανίζεται η ανίχνευση της οριζόντιας θέσης του στόματος και των ρουθουνιών.



Εικόνα B1.5: Εντοπισμός της οριζόντιας θέσης της μύτης και του στόματος

2.1 Εντοπισμός επιπλέον σημείων με χρήση των φίλτρων Gabor

Στη περιοχή του προσώπου υπάρχουν και μεμονωμένα σημεία τα οποία δεν μπορούν να εντοπιστούν με τη διαδικασία που περιγράφηκε νωρίτερα. Χαρακτηρίζονται από περιγραφές σε επίπεδο pixel και όχι σε επίπεδο περιοχής. Τέτοια σημεία είναι οι κόγχες των ματιών και του στόματος και μπορούν να ανιχνευθούν με τη βοήθεια φίλτρων Gabor. Η ανίχνευση των σημείων που αντιστοιχούν στις ίριδες και το κέντρο του στόματος μπορούν να αποτελέσουν την αφετηρία για τον εντοπισμό των ανωτέρω σημείων.

Τα φίλτρα Gabor μπορούν να περιγράψουν την περιοχή που περιβάλλει ένα pixel μιας εικόνας. Αυτό γίνεται αν χρησιμοποιήσουμε μια ομάδα από φίλτρα διαφορετικών συχνοτήτων και προσανατολισμών τα οποία έχουν ως αναφορά το συγκεκριμένο pixel. Με τον τρόπο αυτό, τα φίλτρα δίνουν την δυνατότητα σχηματισμού ενός διανύσματος χαρακτηριστικών που δημιουργείται από τις αποκρίσεις τους.

Η ιδιότητά τους αυτή τα έχει καταστήσει βασική επιλογή για την ανίχνευση μεμονωμένων σημείων στα χαρακτηριστικά προσώπου. Σχεδόν σε όλους τους αλγόριθμους ανίχνευσης χαρακτηριστικών σημείων προσώπου τα φίλτρα Gabor αποτελούν το τελικό στάδιο του αλγόριθμου.

Τα φίλτρα έχουν την ακολουθεί μορφή:

$$\Psi_{\underline{k}}(\underline{x}) = \frac{k^2}{\sigma^2} \cdot \exp\left(-\frac{k^2 x^2}{2\sigma^2}\right) \cdot [\exp(-ikx) - \exp(\sigma^{-2})] \quad (B1.17)$$

όπου \underline{x} είναι η θέση κάποιου pixel της εικόνας, \underline{k} είναι η κεντρική συχνότητα, σ είναι το πλάτος του γαουσιανού παραθύρου του πλάτους της συνάρτησης. Το διάνυσμα χαρακτηριστικών έχει σαν στοιχεία του τις αποκρίσεις των φίλτρων στις διάφορες κεντρικές συχνότητες για μια συγκεκριμένη θέση. Η διαφορά φάσης σε ένα στοιχείο του διανύσματος κατά την σύγκριση των δυο εικόνων, μεταφράζεται σε τοπική μεταβολή ως προς την αντίστοιχη κατεύθυνση. Οι μεταβολές προς κάθε κατεύθυνση εκτιμούνται και συνδυάζονται για την δημιουργία μιας κοινής μεταβολής θέσης με βάση το κριτήριο του ελαχίστου τετραγωνικού σφάλματος. Μόνο στα σημεία τα οποία ισχύει:

$$\xi = \sqrt{\sum_j a_j^2} > \xi_{TH} \quad (B1.18)$$

λαμβάνεται υπόψη η μεταβολή θέσης, όπου j είναι ο δείκτης που κινείται στα στοιχεία ενός διανύσματος και a δείχνει το πλάτος των στοιχείων. Η απόκριση φάσης στην χαμηλότερη χωρική συχνότητα μας δίνει την

αρχική εκτίμηση της διαφοράς θέσης ενός χαρακτηριστικού σημείου. Η θέση του χαρακτηριστικού σημείου ξαναυπολογίζεται με βάση αυτή την εκτίμηση και η διαδικασία επαναλαμβάνεται στις υψηλότερες χωρικές συχνότητες, προσδίδοντας συνεχώς μεγαλύτερη ακρίβεια. Η διαφορά θέσης που προκύπτει αντιστοιχίζεται με ένα βαθμό εμπιστοσύνης μεταξύ των δυο συγκρινόμενων διανυσμάτων. Υψηλής εμπιστοσύνης διαφορές θέσης χρησιμοποιούνται για τον εντοπισμό των σημείων. Προκειμένου να ανιχνεύσουμε χαρακτηριστικά σημεία της εικόνας μπορούμε να εφαρμόσουμε την τεχνική αυτή συγκρίνοντας την εκάστοτε εικόνα με πρότυπο. Επίσης η τεχνική αυτή μπορεί να χρησιμοποιηθεί σε εκτίμηση κίνησης προσώπου σε ακολουθίες βίντεο, καθώς και κατά την απεικόνιση των προσώπων με γράφους των οποίων οι κόμβοι φέρουν περιγραφή των αντίστοιχων αποκρίσεων των φίλτρων στις διάφορες συχνότητες και οι ακμές περιγράφουν τοπογραφικές σχέσεις. Η τελευταία χρήση σχετίζεται με ταίριασμα ελαστικών γράφων και βρίσκει εφαρμογή στην αναγνώριση και την απομόνωση προσώπων ή άλλων αντικειμένων σε σύνθετες σκηνές.

Μια άλλη μορφή της δισδιάστατης συνάρτησης Gabor είναι:

$$g(x, y; u_0, v_0) = \exp \left\{ - \left[\frac{x^2}{2\sigma_x^2} + \frac{y^2}{2\sigma_y^2} \right] + 2\pi j [u_0 x + v_0 y] \right\} \quad (\text{B1.19})$$

όπου τα σ_x και σ_y συμβολίζουν τα χωρικά πλάτη της γαουσιανής περιβάλλουσας και στην ουσία ρυθμίζουν την επιλεκτικότητα των αξόνων. Επίσης (u_0, v_0) είναι η εκάστοτε κεντρική συχνότητα της μιγαδικής ημιτονοειδούς. Οι συναρτήσεις Gabor σχηματίζουν μια πλήρη, μη ορθογωνική βάση. Η αναπαράσταση αυτών σε μορφή wavelet είναι:

$$\Phi_\lambda(x, y, \theta) = \exp \left\{ - \lambda^2 (x'^2 + y'^2) \right\} + j\pi x' \quad (\text{B1.20})$$

όπου $x' = x \cdot \cos\theta + y \cdot \sin\theta$ και $y' = x \cdot \sin\theta + y \cdot \cos\theta$. Η παράμετρος θ δείχνει τον προσανατολισμό του φίλτρου ως προς τον χώρο. Ο προσανατολισμός αφορά στην εξαγωγή χαρακτηριστικών ως προς κάποια επιθυμητή κατεύθυνση.

Η παράμετρος προσανατολισμού θ καθορίζει την κατεύθυνση των ανιχνεύσιμων ακμών. Υπάρχει ιδιαίτερη ευαισθησία σε ακμές μικρού μήκους, ενώ παράλληλα μειώνεται η απόκριση για αύξηση των μηκών των γραμμών.

Ένα άλλο σημαντικό χαρακτηριστικό για τις συναρτήσεις Gabor διατυπώνεται παρακάτω. Για την εκθετική συνάρτηση που εκφράζει την περιβάλλουσα, ισχύει η ιδιότητα ότι το γινόμενο του πεδίου ορισμού της συνάρτησης με το εύρος ζώνης του μετασχηματισμού Fourier είναι ελάχιστο. Αυτό συνεπάγεται βέλτιστο εντοπισμό χαρακτηριστικών, δηλαδή η διακριτική ικανότητα γίνεται βέλτιστη τόσο ως προς τον φυσικό χώρο όσο και ως προς τον χώρο των συχνοτήτων ταυτόχρονα.

3. ΕΝΤΟΠΙΣΜΟΣ ΣΗΜΕΙΩΝ ΣΕ ΕΙΚΟΝΕΣ ΠΡΟΦΙΛ

Χαρακτηριστικά σημεία από την οριογραμμή του προφίλ είχαν χρησιμοποιηθεί στις πρώτες μελέτες της αναγνώρισης προσώπων. Στις σύγχρονες αντιμετωπίσεις η χρήση τους είναι σημαντικά περιορισμένη.

Μεγαλύτερη αξία έχουν σε περιβάλλον δημιουργίας – προσαρμογής συνθετικών μοντέλων του ανθρώπινου κεφαλιού. Στην τρέχουσα παράγραφο περιγράφεται συνοπτικά μια διαδικασία εντοπισμού κάποιων χαρακτηριστικών σημείων του προφίλ. Δεδομένου ότι όλα τα σημεία βρίσκονται στην οριογραμμή του προφίλ, οι εικόνες μπορούν να είναι δυαδικές –ασπρόμαυρες. Για τη μετατροπή τους, μπορεί να χρησιμοποιηθεί η διαδικασία υπολογισμού ακμών της παραγράφου 1. Σε κλασικές τεχνικές εντοπισμού σημείων στο προφίλ, η οριογραμμή θεωρείται ως ένα μονοδιάστατο σήμα και τα επιμέρους σημεία υπολογίζονται με τεχνικές εντοπισμού τοπικών μεγίστων και ελαχίστων σε καμπύλες.

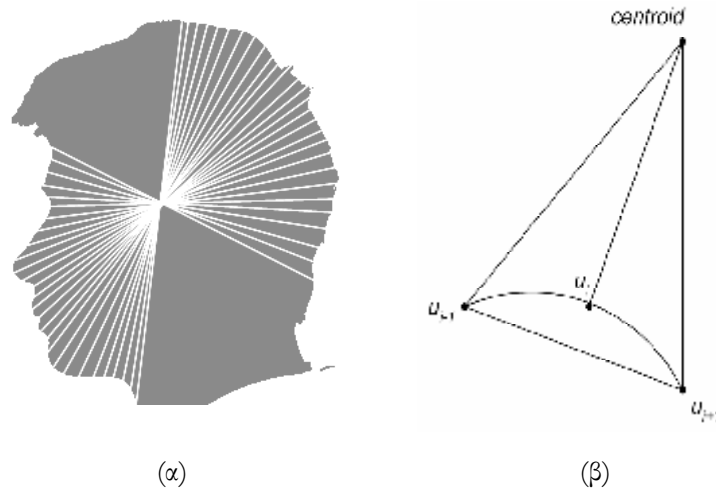
Ένας εναλλακτικός τρόπος υπολογισμού είναι ο γεωμετρικός. Θεωρώντας ως σημείο αναφοράς το κέντρο βάρους της εικόνας προφίλ, χαράσσουμε ευθείες που διέρχονται από αυτό –βλέπε Εικόνα B1.6(α). Με βάση τις τριγωνικές διαμερίσεις που ορίζονται από δύο διαδοχικές ευθείες μπορούμε να εκτιμήσουμε τη καμπυλότητα κάθε σημείου της οριογραμμής. Έστω u_{i-1} , u_i , u_{i+1} τρία διαδοχικά σημεία της οριογραμμής με αντίστοιχες αποστάσεις από το κέντρο βάρους r_{i-1} , r_i , r_{i+1} . Έστω επίσης l_i το ευθύγραμμο τμήμα που συνδέει τα σημεία u_{i-1} , u_{i+1} και m_i το σημείο τομής των ευθυγράμμων τμημάτων r_i και l_i . Το μέτρο της καμπυλότητας του σημείου u_i μπορεί να υπολογιστεί από τη σχέση –βλέπε και Εικόνα B1.6(β):

$$D_i = \|u_i - l_i\| \quad (\text{B1.21})$$

Το πρόσημο της καμπυλότητας προκύπτει από τη σχέση:

$$s_i = \text{sign}(r_i - r_{m_i}) \quad (\text{B1.22})$$

όπου r_{m_i} είναι η απόσταση του σημείου m_i από το κέντρο βάρους.



Εικόνα B1.6: Εντοπισμός χαρακτηριστικών σημείων στο προφίλ

Η διαδικασία πραγματοποιείται σειριακά και διακόπτεται κάθε φορά που υπάρχει αλλαγή του πρόσημου της καμπυλότητας. Δύο διαδοχικά σημεία μηδενικής καμπυλότητας μπορούν να αποτελέσουν σημεία

αναφοράς για την εκτίμηση της καμπυλότητας όλων των ενδιάμεσων σημείων σύμφωνα με τις σχέσεις B1.21 και B1.22. Τα σημεία με τη μέγιστη τοπική καμπυλότητα μπορούν να θεωρηθούν ως τα χαρακτηριστικά σημεία της οριογραμμής.

4.3-Δ ΑΝΑΠΑΡΑΣΤΑΣΗ

Η εξαγωγή χαρακτηριστικών σημείων της επιφάνειας του προσώπου και της οριογραμμής του προφίλ μπορεί να χρησιμοποιηθεί για την προσαρμογή τρισδιάστατων μοντέλων κεφαλιού σε συγκεκριμένες περιπτώσεις –δημιουργία ατομικών μοντέλων. Στην Εικόνα Γ6.1 υπενθυμίζεται ποια σημεία θεωρείται ότι μπορούν να ανιχνευθούν αυτόματα, σε εικόνες όπως περιγράφηκε στις προηγούμενες παραγράφους.



Εικόνα B1.7: Χαρακτηριστικά σημεία στη εμπρόσθια όψη και στο προφίλ

Χαρακτηριστικά στην όψη profile	Περιγραφή	Χαρακτηριστικά πρόσοψης	Περιγραφή
Γωνία 1-2-3	Χαρακτηριστικό του σαγονιού	Λόγος $d(14,15)/d(4,9)$	Χαρακτηριστικό του πλάτους του προσώπου
Γωνία 7-8-9	Χαρακτηριστικό της μύτης	Λόγος $d(12,13)/d(14,15)$	Απόσταση ανάμεσα στα μάτια
Λόγος $d(8,10)/d(2,8)$	Μεγάλο ή μικρό μέτωπο	Λόγος $d(5a,5b)/d(14,15)$	Πλάτος του στόματος
Γωνία 5-6-7	Χαρακτηριστικό άνω χειλους	Λόγος $d(4,6)/d(14,15)$	Ύψος του στόματος
Γωνία 3-4-5	Χαρακτηριστικό κάτω χειλους	Λόγος $d(12a,12b)/d(14,15)$	Πλάτος του ματιού
Γωνία 4-5-6	Γωνία ανάμεσα στα χείλη	Λόγος $d(12c,12d)/d(14,15)$	Ύψος του ματιού
		Λόγος $d(8,9)/d(14,15)$	Μήκος της μύτης
		Λόγος $d(8a,8b)/d(14,15)$	Πλάτος της μύτης

Πίνακας Β1.1: Περιγραφή χαρακτηριστικών υψηλότερου επιπέδου με βάση μεμονωμένα σημεία

Τα σημεία που εντοπίζονται είναι ένα υποσύνολο των παραμέτρων FDP που προβλέπονται στο πρότυπο MPEG-4. Βρίσκονται κατά μείζονα λόγο συγκεντρωμένα γύρω από το στόμα και τα μάτια (στην όψη frontal), γεγονός που ευνοεί ιδιαίτερα τις διαδικασίες τροποποίησης ενός γενικού μοντέλου ανθρώπινου προσώπου ή εντοπισμού ενός προσώπου από μια σχετική βάση δεδομένων. Στον Πίνακα Γ6.1 εμφανίζεται πως με βάση τα μεμονωμένα σημεία στη εμπρόσθια όψη και το προφίλ μπορεί να δοθεί μια περιγραφή υψηλότερου επιπέδου για κάποια ανατομικά χαρακτηριστικά του προσώπου. Με βάση των Πίνακα Γ6.1 πραγματοποιείται μετατροπή των χαρακτηριστικών του τρισδιάστατου μοντέλου. Για λεπτομερέστερη επεξήγηση ο αναγνώστης παραπέμπεται στις εργασίες [31]-[33].

ΕΝΤΟΠΙΣΜΟΣ ΠΡΟΣΩΠΩΝ ΣΕ ΕΓΧΡΩΜΕΣ ΕΙΚΟΝΕΣ ΚΑΙ ΑΚΟΛΟΥΘΙΕΣ ΒΙΝΤΕΟ

Ο εντοπισμός προσώπου δεν είναι πλέον άρρηκτα συνδεδεμένος με την αναγνώριση προσώπων. Αντίθετα αποτελεί ένα πολύ σημαντικό στοιχείο στο πλαίσιο πολλών εφαρμογών πολυμέσων όπως δεικτοδότηση, κατηγοριοποίηση σκηνών σε βίντεο και ανάλυση και περίληψη ειδήσεων. Εμπνευσμένοι από την εργασία [1], πολλοί ερευνητές παρουσίασαν εργασίες για εντοπισμό προσώπων με βάση τα χαρακτηριστικά χρώματος του ανθρώπινου δέρματος. Οι περισσότεροι από τους αλγόριθμους αυτούς αντιμετωπίζουν προβλήματα γενίκευσης εξαιτίας της μοντελοποίησης χρώματος που επιχειρούν. Επιπλέον το στάδιο επαλήθευσης που χρησιμοποιούν εξαρτάται αποκλειστικά από χαρακτηριστικά σχήματος και όχι υφής γεγονός που περιορίζει σημαντικά την αξιοπιστία του εντοπισμού. Στο κεφάλαιο αυτό παρουσιάζεται ένα Γκαουσιανό μοντέλο, για την προσέγγιση της πιθανοτικής κατανομή χρώματος του δέρματος, του οποίου οι παράμετροι επαναπροσδιορίζονται με βάση την τρέχουσα εικόνα / πλαίσιο εισόδου. Με τον τρόπο αυτό τα προβλήματα γενίκευσης περιορίζονται σημαντικά. Επιπλέον το στάδιο επαλήθευσης, το οποίο εφαρμόζεται στα εντοπισμένα τμήματα δέρματος, βασίζεται σε μια παραλλαγή της διαδικασίας ταύτισης με πρότυπο.

1. ΕΙΣΑΓΩΓΗ

Στο παρελθόν ο όρος *face detection-εντοπισμός προσώπου* ήταν ισχυρά συνδεδεμένος με το πρόβλημα της αναγνώρισης προσώπων γεγονός που είχε βαθιά επίδραση στους αλγορίθμους που αναπτύχθηκαν. Για να επιτευχθεί η απαιτούμενη ακρίβεια εντοπισμού πρέπει να τεθούν πολύ αυστηροί [2]. Επιπλέον η μεγάλη πλειονότητα των αλγορίθμων βασίζονταν σε μονόχρωμες εικόνες και χρησιμοποιούντο τεχνικές όπως η ταύτιση με πρότυπο, οι γεωμετρικές ροπές και η ανίχνευση των επιμέρους χαρακτηριστικών του προσώπου όπως τα μάτια, η μύτη και το στόμα [3][4].

Πρόσφατα η ταχεία ανάπτυξη των εφαρμογών πολυμέσων πρόσθεσε αξία στους αλγόριθμους εντοπισμού προσώπου και τους αποσύνδεσε από το πρόβλημα της αναγνώρισης προσώπου. Για εφαρμογές όπως η δεικτοδότηση και ανάκληση εικόνων και βίντεο [5][6], η ταξινόμηση σκηνών βίντεο και η περίληψη ειδήσεων [7], ο εντοπισμός προσώπων αποτελεί ένα πολύτιμο εργαλείο. Οι σύγχρονες εφαρμογές, παρόλα αυτά απαιτούν γρήγορες υλοποιήσεις με επαρκή ακρίβεια, παρά εξαντλητικές διαδικασίες με υψηλή ακρίβεια στον εντοπισμό. Αποτέλεσμα του γεγονότος αυτού είναι οι αλγόριθμοι ανίχνευσης, που χρησιμοποιήθηκαν αποτελεσματικά στο πλαίσιο συστημάτων αναγνώρισης προσώπων, να μην μπορούν να χρησιμοποιηθούν ή να χρειάζονται επανασχεδιασμό. Τέλος πολλές από τις εφαρμογές πολυμέσων που σχετίζονται με εικόνες και βίντεο εμπλέκουν χαρακτηριστικά χρώματος σε αντίθεση με τους παραδοσιακούς αλγορίθμους ανίχνευσης προσώπων.

Η εργασία [1] ενέπνευσε πολλούς ερευνητές για την υλοποίηση αλγορίθμων ανίχνευσης προσώπων οι οποίοι βασίζονται σε χαρακτηριστικά χρώματος. Η βασική ιδέα της εργασίας [1] είναι η μοντελοποίηση των χρωματικών χαρακτηριστικών του ανθρώπινου δέρματος στο χώρο χρωμάτων *YCrCb*. Οι περισσότεροι από τους αλγορίθμους αυτούς εμπλέκουν κάποια στάδια μετεπεξεργασίας για να επιτύχουν αξιοσημείωτα

αποτελέσματα [7]. Παρόλα αυτά δεν είναι σαφές αν τα στάδια μετεπεξεργασίας είναι αρκετά για να οδηγήσουν σε ένα επαρκές αποτέλεσμα εντοπισμού προσώπου τους αλγόριθμους που βασίζονται στα χαρακτηριστικά χρώματος. Παρόλο που ο υπόχωρος που καταλαμβάνεται από τις χρωματικές συνιστώσες του δέρματος στο χρωματικό πεδίο C_r-C_b είναι πράγματι μικρός δεν μπορεί να μοντελοποιηθεί με ένα τέτοιο γενικό τρόπο ώστε να είναι αποδοτικός για όλες τις εικόνες οι οποίες περιέχουν πρόσωπα. Για να βελτιώσουμε την ικανότητα γενίκευσης του μοντέλου πρέπει να το κάνουμε πιο «χαλαρό» το οποίο όμως οδηγεί σε αυξημένο ποσοστό *false alarms-λανθασμένων εντοπισμών*. Από τη άλλη πλευρά ένα «αυστηρό» μοντέλο οδηγεί σε ένα αυξημένο αριθμό *dismissals-απωλειών εντοπισμού*. Επιπλέον η επίδραση του γενικού φωτισμού όπως εκφράζεται μέσα από το κανάλι Y δεν είναι εντελώς αμελητέα.

Στο κεφάλαιο αυτό αντιμετωπίζεται το πρόβλημα του εντοπισμού προσώπου συνδυάζοντας τις χρωματικές συνιστώσες του δέρματος, η πιθανοτική κατανομή των οποίων προσεγγίζεται με ένα δισδιάστατο Γκαουσιανό μοντέλο, και την εφαρμογή μιας διαδικασίας ταύτισης με πρότυπο. Το συνολικό σχήμα βελτιώνει τη αποδοτικότητα της ανίχνευσης προσώπου με δύο διαφορετικούς τρόπους: (α) Βελτιώνοντας την ικανότητα γενίκευσης του χρωματικού μοντέλου για το δέρμα η οποία επιτυγχάνεται με επανεκτίμηση των παραμέτρων της Γκαουσιανής κατανομής με βάση την τρέχουσα εικόνα εισόδου και, (β) με τη χρήση ενός σταδίου επαλήθευσης το οποίο εμπλέκει και πληροφορία υψής εκτός από χαρακτηριστικά χρώματος και σχήματος.

2. ΕΝΤΟΠΙΣΜΟΣ ΤΩΝ ΠΙΘΑΝΟΤΕΡΩΝ ΤΜΗΜΑΤΩΝ ΔΕΡΜΑΤΟΣ

Όπως έχει αναφερθεί σε κλασικές μελέτες [8][9] οι χρωματικές συνιστώσες του δέρματος καταλαμβάνουν μια μικρή περιοχή του χρωματικού επιπέδου C_r-C_b του μοντέλου χρωμάτων $[Y, C_r, C_b]$. Ο Wang 0, βασισμένος στην ιδέα αυτή παρουσίασε ένα αλγόριθμο ανίχνευσης προσώπων ο οποίος ενέπνευσε πολλούς ερευνητές. Με ένα παρόμοιο τρόπο προσεγγίζουμε την πιθανοτική κατανομή των χρωματικών συνιστωσών του δέρματος μέσω μιας δισδιάστατης Γκαουσιανής συνάρτησης πυκνότητας πιθανότητας. Η επιλογή της Γκαουσιανής κατανομής στηρίζεται στο κεντρικό οριακό θεώρημα και στην τυχαιότητα η οποία χαρακτηρίζει την χρωματική υφή του δέρματος στους ανθρώπους. Επεκτείνοντας την ιδέα του Wang χρησιμοποιούμε ένα απλό γραμμικό μοντέλο ανατροφοδότησης για την επανεκτίμηση των παραμέτρων της κατανομής με βάση την νέα πληροφορία η οποία εισέρχεται στο σύστημα μέσω της τρέχουσας εικόνας / καρέ.

Υποθέτοντας ότι το μέσο διάνυσμα μ_0 και ο πίνακας συμμεταβλητότητας \mathbf{C} έχουν ευσταθώς εκτιμηθεί από κάποιο σύνολο μάθησης, η πιθανότητα ενός προτύπου εισόδου \mathbf{x} –το οποίο εκφράζει τις τιμές των χρωματικών συνιστωσών C_r, C_b κάποιου pixel – δίνεται από τη σχέση:

$$P(\mathbf{x} | \mu_0, \mathbf{C}) = \frac{\exp\left\{-\frac{1}{2}(\mathbf{x} - \mu_0)^T \mathbf{C}^{-1}(\mathbf{x} - \mu_0)\right\}}{(2\pi)^{\frac{k}{2}} \cdot |\mathbf{C}|^{\frac{1}{2}}} \quad (\text{B2.1})$$

όπου k = αριθμός των χρωματικών συνιστωσών = 2

Για την εκτίμηση των μ_0 και C χρησιμοποιήσαμε ως σύνολο μάθησης pixels προσώπων από απλά βιντεοκλίπς, έγχρωμες εικόνες, ατομικές βιντεοκάμερες και ψηφιακές φωτογραφικές μηχανές.

Παρόλο που η χρήση του ανωτέρω Γκαουσιανού μοντέλου, για την ταξινόμηση των pixels ως pixels δέρματος ή μη, είναι αρκετά αποδοτική για καλύτερη απόδοση και ικανότητα γενίκευσης πραγματοποιείται επανεκτίμηση του μέσου διανύσματος με βάση την τρέχουσα εικόνα / καρέ.

Έστω I το σύνολο των pixels της εικόνας εισόδου, χωρικής ανάλυσης $M \times N$ και $p(\mathbf{x})$ η πιθανότητα του pixel \mathbf{x} να είναι pixel δέρματος τότε τα σύνολα X και X^c ικανοποιούν τις σχέσεις:

$$X \subset I, X^c \subset I, X \cup X^c = I$$

και το X ορίζεται όπως παρακάτω:

$$X: \{x \in I \mid p(\mathbf{x}) \geq C_I\}$$

Όπου $C_I = \mu_x + \sigma_x$, $\mu_x = \frac{1}{M \cdot N} \sum_{x \in I} p(x)$ αντιπροσωπεύει τη μέση πιθανότητα των pixels της εικόνας και

$$\sigma_x = \frac{1}{M \cdot N - 1} \sqrt{\sum_{x \in I} \{p(x) - \mu_x\}^2}$$
 είναι η τυπική τους απόκλιση.

Αν L είναι ο αριθμός των στοιχείων του συνόλου X τότε $\mu = \frac{1}{L} \sum_{x \in X} x$ αποτελεί το μέσο διάνυσμα των χρωματικών συνιστωσών των pixels δέρματος όπως αυτό εκτιμάται από το σύνολο X .

Σύμφωνα με τα ανωτέρω το αρχικό Γκαουσιανό μοντέλο, όπως εκφράζεται από την εξίσωση (B2.1), χρησιμοποιείται σε ένα πρώτο πέρασμα για την ταξινόμηση των pixels και τα pixels τα οποία έχουν ταξινομηθεί ως pixels δέρματος χρησιμοποιούνται για την επανεκτίμηση του διανύσματος μ_0 σύμφωνα με την εξίσωση:

$$\mu_0 = (1 - m) \cdot \mu_0 + m \cdot \mu \tag{B2.2}$$

όπου μ είναι το εκτιμώμενο από την τρέχουσα εικόνα μέσο διάνυσμα των χρωματικών συνιστωσών των pixels και m είναι μια σταθερά μνήμης για το γραμμικό μοντέλο ανατροφοδότησης.

Με βάση τον ορισμό του το σύνολο X δεν μπορεί να είναι κενό, περίπτωση που μπορεί να συμβεί αν το C_I ληφθεί σταθερό και εξαιτίας της μη ικανότητας γενίκευσης του αρχικού Γκαουσιανού μοντέλου. Κατά συνέπεια είναι πάντοτε δυνατή η επανεκτίμηση των παραμέτρων της Γκαουσιανής κατανομής. Από τη άλλη πλευρά, για το σύνολο X δεν μπορεί να αποκλειστεί η πιθανότητα να μην περιλαμβάνει pixels δέρματος και κατά συνέπεια η εκτίμηση του μ να είναι εντελώς λανθασμένη. Για το λόγο αυτό η αρχική εκτίμηση των παραμέτρων του μοντέλου είναι πάρα πολύ σημαντική και επιπλέον η επιλογή του m πρέπει να γίνεται πολύ προσεκτικά ώστε να μην υπάρξει αποσυντονισμός αλλά απλή ρύθμιση του αρχικού μοντέλου.

Η τελική ταξινόμηση των pixels πραγματοποιείται χρησιμοποιώντας το τροποποιημένο Γκαουσιανό μοντέλο και εφαρμόζοντας τις σχέσεις:

$$\frac{p(\mathbf{x} | \omega_1)}{p(\mathbf{x} | \omega_0)} > th \Rightarrow \mathbf{x} \in \omega_1 \quad (\text{B2.3})$$

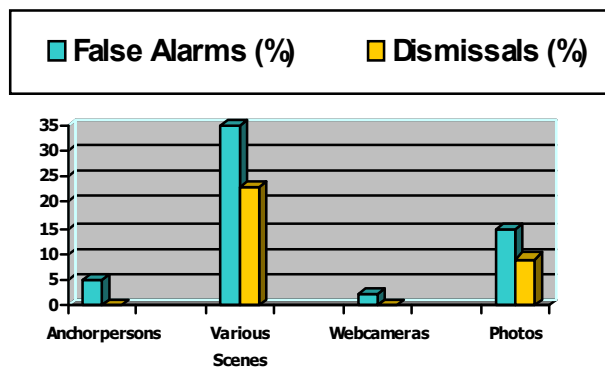
$$\frac{p(\mathbf{x} | \omega_1)}{p(\mathbf{x} | \omega_0)} < th \Rightarrow \mathbf{x} \in \omega_0 \quad (\text{B2.4})$$

$$th = \frac{c_{01} \cdot p(\omega_0)}{c_{10} \cdot p(\omega_1)} \quad (\text{B2.5})$$

όπου ω_1 θεωρούμε την κλάση των pixel δέρματος και ω_0 τη κλάση των υπολοίπων. Οι πιθανότητες $p(\omega_0)$ και $p(\omega_1)$ εκφράζουν τις α priori πιθανότητες των κλάσεων ω_0 και ω_1 αντίστοιχα. Οι σταθερές c_{01} και c_{10} εκφράζουν το κόστος της λανθασμένης ταξινόμησης κάποιου pixel στην κλάση ω_1 –*false alarm*- και ω_0 –*dismissal*- αντίστοιχα.

Στο τέλος της ανωτέρω διαδικασίας δημιουργείται μια δυαδική εικόνα στην οποία τα τμήματα δέρματος αντιπροσωπεύουν το πρώτο πλάνο και το υπόλοιπο κομμάτι της εικόνας το φόντο.

Η εξίσωση (B2.2) χρησιμοποιείται επίσης για την προσαρμογή του μοντέλου σε ένα πλήρως δυναμικό περιβάλλον. Συγκεκριμένα σε ακολουθίες βίντεο οι παράμετροι του μοντέλου επανεκτιμούνται καρέ προς καρέ παρέχοντας τη δυνατότητα ευσταθούς παρακολούθησης της κίνησης των τμημάτων δέρματος.



Σχήμα B2.1: Αποτελέσματα επιτυχούς ανίχνευσης τμημάτων δέρματος για τέσσερις διαφορετικές κατηγορίες εικόνων

Στο Σχήμα B2.1 παρουσιάζεται μια σύγκριση ανάμεσα στην αποτελεσματικότητα εντοπισμού τμημάτων δέρματος για διάφορες κατηγορίες έγχρωμων εικόνων. Παρατηρούμε ότι για ελεγχόμενες συνθήκες λήψης όπως αυτές που αντιστοιχούν σε σκηνές από τηλεοπτικά studio ή λήψη σε περιβάλλον γραφείου (ατομικές βιντεοκάμερες) ο αλγόριθμος που περιγράφηκε παραπάνω εμφανίζει υψηλά ποσοστά επιτυχίας. Σε περιπτώσεις ανομοιομορφων συνθηκών φωτισμού όπως αυτές που αντιστοιχούν σε σκηνές από τηλεοπτικά προγράμματα (various scenes -ταινίες, ειδήσεις κλπ) έχουμε αρκετά υψηλό ποσοστό αποτυχίας ανίχνευσης και ακόμα υψηλότερο ποσοστό εσφαλμένων εντοπισμών. Στη περίπτωση απλών φωτογραφιών έχουμε ενδιάμεσα ποσοστά επιτυχίας δεδομένου ότι οι συνθήκες φωτισμού δεν παρουσιάζουν τόσο μεγάλη ανομοιογένεια αλλά ούτε και είναι πλήρως ελεγχόμενες.

3. ΑΠΟΜΟΝΩΣΗ ΤΩΝ ΤΜΗΜΑΤΩΝ ΔΕΡΜΑΤΟΣ

Το αποτέλεσμα του πρώτου σταδίου του αλγορίθμου μπορεί να δώσει τμήματα δέρματος τα οποία δεν είναι συμπαγή και επιπλέον το κάθε ένα από αυτά πρέπει να ελεγχθεί ξεχωριστά για το αν αποτελεί πρόσωπο ή όχι. Η εφαρμογή μορφολογικών τελεστών όπως οι τελεστές *opening* και *closing* [11] μπορούν να εφαρμοστούν για την επίτευξη συμπαγών αντικειμένων. Μετά τη διαδικασία του φιλτραρίσματος με τους ανωτέρω τελεστές ο μετασχηματισμός απόστασης (*morphological distance transform*) και τεχνικές κατανομής ιστογράμματος εφαρμόζονται για την απομόνωση των επιμέρους μη συνδεδεμένων τμημάτων [8] σύμφωνα με το μέγεθος τους.

Η διαδικασία συνοψίζεται στα παρακάτω βήματα:

Έστω X το σύνολο pixels τα οποία έχουν ταξινομηθεί ως pixel δέρματος. Έστω επίσης ότι το σύνολο X αποτελείται από $S_i, i = 1 \dots n$ συμπαγή υποσύνολα τέτοια ώστε:

$$S_i \cap S_j = \emptyset \text{ για όλα τα } i \neq j, \text{ και } \bigcup_{i=1}^n S_i = X.$$

Αν $D(S_i)$ είναι ο μετασχηματισμός απόστασης του συνόλου S_i –βλέπε Εικόνα B2.1 (γ)– τότε μια ταξινόμηση $\{S_1 \geq S_2 \geq \dots \geq S_n\}$ των συνόλων από $S_i, i = 1 \dots n$, μπορεί να οριστεί με βάση τη σχέση $d(S_i) = \max\{D(S_i)\}$. Σύμφωνα με τη σχέση αυτή ισχύει $S_i > S_j$ αν $d(S_i) > d(S_j)$.

Έστω τώρα ότι με βάση το παραπάνω κριτήριο ταξινόμησης ισχύει $\{S_1 > S_2 \geq \dots \geq S_n\}$. Για την απομόνωση του συνόλου S_1 ακολουθείται η επόμενη διαδικασία:

Αν $q = \max_i \{D(S_i)\}$ τότε προφανώς ισχύει $q = d(S_1)$. Ορίζουμε το σύνολο M (markers) με βάση τη σχέση:

$$M: \{\mathbf{x} \in X \mid D(\mathbf{x}) \geq q\}, \text{ όπου } D(\mathbf{x}) \text{ η τιμή του μετασχηματισμού απόστασης στο σημείο } \mathbf{x}.$$

Προφανώς ισχύει $M \subset S_1$. Χρησιμοποιώντας το σύνολο M σε μια διαδικασία *opening by reconstruction* [11] πάνω στο σύνολο X καταλήγουμε στα σύνολα X_1, X_2 τα οποία ικανοποιούν τις σχέσεις $X_1 = S_1$ και $X_2 = X - S_1$ –βλέπε Εικόνα B2.1(δ) και (ε) αντίστοιχα.

Θεωρώντας το σύνολο X_2 ως το νέο σύνολο X και επαναλαμβάνοντας τη ανωτέρω διαδικασία απομονώνουμε διαδοχικά τα υποσύνολα S_2, S_3, \dots, S_n .

Στη περίπτωση που για κάποια σύνολα $S_i, S_j, i \neq j$, ισχύει η σχέση $d(S_i) = d(S_j)$, τότε αν n_m είναι ο αριθμός των στοιχείων του συνόλου M θα ισχύει $n_m \geq 2$ και δεν θα ισχύει καμία από τις σχέσεις $M \subset S_i, M \subset S_j$. Για να ειρυνιστεί η ανωτέρω περίπτωση αντικαθιστούμε το σύνολο M με το σύνολο M_1 για το οποίο ισχύει $n_{m1} = 1$ και $M_1 \subseteq M$. Η επιλογή αυτή απαγορεύει την απομόνωση των συνόλων S_i και S_j ταυτόχρονα –τα οποία σε μια τέτοια περίπτωση θα θεωρούνταν ως ένα. Φυσικά μπορεί να ισχύουν

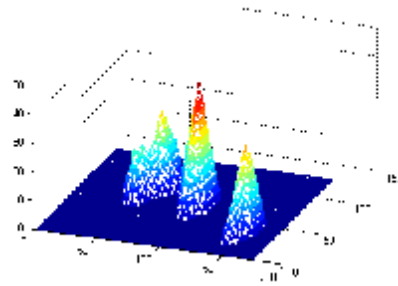
ταυτόχρονα οι σχέσεις $n_m \geq 2$ και $M \subset S_i$ αλλά σε αυτή την περίπτωση η επιλογή του M , αντί M πάλι θα οδηγήσει σε απομόνωση του S_i —έστω όχι και τόσο γρήγορα.



(α)



(β)



(γ)



(δ)



(ε)



(ζ)

Εικόνα B2.1: (α) Αρχική εικόνα (β) εντοπισμός των πιθανών τμημάτων δέρματος (γ) ο μετασχηματισμός απόστασης (δ) ανακατασκευασμένο τμήμα της (β) με χρήση ως marker του σημείου με τη μέγιστη τιμή του μετασχηματισμού απόστασης (ε) μη ανακατασκευασμένο τμήμα της (β), (ζ) τα ορθογώνια που περιλαμβάνουν τα επιμέρους τμήματα και που οδηγούνται στο επόμενο στάδιο

Τα απομονωμένα τμήματα δέρματος τα οποία προκύπτουν από την παραπάνω διαδικασία έχουν ακανόνιστα σχήματα. Κάποιου είδους φιλτράρισμα με βάση τα χαρακτηριστικά σχήματος μπορεί να εφαρμοστεί, ώστε να απαλειφθούν τα τμήματα εκείνα τα οποία είναι απίθανο να αντιστοιχούν σε πρόσωπα. Για κάθε ένα από τα υπόλοιπα τμήματα το ορθογώνιο που τα περιλαμβάνει αποκρίπτεται από την αρχική

εικόνα και οδηγείται στο επόμενο στάδιο. Η Εικόνα B2.1 διευκρινίζει την προηγούμενη διαδικασία. Τα τμήματα δέρματος φαίνονται στην Εικόνα B2.1(β) και τα αντίστοιχα ορθογώνια επιδεικνύονται στην Εικόνα B2.1(ζ).

4. ΦΙΛΤΡΑΡΙΣΜΑ ΜΕ ΒΑΣΗ ΤΟ ΣΧΗΜΑ

Τα απομονωμένα τμήματα δέρματος μπορεί να ανήκουν σε πρόσωπα, άλλα τμήματα ανθρώπινου δέρματος ή σε άλλα αντικείμενα με χρώμα παρόμοιο με αυτό του δέρματος. Δεδομένου ότι η διαδικασία ταύτισης με πρότυπο που ακολουθεί είναι πάντοτε υπολογιστικά πολύπλοκη, όσο λιγότερα τμήματα δέρματος εξεταστούν με τον τρόπο αυτό τόσο ταχύτερη θα είναι η συνολική εκτέλεση του αλγορίθμου. Στην παράγραφο αυτή περιγράφεται μια διαδικασία με την οποία κάποια πιθανά τμήματα δέρματος με πολύ ακανόνιστο σχήμα απομακρύνονται και δεν οδηγούνται στο επόμενο στάδιο.

Ιδανικά ένα πρόσωπο έχει αναλογίες που πλησιάζουν με αυτές μιας έλλειψης με λόγο μεγάλου προς μικρό άξονα από 1.2 έως 1.6. Η ομοιότητα ενός σχήματος με μια έλλειψη μπορεί να χρησιμοποιηθεί ως κριτήριο σχήματος για τα πιθανά τμήματα δέρματος. Τάυτιση σχημάτων, υπό τυχαίες μορφοποιήσεις, μπορεί να πραγματοποιηθεί με αρκετές τεχνικές όπως *active contour models –snakes-* ή παραμορφώσιμα πρότυπα -*deformable templates* [30]. Επιπλέον η συμπαγής κίνηση των αντικείμενων ή οι *affine* μετασχηματισμοί όπως η μετατόπιση, η στρόφιξη, η αλλαγή κλίμακας και η στρέβλωση μπορούν να απαλειφθούν χρησιμοποιώντας *affine invariants* ή κανονικοποίηση καμπύλης [29].

Τα περιγράμματα των πιθανών τμημάτων δέρματος δεν είναι σε καμία περίπτωση ομαλά. Ακόμα και μετά την προσέγγιση τους με *splines* για παράδειγμα οι ανωτέρω διαδικασίες ταύτισης σχήματος δίνουν σχετικά φτωχά αποτελέσματα. Επιπλέον η διαδικασία φιλτραρίσματος με βάση το σχήμα πρέπει να είναι σχετικά γρήγορη για να αξίζει να εφαρμοστεί. Για το σκοπό αυτά λαμβάνουμε υπόψη μόνο ολικά χαρακτηριστικά τα σχημάτων. Συγκεκριμένα ορίζουμε την κυκλικότητα ή βαθμό συμπαγότητας -*compactness*- με βάση την περίμετρο και το εμβαδόν ενός σχήματος και σύμφωνα με τη σχέση:

$$g_x = \frac{4\pi \cdot a_x}{r_x^2} \quad (B2.7)$$

όπου r_x είναι η περίμετρος –αριθμός των pixels του περιγράμματος- και a_x είναι το εμβαδόν που καλύπτει –συνολικός αριθμός των pixels- του τμήματος X . Σημειώνεται ότι η μέγιστη κυκλικότητα λαμβάνεται για τμήματα με κυκλικό σχήμα και αντιστοιχεί στην τιμή ένα, και με βάση τον ορισμό της η κυκλικότητα είναι κανονικοποιημένη στο διάστημα $[0,1]$.

Ένα άλλο γενικό χαρακτηριστικό του σχήματος λαμβάνεται από το βαθμό επιμήκυνσής του -*elongation*- ο οποίος εκφράζει το λόγο των βασικών αξόνων του, όπως αυτός λαμβάνεται από τον διακριτό μετασχηματισμό *Kathunen-Loeve* του περιγράμματος. Συγκεκριμένα έστω τα $N \times 1$ διανύσματα \mathbf{x} και \mathbf{y} που εκφράζουν τις συντεταγμένες των σημείων του περιγράμματος του τμήματος X . Ο 2×2 πίνακας συμμεταβλητότητας των σημείων αυτών σε σχέση με το κέντρο μάζας τους (μ_x, μ_y) δίνεται από τη σχέση:

$$\mathbf{C} = \frac{1}{N} [\mathbf{x} - \mu_x \quad \mathbf{y} - \mu_y]^T [\mathbf{x} - \mu_x \quad \mathbf{y} - \mu_y] \quad (\text{B2.8})$$

Τα δύο ιδιοδιανύσματα του πίνακα C εκφράζουν τους βασικούς άξονες του σχήματος. Ο λόγος των ιδιοτιμών του εκφράζει τον βαθμός επιμήκυνσης του:

$$\ell_x = \sqrt{\lambda_2 / \lambda_1} \quad (\text{B2.9})$$

οπού λ_1, λ_2 είναι η μέγιστη και η ελάχιστη ιδιοτιμή του πίνακα C αντίστοιχα. Στην πραγματικότητα η σχέση (B2.9) εκφράζει το αντίστροφο της επιμήκυνσης –δεδομένου ότι λαμβάνεται ο λόγος της μικρότερης προς την μεγαλύτερη ιδιοτιμή και όχι το αντίστροφο.

Τα παραπάνω γενικά χαρακτηριστικά σχήματος είναι αριετά ευσταθή, σε σχέση με το θόρυβο που μπορεί να προκύψει κατά τη διαδικασία κατάτμησης των πιθανών τμημάτων δέρματος, και αμφότερα είναι κανονικοποιημένα στο διάστημα $[0,1]$. Είναι επίσης ανεπηρέαστα σε αλλαγή κλίμακας, μετατόπιση και στροφή περί τον άξονα \mathbf{z} . Πειραματικά αποτελέσματα έδειξαν ότι οι τυπικές τιμές, για τμήματα που ανήκουν σε πρόσωπα, κυμαίνονται από 0.44 έως 0.79 για την κυκλικότητα και 0.59 έως 0.91 για την επιμήκυνση. Για το φιλτράρισμα με βάση το σχήμα τα ανωτέρω χαρακτηριστικά πρέπει να ικανοποιούνται ταυτόχρονα. Για παράδειγμα, ένα τμήμα με σχήμα σταυρού μπορεί να έχει την κατάλληλη επιμήκυνση αλλά θα απορριφθεί εξαιτίας χαμηλής τιμής στην κυκλικότητα. Αντίστοιχα ένα κυκλικό σχήμα με το περιγράμματό του επηρεασμένο από θόρυβο μπορεί να δώσει σωστή τιμή κυκλικότητας αλλά θα απορριφθεί εξαιτίας πολύ μεγάλης τιμής στην επιμήκυνση.

5. ΕΝΤΟΠΙΣΜΟΣ ΠΡΟΣΩΠΩΝ ΕΝΤΟΣ ΤΩΝ ΤΜΗΜΑΤΩΝ

Στο τρίτο στάδιο του αλγορίθμου για όλα τα απομονωμένα πιθανά τμήματα δέρματος, όπως αυτά περιγράφονται από το ορθογώνιο που τα περικλείει, εφαρμόζεται η επόμενη διαδικασία εντοπισμού προσώπων με βάση της αρχής της ταύτισης με πρότυπο:

Έστω $M(u, \theta)$ είναι ένα πρότυπο προσώπου σε κλίμακα $u(h, v)$, η οποία περιγράφεται από την οριζόντια χωρική ανάλυση h και την κάθετη v , και γωνία κλίσης (στροφή ως προς τον άξονα \mathbf{z}) θ . Αν F είναι μια περιοχή της εικόνας η οποία πιθανόν να περιλαμβάνει κάποιο πρόσωπο σε τυχαία κλίμακα, θέση και κλίση, και A μια υποπεριοχή της F τότε χρησιμοποιούμε την επόμενη μετρική για να βρούμε την ελάχιστη συσχέτιση ανάμεσα στα σύνολα A και M σε κλίμακα u και κλίση θ :

$$r(u, \theta) = \min_{A \subset F} \left\{ \frac{|A - M(u, \theta)|}{\rho \cdot a \cdot b} \right\} \quad (\text{B2.10})$$

όπου $\rho = 1 - c \cdot \left| \frac{h}{v} - \frac{2}{3} \right|$, χρησιμοποιείται για να συμπεριλάβει την ανατομία του προσώπου στη μετρική, c είναι μια σταθερά ($0 < c < 0.5$) και $a = \text{mean}(A)$, $b = \text{mean}(M)$ χρησιμοποιούνται για ανεξαρτησία από τις γενικότερες συνθήκες φωτισμού. Αν η τιμή του r είναι χαμηλότερη από κάποιο δοσμένο κατώφλι τότε η περιοχή F γίνεται αποδεικτή ως περιοχή που περιλαμβάνει πρόσωπο. Η βέλτιστη κλίμακα και κλίση

λαμβάνεται από τη σχέση $[U, \Theta] = \arg \min \{r(u, \theta)\}$ και η τελική ανίχνευση πραγματοποιείται χρησιμοποιώντας το πρότυπο $M(U, \Theta)$ και την μετρική της εξίσωσης (3). Η υποπεριοχή A^* που εντοπίζεται αποτελεί την περιοχή προσώπου του συνόλου F .

Η παραπάνω διαδικασία, αν εφαρμοστεί σε μια ολόκληρη εικόνα, και χωρίς κάποια εκ των προτέρων γνώση για την κλίμακα και τη θέση του αναζητούμενου προσώπου είναι πολύ χρονοβόρα. Χρησιμοποιώντας την πληροφορία που προκύπτει από τη διαδικασία απομόνωσης των τμημάτων δέρματος –βλέπε Παράγραφο 3– η βέλτιστη κλίμακα U προσεγγίζεται από τη οριζόντια χωρική ανάλυση h η οποία είναι άνω φραγμένη από το πλάτος του ορθογωνίου που περιλαμβάνει κάποιο τμήμα δέρματος και τη κάθετη ανάλυση v η οποία είναι άνω φραγμένη είτε από το μήκος του ορθογωνίου είτε από τη ποσότητα $2*b$ σύμφωνα με την ανατομία του ανθρώπινου προσώπου. Με την εκτίμηση της βέλτιστης κλίμακας U η υπολογιστική πολυπλοκότητα της διαδικασίας εντοπισμού περιορίζεται σημαντικά. Επιπλέον σε ακολουθίες βίντεο το πρόσωπο ανιχνεύεται στο πρώτο καρέ και η πληροφορία για τη κλίμακα και την κλίση της μπορεί να χρησιμοποιηθεί στα επόμενα καρέ για να ελαχιστοποιήσει την προσπάθεια ανίχνευσης.

6. ΠΕΙΡΑΜΑΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ

Για την επαλήθευση της αποδοτικότητας του αλγορίθμου πραγματοποιήσαμε τέσσερα διαφορετικά πειράματα. Στα δύο πρώτα ελέγχεται η αποδοτικότητα του εντοπισμού προσώπου σε μονόχρωμες εικόνες και ακολουθίες βίντεο. Τα πειράματα αυτά συμπεριλήφθηκαν ώστε να είναι εφικτή η πραγματοποίηση συγκρίσεων όσον αφορά στην υπολογιστική πολυπλοκότητα. Τα άλλα δύο πειράματα αναφέρονται σε έγχρωμες εικόνες και ακολουθίες βίντεο αντίστοιχα.

6.1 Εντοπισμός προσώπου σε μονόχρωμες εικόνες

Στο πείραμα αυτό δεν υπάρχει καμία ένδειξη όσον αφορά τη θέση του προσώπου στην εικόνα ούτε και για την κλίση και κλίμακά του. Ο μόνος τρόπος εντοπισμού είναι η εφαρμογή της διαδικασίας που περιγράφηκε στην Παράγραφο 5, σε όλη την εικόνα.

Υλικό: Τα δεδομένα –φωτογραφίες– που χρησιμοποιήθηκαν στο πείραμα προέρχονται από τη βάση προσώπων CMU, η οποία περιέχει εικόνες από 20 ανθρώπους (άνδρες και γυναίκες). Υπάρχουν 32 διαφορετικές φωτογραφίες χωρικής ανάλυσης 120x128 για κάθε άτομο, οι οποίες περιέχουν διαφορετικές εκφράσεις και στις οποίες τα πρόσωπα βρίσκονται σε διαφορετική κλίση και περιστροφή. Για την επαλήθευση της διαδικασίας ανίχνευσης χρησιμοποιήσαμε μόνο τις φωτογραφίες που κοιτάζουν κατευθείαν στη κάμερα –όχι στραμμένες. Στο ίδιο πείραμα χρησιμοποιήσαμε επίσης και μερικές άλλες εικόνες προσώπων από βίντεοκλιπς και οι οποίες παρουσιάζουν μια σημαντική ποικιλία όσον αφορά το φόντο τους.

Αποτελέσματα: Η αποδοτικότητα του αλγορίθμου επιδεικνύεται στον Πίνακα B2.1. Ο ολικός αριθμός των εικόνων που χρησιμοποιήθηκαν είναι 100 και για τις προσομοιώσεις χρησιμοποιήθηκε ένας υπολογιστής Pentium II / 233 MHz. Δεδομένου ότι σε όλες οι εικόνες περιέχεται ένα μόνο πρόσωπο αποδεχόμαστε ως περιοχή προσώπου την περιοχή με τη μέγιστη ομοιότητα με το πρότυπο –στη εκτιμηθείσα βέλτιστη

κλίμακα και κλίση. Για τον ίδιο λόγο το ποσοστό απώλειας -dismissal rate- είναι το ίδιο με το ποσοστό λανθασμένης ανίχνευσης -false alarm rate.

Dismissal Rate (DR) (%)	False Alarm Rate (FAR) (%)	Mean Execution Time (MET) (sec)
5	5	131.7

Πίνακας B2.1: Εντοπισμός προσώπου σε μονόχρωμες εικόνες

Σχολιασμός: Η αποτελεσματικότητα του αλγορίθμου όσον αφορά τον ορθό εντοπισμό είναι σχεδόν τέλεια. Παρόλα αυτά η όλη διαδικασία είναι εξαιρετικά χρονοβόρα. Αποτυχία ανίχνευσης εμφανίζουν οι εικόνες στις οποίες τα πρόσωπα έχουν γενειάδα ή / και γυαλιά ενώ λανθασμένος εντοπισμός παρουσιάζεται σε περιπτώσεις εικόνων με υψηλό συχνοτικό περιεχόμενο.



Εικόνα B2.2: Παραδείγματα ανίχνευσης προσώπου σε μονόχρωμες εικόνες

6.2 Εντοπισμός προσώπου σε μονόχρωμες ακολουθίες βίντεο

Πρόκειται για μια περίπτωση αντίστοιχη των στατικών εικόνων. Η διαδικασία που περιγράφηκε στη Παράγραφο 5 εφαρμόζεται μόνο στο πρώτο καρέ της ακολουθίας. Στα επόμενα καρέ υπάρχει πλέον σοβαρή ένδειξη για την θέση και κλίμακα του προσώπου –όχι όμως και για την κλίση του η οποία αλλάζει αρκετά γρήγορα. Μετά από κάθε καρέ πραγματοποιείται επανεκτίμηση της κλίμακας και της θέσης του προσώπου, ενώ όταν ο βαθμός ομοιότητας με το πρότυπο στη βέλτιστη κλίμακα αλλάξει απότομα, τότε ενεργοποιείται η διαδικασία της Παραγράφου 5.

Υλικό: Τα δεδομένα –ακολουθίες βίντεο- που χρησιμοποιήθηκαν στο πείραμα προέρχονται κυρίως από βιντεοκλίπς καταγραμμένα από το πρόγραμμα διαφόρων Ελληνικών καναλιών. Το μέγεθος των καρέ είναι σύμφωνα με το CIF format και πριν τη διαδικασία της ανίχνευσης πραγματοποιείται υποδειγματοληψία. Ο ρυθμός εγγραφής ήταν 10 καρέ ανά δευτερόλεπτο. Πρέπει να τονιστεί ότι η πληροφορία χρώματος αγνοήθηκε ώστε να καταλήξουμε σε μονόχρωμες ακολουθίες.

	DR (%)	FAR (%)	MET (sec)
Πρώτο καρέ	10	10	225.4
Επόμενα Καρέ	10	10	5

Πίνακας B2.2: Εντοπισμός προσώπου σε ακολουθίες βίντεο απόχρωσης του γκρι

Αποτελέσματα: Η αποδοτικότητα του αλγόριθμου επιδεικνύεται στον Πίνακα B2.2. Ο ολικός αριθμός των εικόνων που χρησιμοποιήθηκαν είναι 300 -10 ακολουθίες 30 καρέ έκαστη- και για τις προσομοιώσεις χρησιμοποιήθηκε ο ίδιος υπολογιστής όπως στο προηγούμενο πείραμα. Μόνο ένα πρόσωπο εντοπίζεται και παρακολουθείται σε κάθε ακολουθία.

Σχολιασμός: Όπως φαίνεται εμφανώς στον Πίνακα B2.2, με ειμετάλλευση της πληροφορίας για τη θέση και κλίμακα που παρέχεται με την εφαρμογή του αλγορίθμου στο πρώτο καρέ, ο μέσος χρόνος εκτέλεσης μειώνεται περίπου 40 φορές στα επόμενα καρέ. Βέβαια, πρέπει να αναφερθεί ότι στις ακολουθίες που χρησιμοποιήθηκαν δεν είχαμε αλλαγή πλάνου και επομένως η ενεργοποίηση ολόκληρης της διαδικασίας δεν χρειάστηκε σε κανένα ενδιάμεσο στάδιο. Ορθή ανίχνευση του προσώπου στο πρώτο καρέ εξασφαλίζει το σωστό εντοπισμό του και στα επόμενα, ενώ αντίστοιχα αποτυχία ανίχνευσης στο πρώτο καρέ παραπλανά και τη διαδικασία στα επόμενα.



Εικόνα B2.3: Παραδείγματα εντοπισμού προσώπου σε μονόχρωμες ακολουθίες βίντεο

6.3 Εντοπισμός προσώπου σε έγχρωμες φωτογραφίες

Ολόκληρη η διαδικασία όπως περιγράφηκε στις Παραγράφους 2 έως 5, εφαρμόστηκε κατά την εκτέλεση του πειράματος αυτού. Πρώτα εντοπίζονται τα τμήματα δέρματος και στη συνέχεια σε κάθε τμήμα ξεχωριστά εφαρμόζεται η διαδικασία της ανίχνευσης με τη χρήση ταύτισης με το πρότυπο.

Πηγή: Τα δεδομένα –έγχρωμες φωτογραφίες- που χρησιμοποιήθηκαν στο πείραμα προέρχονται από βιντεοκλίπ τηλεοπτικών προγραμμάτων με χωρική ανάλυση σύμφωνη με το CIF format. Σε κάθε καρέ μπορούν να περιέχονται περισσότερα από ένα πρόσωπα.

Αποτελέσματα: Η απόδοση του αλγορίθμου επιδεικνύεται στον Πίνακα B2.3. Χρησιμοποιήθηκε ένα σύνολο από 100 εικόνες οι οποίες περιείχαν 145 πρόσωπα σε ικανοποιητική χωρική ανάλυση –πολύ μικρά πρόσωπα αγνοήθηκαν. Δεδομένου ότι περισσότερα από ένα πρόσωπα μπορούν να περιέχονται σε κάθε εικόνα, τα ποσοστά απώλειας ανίχνευσης και λανθασμένης ανίχνευσης είναι διαφορετικά. Η Εικόνα B2.5 επιδεικνύει μια εικόνα στην οποία έχουν εντοπιστεί τρία τμήματα δέρματος, δύο από τα οποία επιβεβαιώνεται ότι αντιστοιχούν σε πρόσωπα. Πρέπει να τονιστεί ότι μερικά μικρά τμήματα δέρματος τα οποία αντιστοιχούν σε πρόσωπα δεν εντοπίστηκαν στην παραπάνω εικόνα. Αυτό οφείλεται στην επιλογή μας να θεωρήσουμε ως αναγνωρίσιμα, πρόσωπα με χωρική ανάλυση 24x24 pixel τουλάχιστον, η οποία δεν απέχει από την ελάχιστη αναγνωρίσιμη χωρική ανάλυση που προκύπτει από μελέτες στην αναγνώριση προσώπων και αναφέρεται στο πρώτο μέρος της διατριβής.

Σχολιασμός: Η αποτυχία εντοπισμού όλων των τμημάτων δέρματος είναι ο βασικότερος παράγοντας για τις απώλειες ανίχνευσης, ενώ λανθασμένοι εντοπισμοί οφείλονται βασικά στη διαδικασία ταύτισης με πρότυπο και κύρια στη κατωφλίωση. Πρέπει να τονιστεί πάντως ότι σε αραιές περιπτώσεις ο συνδυασμός του εντοπισμού του δέρματος και της ταύτισης με το πρότυπο αυξάνει το συνολικό ποσοστό επιτυχημένων ανιχνεύσεων. Η Εικόνα B2.4(β) παρουσιάζει ένα τέτοιο παράδειγμα. Εφαρμόζοντας ταύτιση με πρότυπο σε ολόκληρη την εικόνα –μετά από μετατροπή της σε μονόχρωμη- οδηγούμαστε σε αποτυχία ανίχνευσης του πολύ έντονα φωτισμένου προσώπου. Το μεγάλο ορθογώνιο –κόκκινο- που περιβάλλει το τμήμα δέρματος της Εικόνας B2.4(β) περιορίζει την περιοχή που πρέπει να ελεγχθεί και προστατεύει από λανθασμένη ανίχνευση εξαιτίας του υψίσυχνου περιεχομένου της υπόλοιπης εικόνας. Όπως συνάγεται από τον Πίνακα B2.3 η σημαντικότερη πτώση του μέσου χρόνου εκτέλεσης είναι το ισχυρότερο πλεονέκτημα της χρήσης του συνδυασμού των δύο τεχνικών.

DR (%)	FAR (%)	MET (sec)
4	2	1.7

Πίνακας B2.3: Εντοπισμός προσώπου σε έγχρωμες φωτογραφίες



Εικόνα B2.4: Παραδείγματα εντοπισμού προσώπου σε έγχρωμες φωτογραφίες



Εικόνα B2.5: Εντοπισμός πολλαπλών προσώπων

6.4 Εντοπισμός προσώπου σε έγχρωμες ακολουθίες βίντεο

Το πείραμα αυτό είναι αντίστοιχο με αυτό των έγχρωμων στατικών εικόνων. Η διαφορά βρίσκεται στο γεγονός ότι ο μέσος χρόνος εκτέλεσης μειώνεται περαιτέρω, δεδομένου ότι δεν πραγματοποιείται ψάξιμο για τα τμήματα δέρματος παρά μόνο στο πρώτο καρέ, ενώ στα επόμενα γίνεται παρακολούθηση των

τμημάτων αυτών. Μόνο όταν η πιθανότητα κάποιου από τα παρακολουθούμενα τμήματα να περιέχει πρόσωπο μειωθεί σημαντικά, ενεργοποιείται η διαδικασία εντοπισμού των τμημάτων δέρματος.

Υλιό: Χρησιμοποιήθηκαν οι ίδιες ακολουθίες με αυτές του πειράματος Β –Παράγραφος 6.2- με κάποιες επιπλέον προσθήκες που περιελάμβαναν αλλαγές πλάνων.

Αποτελέσματα: Η απόδοση του αλγορίθμου επιδεικνύεται στον Πίνακα Β2.4. Χρησιμοποιήθηκαν συνολικά 600 εικόνες -20 ακολουθίες με 30 καρτέ έιαστη. Τα τμήματα δέρματος που ανιχνεύονται στο πρώτο καρτέ ελέγχονται ως προς το περιεχόμενο τους μέσω ταύτισης με το πρότυπο. Στα επόμενα καρτέ παρακολουθούνται μόνο εκείνα τα τμήματα που έχει επαληθευτεί ότι περιέχουν πρόσωπο.

	DR (%)	FAR (%)	MET (sec)
Πρώτο καρτέ	4	2	2.1
Επόμενα Καρτέ	4	2	1.4

Πίνακας Β2.4: Εντοπισμός προσώπου σε έγχρωμες ακολουθίες βίντεο

Σχολιασμός: Όπως και στο πείραμα Β επιτυχημένη ανίχνευση στο πρώτο καρτέ εξασφαλίζει την επιτυχημένη ανίχνευση και στα επόμενα ενώ το ανάποδο συμβαίνει σε απώλεια εντοπισμού στο πρώτο καρτέ. Εκμετάλλευση της χρονικής συσχέτισης ανάμεσα στα καρτέ της ακολουθίας οδηγεί σε μείωση του μέσου χρόνου εκτέλεσης μετά το πρώτο καρτέ κατά 25%.

7. ΣΥΜΠΕΡΑΣΜΑΤΑ

Η προτεινόμενη στο κεφάλαιο αυτό τεχνική παρέχει μια αποδοτική μεθοδολογία για την ανίχνευση προσώπων σε εικόνες και ακολουθίες βίντεο και μπορεί να χρησιμοποιηθεί εκτός των εφαρμογών επικοινωνίας ανθρώπου μηχανής και σε μια πλειάδα από εφαρμογές πολυμέσων όπως η κατάκτηση και περιήληψη ακολουθιών βίντεο, η πλοήγηση σε βάσεις δεδομένων εικόνων και βίντεο, η δεικτοδότηση και περιήληψη τηλεοπτικών ειδήσεων, η αλληλοδραστική ανάληψη, από βάσεις πολυμέσων, με βάση το περιεχόμενο και. Εμπλέκει απλά χαρακτηριστικά χρώματος σε συνδυασμό με πληροφορία υψής. Η αποτελεσματικότητα της μεθόδου αυξάνεται με τη χρήση μιας διαδικασίας ταύτισης με πρότυπο με κόστος την αύξηση της υπολογιστικής πολυπλοκότητας. Παρόλα αυτά, όπως επιδείχθηκε με μια σειρά από πειράματα, η υπολογιστική πολυπλοκότητα μειώνεται έως και 100 φορές σε σχέση με την αμιγή εφαρμογή της ταύτισης με πρότυπο, εάν εφαρμοστεί ο συνδυασμός της ανίχνευσης τμημάτων δέρματος και της ταύτισης με πρότυπο.

ΑΝΑΛΥΣΗ ΔΕΛΤΙΩΝ ΕΙΔΗΣΕΩΝ ΜΕ ΑΝΙΧΝΕΥΣΗ ΠΡΟΣΩΠΩΝ

Η ανάλυση και δεικτοδότηση μεταδιδόμενων ή ψηφιοποιημένων αποθηκευμένων αρχείων από δελτία ειδήσεων αποτελεί ένα σημαντικό εργαλείο στο πλαίσιο πολλών αλληλοδραστικών εφαρμογών πολυμέσων όπως η ανάγλυση με βάση το περιεχόμενο, η εκπομπή ειδήσεων κατά απαίτηση (*news-on-demand –NoD*), η πλοήγηση σε βάσεις δεδομένων και η αυτόματη εξαγωγή περίληψης ειδήσεων. Οι τρέχουσες τεχνικές στην ανάλυση δελτίων ειδήσεων είναι μεν αποδοτικές αλλά στηρίζονται σε μεγάλο ποσοστό σε πληροφορία κειμένου –οπότιτλοι ή *teletext*. Στο κεφάλαιο αυτό παρουσιάζεται μια τεχνική χρονικής ανάλυσης ειδήσεων με βάση το οπτικό περιεχόμενο η οποία στηρίζεται στο γεγονός ότι οι σκηνές στο *studio* περιλαμβάνουν πάντοτε τον παρουσιαστή. Εφαρμόζοντας προσεγγίσεις εντοπισμού προσώπων όπως αυτές παρουσιάστηκαν στα προηγούμενα κεφάλαια μπορούμε να διαχωρίσουμε της σκηνές του *studio* από τις αντίστοιχες εξωτερικές. Η προτεινόμενη τεχνική μπορεί να χρησιμοποιηθεί είτε ως κύρια είτε σε συνδυασμό με αντίστοιχες τεχνικές που στηρίζονται στην αναγνώριση φωνής ή σε πληροφορία κειμένου.

Από τα συμπεράσματα της εργασίας που παρουσιάζεται στο κεφάλαιο αυτό μπορεί να γίνει αντιληπτό μέχρι ποιον σημείου και κάτω από ποιες συνθήκες μπορεί κάποιος να αναμένει αξιόπιστο εντοπισμό του προσώπου σε ακολουθίες βίντεο. Επιπλέον γίνεται νύξη πως μπορεί να παρακολουθηθεί η κίνηση του προσώπου σε διαδοχικά καρέ και πως κάποια άλλα στοιχεία όπως η μεταβολή ιστογράμματος μπορούν να δράσουν επικουρικά στη διαδικασία αυτή.

1. ΕΙΣΑΓΩΓΗ

Η μαζική παραγωγή δεδομένων πολυμέσων από πληθώρα πηγών όπως τα τηλεοπτικά προγράμματα, οι κινηματογραφικές ταινίες, τα ψηφιοποιημένα αρχεία και οι αλληλοδραστικές ιστοσελίδες, θέτουν την ανάγκη ανάπτυξης νέων εργαλείων για αυτόματη ανάλυση, τμηματοποίηση και δεικτοδότηση με βάση το περιεχόμενο [13]. Τέτοια εργαλεία είναι σχεδόν απαραίτητα σε πολυμεσικές εφαρμογές όπως η αυτόματη εξαγωγή περίληψης, η ανάγλυση με βάση το περιεχόμενο και η πλοήγηση, όπως αυτές προσδιορίζονται από τα σύγχρονα πρότυπα κωδικοποίησης MPEG-4 και MPEG-7 [17].

Τα «ζωντανά» τηλεοπτικά δελτία ειδήσεων και τα αντίστοιχα ψηφιοποιημένα αρχεία είναι δύο πολύ σημαντικές πηγές πολυμεσικών δεδομένων για δύο λόγους: (α) Αμφότερες οι οπτικές και ακουστικές πηγές ειδήσεων μεταδίδουν μη δομημένα, σε σχέση με το περιεχόμενο, δεδομένα, όπως κάθε οπτικοακουστική πηγή [22]. Παρόλα αυτά, η τυπική οργάνωση των ειδήσεων σε μικρότερα τμήματα στα οποία εμπλέκονται σχεδόν διαδοχικά ο παρουσιαστής, κάποιος εξωτερικός *reporter* και εξωτερικές σκηνές, επιτρέπουν ευκολότερο διαχωρισμό σε χρονικά τμήματα τα οποία είναι νοηματικά αυτόνομα. Τα χρονικά διαμερισμένα τμήματα μπορούν σε επόμενο στάδιο να χρησιμοποιηθούν για εξαγωγή πρωτογενών οπτικοακουστικών χαρακτηριστικών και νοηματικής περιγραφής. (β) Πλήρως αυτοματοποιημένα συστήματα δεικτοδότησης και σχολιασμού ειδήσεων αρχείου μπορούν να χρησιμοποιηθούν από αναλυτές δεδομένων σε κρατικές υπηρεσίες ή ραδιοτηλεοπτικούς σταθμούς, από εταιρίες παροχής υπηρεσιών και τέλος από τους απλούς χρήστες ραδιοτηλεοπτικών υπηρεσιών [19]. Ο χαρακτηρισμός ψηφιακών ειδήσεων

σε αρχεία γίνεται στην παρούσα φάση με χειρονακτικές μεθόδους οι οποίες είναι και χρονοβόρες και δαπανηρές.

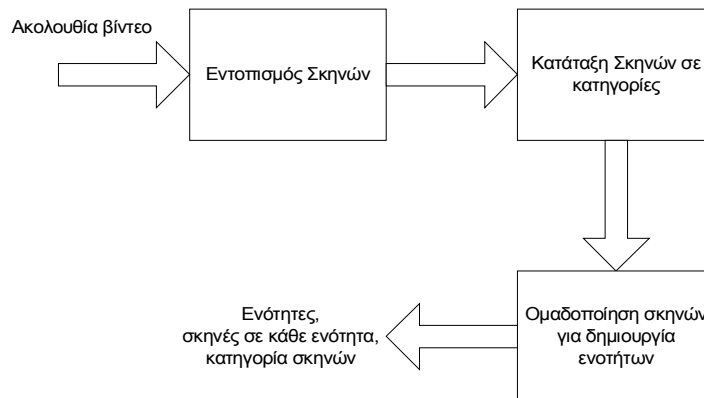
Διάφορα πρωτότυπα συστήματα για αυτόματο ή ημιαυτόματο διαμερισμό και σχολιασμό ψηφιακών αρχείων ειδήσεων έχουν εμφανιστεί στη διεθνή βιβλιογραφία τα τελευταία χρόνια. Τα συστήματα αυτά απευθύνονται σε εφαρμογές όπως η αλληλοδραστική πλοήγηση σε βάσεις ψηφιακού βίντεο, η εκπομπή ειδήσεων κατά απαίτηση (NoD), και η ανάκληση με βάση το περιεχόμενο [15][18][20][22]. Οι περισσότερες από τις τεχνικές αυτές στηρίζονται σε μεγάλο βαθμό σε δεδομένα κειμένου ή γλωσσικά στοιχεία τα οποία μεταδίδονται μέσω κειμένων teletext ή υποτίτλων. Η χρήση αμιγώς οπτικοακουστικών στοιχείων για τις παραπάνω εφαρμογές είναι καθολικά αποδεκτό ότι αποτελεί την τάση του μέλλοντος [21]. Επιπλέον η έλλειψη πληροφορίας κειμένου από ένα μεγάλο αριθμό από αρχειοθετημένα δελτία ειδήσεων παρελθόντων ετών δεν μπορεί σε καμία περίπτωση να αγνοηθεί. Ερευνητές από το πεδίο της τεχνητής νοημοσύνης θα κληθούν να ενοποιήσουν όλες τις περιγραφές περιεχομένου είτε αυτές βασίζονται σε οπτικοακουστικά στοιχεία είτε δεδομένα κειμένου με τελικό στόχο την νοηματική κατάτμηση των τηλεοπτικών δελτίων ειδήσεων. Οι τεχνικές που στηρίζονται σε οπτικοακουστικά στοιχεία για την κατάτμηση ειδήσεων περιορίζονται στην ανίχνευση σιωπής ή μύρης οθόνης [19] και η ολοκλήρωση των στοιχείων για την λήψη κάποιας απόφασης, όσον αφορά στον χρονικό διαμερισμό, επιτυγχάνεται με τεχνικές δυναμικού προγραμματισμού [20], με χρήση μηχανών πεπερασμένων καταστάσεων [19] και Hidden Markov Models [16].

Στο κεφάλαιο αυτό παρουσιάζεται μια τεχνική για χρονικό διαμερισμό τηλεοπτικών ειδήσεων, με βάση οπτικά στοιχεία, η οποία στηρίζεται στο γεγονός ότι τα δελτία ειδήσεων αναλώνονται σε μεγάλο βαθμό στην καταγραφή της ανθρώπινης δραστηριότητας και η παρουσία προσώπων σε αυτά είναι κυρίαρχη. Πρόσωπα σε δελτία ειδήσεων εμφανίζονται σε τρεις κατηγορίες σκηνών: (α) πλάνα στο studio στα οποία υπάρχουν ελεγχόμενες συνθήκες λήψης, λίγα πρόσωπα, με ικανοποιητική χωρική ανάλυση, και σε κοντινή απόσταση από την κάμερα, (β) πλάνα από ρεπορτάζ, στα οποία σχεδόν πάντα υπάρχει ο ρεπόρτερ και κάποιο άλλο πρόσωπο ταυτόχρονα ή σε εναλλαγή, και στα οποία οι συνθήκες λήψης δεν είναι οι ιδανικές, ενώ τα πρόσωπα μπορούν να έχουν οποιαδήποτε οριοθέτηση, (γ) εξωτερικά πλάνα στα οποία δεν είναι υποχρεωτικό να υπάρχουν πρόσωπα και στα οποία οι συνθήκες λήψης είναι τελείως απροσδιόριστες.

Όπως φαίνεται από τα προηγούμενα σχόλια η ευκολία ανίχνευσης των προσώπων στις τρεις παραπάνω κατηγορίες σκηνών είναι φθίνουσα, γεγονός που επιβεβαιώνεται από τα πειραματικά αποτελέσματα που παρατίθενται στο τέλος του κεφαλαίου. Εκτός από την ανίχνευση προσώπου η προτεινόμενη τεχνική κατάτμησης ειδήσεων χρησιμοποιεί την πρότερη γνώση, για τη διαμόρφωση των δελτίων, υπό μορφή κανόνων καθώς και κάποια απλά χαρακτηριστικά για ανίχνευση αλλαγής πλάνου, παρακολούθηση της κίνησης και μεταβολή στο μέγεθος των προσώπων. Η τεχνική μπορεί να εφαρμοστεί είτε αυτόνομα είτε σε συνδυασμό με τεχνικές που χρησιμοποιούν πληροφορία κειμένου ή ακουστικά στοιχεία.

2. ΑΝΑΛΥΣΗ ΤΟΥ ΠΡΟΤΕΙΝΟΜΕΝΟΥ ΣΥΣΤΗΜΑΤΟΣ

Το διάγραμμα ροής του προτεινόμενου αλγόριθμου επιδεικνύεται στο Σχήμα Β3.1. Η είσοδος του συστήματος είναι μια ψηφιακή ακολουθία βίντεο από κάποιο δελτίο ειδήσεων και η έξοδος είναι ένα σύνολο από ενότητες για τις οποίες καταγράφεται ο χρόνος έναρξης και τέλους και για κάθε ενότητα το σύνολο των πλάνων που την απαρτίζουν και σε ποια κατηγορία υπάγεται κάθε σκηνή.



Σχήμα Β3.1: Το μπλοκ διάγραμμα της τεχνικής ανάλυσης τηλεοπτικών δελτίων ειδήσεων.

Η πρότερη γνώση της διάρθρωσης των τηλεοπτικών δελτίων ειδήσεων είναι κρίσιμος παράγοντας για την ανάλυσή τους σε επιμέρους τμήματα. Για το σκοπό αυτό ορίζουμε τις παρακάτω έννοιες:

Θεματική ενότητα: Το σύνολο των πλάνων που αναφέρονται στο ίδιο θέμα του δελτίου ειδήσεων. Για παράδειγμα, μια θεματική ενότητα που αναφέρεται σε κάποια αεροπειρατεία μπορεί να απαρτίζεται από κάποιο εισαγωγικό πλάνο στο studio, στο οποίο ο παρουσιαστής αρχίζει την παρουσίαση του θέματος, κάποια πλάνα από ρεπόρτερ, πλάνα αρχείου, εξωτερικά πλάνα και μακρινή λήψη του αεροσκάφους, επιστροφή στο studio κοκ.

Ενότητα: Το σύνολο των πλάνων μεταξύ δύο διαδοχικών εμφανίσεων του παρουσιαστή στο studio – συμπεριλαμβανομένου του πρώτου πλάνου με τον παρουσιαστή. Σε πολλές περιπτώσεις η διάρκεια των ενότητων συμπίπτει με τη διάρκεια των θεματικών ενότητων. Σε αρκετές πάντως περιπτώσεις η διάρκεια μιας ενότητας είναι μικρότερη από την αντίστοιχη της θεματικής ενότητας ενώ σπανιότερα μια θεματική ενότητα μπορεί να μην έχει εξωτερικά πλάνα ή πλάνα με ρεπόρτερ και επομένως αποτελεί υποσύνολο μιας ενότητας. Το μεγαλύτερο πρόβλημα σε σχέση με τον ορισμό των ενότητων και των θεματικών ενότητων προκύπτει σε περιπτώσεις συζητήσεων υπό μορφή παραθύρων στις οποίες έχουμε εναλλαγές κοντινών πλάνων όσον αφορά τους συμμετέχοντες στη συζήτηση. Προφανώς ένας βασικός στόχος είναι οι ενότητες να συμπίπτουν με τις θεματικές ενότητες. Πάντως δεν είναι αποτυχία και οι ενότητες να αποτελούν υποσύνολο των θεματικών ενότητων.

Σε μια χονδρική διαίρεση τα πλάνα από κάποιο δελτίο ειδήσεων μπορούν να καταταγούν σε τέσσερις κατηγορίες:

Πλάνα με τον παρουσιαστή στο studio (anchorman shots): Τα βασικά χαρακτηριστικά των πλάνων αυτών είναι:

- Μη μεταβαλλόμενο φόντο
- Ύπαρξη προσώπου με σχετικά μεγάλη χωρική ανάλυση, στο κέντρο του καρέ, το οποίο κοιτάζει ευθεία στη κάμερα
- Μικρή μεταβολή στη θέση του προσώπου ανάμεσα σε διαδοχικά καρέ
- Σχετικά μεγάλη χρονική διάρκεια

Πλάνα με ρεπόρτερ (report/interview shots): Τα βασικά χαρακτηριστικά των πλάνων αυτών είναι:

- Ύπαρξη προσώπου σε γενικά αποδεκτή ανάλυση το οποίο δεν κοιτάζει υποχρεωτικά στη κάμερα
- Συχνές εναλλαγές ανάμεσα στο πρόσωπο του ρεπόρτερ και σε άλλα πρόσωπα
- Μεταβαλλόμενο φόντο –εξωτερικά ρεπορτάζ- ή φόντο που αλλάζει μεταξύ δύο πιθανών καταστάσεων –ρεπορτάζ ή συνέντευξη σε κλειστό χώρο.

Στατικά πλάνα (static shots): Τα βασικά χαρακτηριστικά των πλάνων αυτών είναι:

- Δεν υπάρχουν πρόσωπα σε υψηλή χωρική ανάλυση
- Σταθερό φόντο

Τυπικά σταθερά πλάνα είναι το δελτίο καιρού και οι οικονομικές αναλύσεις.

Εξωτερικά πλάνα (outdoor shot): Τα βασικά χαρακτηριστικά των πλάνων αυτών είναι:

- Απροσδιόριστο και μεταβαλλόμενο φόντο
- Μη υποχρεωτική ύπαρξη προσώπου
- Ταχεία κίνηση προσώπων ανάμεσα σε διαδοχικά καρέ (εφόσον υπάρχουν)
- Μικρή διάρκεια

Το πρώτο στάδιο του αλγορίθμου αφορά τον εντοπισμό των πλάνων και μπορεί να πραγματοποιηθεί με πολύπλοκες τεχνικές που έχουν αναπτυχθεί για το σκοπό αυτό [14]. Παρόλα αυτά μια απλή και γρήγορη λύση για την ανίχνευση αλλαγής πλάνου μπορεί να πραγματοποιηθεί με βάση την κίνηση του φόντου όπως αυτή υπολογίζεται από την απόλυτη διαφορά ανάμεσα σε διαδοχικά καρέ. Η εφαρμογή κάποιου κατάλληλα προσδιορισμένου κατωφλίου μπορεί να διαχωρίσει δύο πλάνα.

Η ταξινόμηση ενός πλάνου σε κάποια κατηγορία γίνεται με βάση την πλειονότητα των καρέ. Για παράδειγμα αν για κάποιο πλάνο ο μεγαλύτερος αριθμός των καρέ ταξινομήθηκε ως καρέ με παρουσιαστή στο studio τότε και το πλάνο υπάγεται στην ίδια κατηγορία.

2.1 Βαθμός μεταβολής ιστογράμματος

Σε ένα τηλεοπτικό δελτίο ειδήσεων τα καρέ που αντιστοιχούν σε πλάνα στο studio αποτελούν ένα μεγάλο μέρος των συνολικών καρέ της ακολουθίας. Ένα βασικό χαρακτηριστικό των καρέ αυτών είναι το σταθερό φόντο, χαρακτηριστικό που μπορεί να χρησιμοποιηθεί για την διάκριση μεταξύ πλάνων στο studio και πλάνων με ρεπόρτερ σε εσωτερικό χώρο –μη μεταβαλλόμενο φόντο. Για το σκοπό αυτό ορίζουμε τον βαθμό μεταβολής ιστογράμματος σύμφωνα με τα παρακάτω:

Έστω το καρέ I με χρωματικές συνιστώσες $R_p, G_p, B_p \in \mathfrak{R}^{n \times m}$ για τις οποίες ισχύει $l_R \leq R(i, j) \leq h_R$, $l_G \leq G(i, j) \leq h_G$, $l_B \leq B(i, j) \leq h_B$, $\forall i, j$, και η διαμέριση $\{S_1, S_2, \dots, S_N\}$ για την οποία ισχύει $\min\{l_R, l_G, l_B\} \leq \min\{S_1\}$, και $\max\{S_N\} \leq \max\{h_R, h_G, h_B\}$.

Ορίζουμε το διάνυσμα v_I χαρακτηριστικών για το καρέ I ως:

$$v_I = [n_1(R_I) \ n_2(R_I) \dots n_N(R_I) \ n_1(G_I) \ n_2(G_I) \dots n_N(G_I) \ n_1(B_I) \ n_2(B_I) \dots n_N(B_I)]^T$$

όπου $n_i(k)$ είναι το πλήθος των pixels της συνιστώσας $k \in \{R_p, G_p, B_p\}$ που αντιστοιχεί στην κλάση S_i .

Ο βαθμός μεταβολής ιστογράμματος ανάμεσα στα καρέ I και J δίνεται από τη σχέση:

$$d = \|v_I - v_J\|.$$

2.2 Μεταβολή μεγέθους και μετατόπιση τμημάτων

Η μετατόπιση του τμήματος με τη μέγιστη πιθανότητα προσώπου μπορεί να χρησιμοποιηθεί για τον περιορισμό των false alarms και για την διαφοροποίηση πλάνων studio στα οποία εμφανίζονται δύο παρουσιαστές. Στη τελευταία περίπτωση μπορεί να μην υπάρχει εμφανής μεταβολή στο ιστόγραμμα οπότε η μεταβολή στο μέγεθος και στη θέση του κυρίαρχου τμήματος δέρματος είναι μια σοβαρή ένδειξη.

Κατά τη διαδικασία κατάτμησης του καρέ σε τμήματα υπάρχει πάντα διαθέσιμο το μέγεθος του κάθε τμήματος. Η μετατόπιση ενός τμήματος υπολογίζεται από τη μεταβολή στη θέση του κέντρου βάρους του ανάμεσα σε δύο διαδοχικά καρέ. Με δεδομένο ότι δεν ενδιαφέρει η στρόφη ως προς τον άξονα z η μετατόπιση του κέντρου βάρους είναι αρκετή για τον υπολογισμό της αλλαγής στη θέση κάποιου τμήματος.

3. ΕΝΤΟΠΙΣΜΟΣ ΚΑΙ ΠΕΡΙΓΡΑΦΗ ΠΡΟΣΩΠΩΝ ΣΤΑ ΚΑΡΕ

Ένα κρίσιμο στοιχείο για τον χαρακτηρισμό των καρέ και την κατάταξη των σκηνών σε κατηγορίες είναι η παρουσία του προσώπου και πληροφορίες σχετικές με αυτό. Η διαδικασία εντοπισμού του προσώπου είναι όμοια με αυτήν που περιγράφηκε στο προηγούμενο κεφάλαιο με εξαίρεση την παράλειψη του σταδίου επαλήθευσης μέσω της ταύτισης με πρότυπο. Η παράλειψη του σταδίου επαλήθευσης έγινε για αύξηση της ταχύτητας εκτέλεσης και ως αποτέλεσμα τα εντοπισμένα τμήματα είναι στην ουσία τμήματα δέρματος και

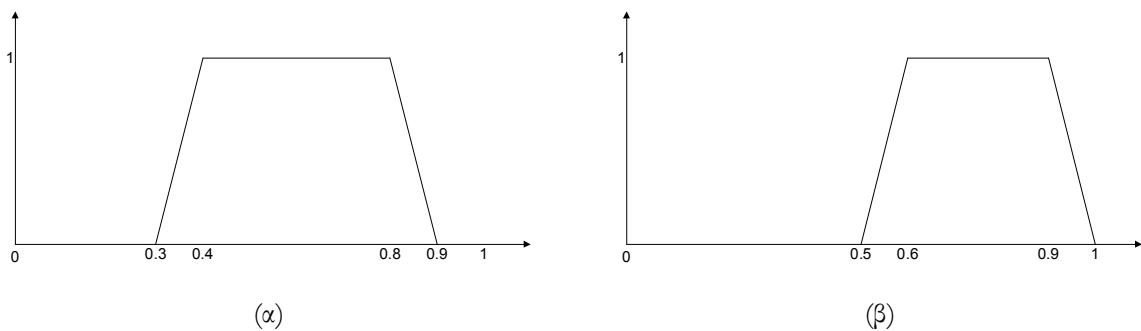
όχι πρόσωπα. Παρόλα αυτά, το σχήμα των τμημάτων λαμβάνεται υπόψη για περιοριστεί η πιθανότητα να ειληφθεί ως πρόσωπο κάποιο τμήμα του ανθρωπίνου σώματος ή κάποιο άσχετο αντικείμενο.

Για κάθε διαχειρισμένο τμήμα S_i που προκύπτει από τη διαδικασία εντοπισμού προσώπου υπολογίζονται τα επόμενα χαρακτηριστικά

- Το κέντρο μάζας του $m(S_i)$
- Το εμβαδόν του $A(S_i)$ (άθροισμα των pixels που καταλαμβάνει)
- Η πιθανότητα δέρματος $c(S_i)$ δέρματος με βάση το γκαουσιανό μοντέλο για το χρώμα δέρματος. Συγκεκριμένα αν $\mathbf{x}(S_i)$ είναι το διάνυσμα που εκφράζει τη μέση τιμή των χρωματικών συνιστωσών C_r, C_b του τμήματος S_i τότε η πιθανότητα δέρματος δίνεται από τη σχέση:

$$c(S_i) = \exp \left\{ -\frac{1}{2} (\mathbf{x}(S_i) - \boldsymbol{\mu}_0)^T \mathbf{C}^{-1} (\mathbf{x}(S_i) - \boldsymbol{\mu}_0) \right\} \quad (\text{B3.1})$$

- Η πιθανότητα προσώπου $f(S_i)$ η οποία αποτελεί συνδυασμό της πιθανότητας δέρματος $c(S_i)$ και χαρακτηριστικών σχήματος. Συγκεκριμένα αν $l(S_i)$ και $g(S_i)$ είναι οι τιμές επιμήκυνσης και κυκλικότητας του τμήματος S_i αντίστοιχα, χρησιμοποιούμε τις συναρτήσεις συμμετοχής του Σχήματος B3.2 για τον υπολογισμό των τιμών $l'(S_i)$ και $g'(S_i)$. Η πιθανότητα προσώπου $f(S_i)$ δίνεται από τη σχέση: $f(S_i) = \sqrt{c(S_i) \cdot g'(S_i) \cdot l'(S_i)}$



Σχήμα B3.2: Συναρτήσεις συμμετοχής για την (α) κυκλικότητα και (β) επιμήκυνση προσώπου.

Η συσχέτιση κάθε τμήματος με μια πιθανότητα προσώπου είναι στοιχείο το οποίο χρησιμοποιείται άμεσα για την κατάταξη των επιμέρους καρτέ σε μια από τις κατηγορίες που ορίστηκαν στην προηγούμενη παράγραφο. Τα υπόλοιπα χαρακτηριστικά αξιοποιούνται σε άλλα μέρη του αλγορίθμου και σχολιάζονται σε επόμενες παραγράφους.

Μετά την κατασκευή του χάρτη πιθανότητας προσώπου –υπολογισμός της πιθανότητας προσώπου για κάθε ένα από τα τμήματα δέρματος που υπάρχουν στο καρτέ- για κάθε καρτέ της ακολουθίας βίντεο, η ανίχνευση κοντινών πλάνων προσώπου είναι άμεσα εφικτή και αποτελεί μια καλή ένδειξη για την παρουσία παρουσιαστών, ρεπόρτερ και ανθρώπων από τους οποίους λαμβάνεται συνέντευξη. Συγκεκριμένα το

μέγεθος των εντοπισμένων προσώπων χρησιμοποιείται ως στοιχείο για την απομάκρυνση των μικρότερων τμημάτων δέρματος, τα οποία μπορεί να αντιστοιχούν σε πρόσωπα στο φόντο, χέρια ή άλλα άσχετα αντικείμενα. Μετά την απομάκρυνση των μικρότερων τμημάτων συνήθως παραμένουν το πολύ τέσσερα πρόσωπα σε κοντινή λήψη σε κάθε καρέ.

4. ΠΕΙΡΑΜΑΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ

Για τα πειράματα χρησιμοποιήσαμε βίντεο με δελτία ειδήσεων από 4 κανάλια (A5, ET1, MEGA και ANT1) τα οποία εκφράζουν διαφορετικές συνθήκες στο studio και φυσικά διαφορετικούς παρουσιαστές. Η διάρκεια των βίντεο είναι 10 λεπτά έκαστο με δειγματοληψία 10 καρέ ανά δευτερόλεπτο και χωρική ανάλυση 384x288x24bpp. Επίδειξη της ανίχνευσης προσώπου –σε μια σκηνή studio- γίνεται στην Εικόνα B3.1 Οι καλές συνθήκες λήψης και το σχετικά ομοιόμορφο φόντο επιτρέπουν τον επακριβή εντοπισμό του προσώπου με χρήση μόνο πληροφορίας χρώματος και σχήματος.



(α)



(β)

Εικόνα B3.1: Εντοπισμός προσώπου (α) Αρχικό καρέ, (β) κατάτμηση με βάση το χρώμα και με ενσωμάτωση χαρακτηριστικών χρώματος.

Χρονική κατάτμηση σε ένα τυπικό απόσπασμα μια ακολουθίας βίντεο επιδεικνύεται στην Εικόνα B3.2. Η συνολική διάρκεια του αποσπάσματος είναι 100 δευτερόλεπτα και περιέχει 15 διαχειριζόμενα πλάνα ένα από τα οποία αντιστοιχεί στον παρουσιαστή. Στην Εικόνα B3.3(α) επιδεικνύεται η μέγιστη πιθανότητα προσώπου σε σχέση με τον αριθμό των καρέ και πριν τη διαδικασία απομάκρυνσης των μικρών τμημάτων δέρματος. Ομοίως στην Εικόνα B3.3(β) παρουσιάζεται η καμπύλη πιθανότητας για το κυρίαρχο –μεγαλύτερο- τμήμα δέρματος μετά την απομάκρυνση των μικρότερων τμημάτων δέρματος. Παρατηρούμε ότι αρκετά καρέ τα οποία περιείχαν μικρά τμήματα δέρματος, με μεγάλη πιθανότητα να αντιστοιχούν σε πρόσωπα, έχουν παραληφθεί και μαζί με αυτά έχει μειωθεί η πιθανότητα των *false alarms*. Η εφαρμογή ενός φίλτρου ενδιάμεσης τιμής –καμπύλη Εικόνας B3.3(γ)- απαλείφει πλάνα που περιέχουν μεν πρόσωπα αλλά έχουν περιορισμένη διάρκεια για να θεωρηθούν ως αυτόνομα. Για παράδειγμα τα πλάνα με τον παρουσιαστή στο studio σπανίως έχουν διάρκεια μικρότερη από δέκα δευτερόλεπτα. Κατάλληλη επιλογή του παραθύρου για το φίλτρο ενδιάμεσης τιμής μπορεί να αντισταθμίσει την ελάχιστη αποδεικτική διάρκεια για κάποιο πλάνο. Γενικά μπορεί κάποιος να ορίσει τρία είδη πλάνων σε σχέση με την παρουσία προσώπων σε αυτά: (α) Πλάνα στα οποία εντοπίζονται πρόσωπα χωρίς να υπάρχουν –κάποια τμήματα με σχήμα και χρώμα που ομοιάζει με αυτό του προσώπου εκλαμβάνονται ως πρόσωπα- και τα οποία οδηγούν σε *false alarms*, (β) πλάνα στα οποία ορθώς εντοπίζονται πρόσωπα και, (γ) διαδοχικά πλάνα στα οποία

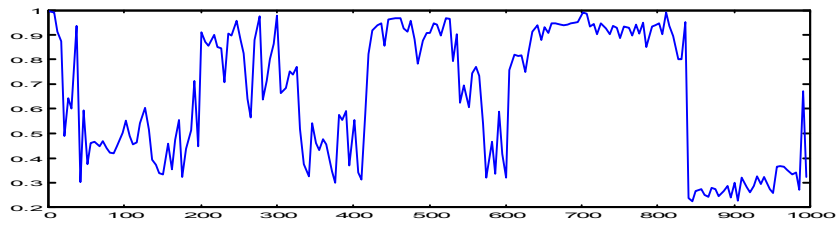
εναλλάσσονται ο παρουσιαστής με κάποιο άλλο πρόσωπο και στα οποία το μέγεθος και η θέση των δύο αυτών προσώπων συμπίπτει. Ένα τέτοιο γεγονός μπορεί να οδηγήσει σε αποτυχία ορθής ταξινόμησης των δύο πλάνων τα οποία υπάγονται σε διαφορετικές κατηγορίες. Η περίπτωση των *false alarms* μπορεί να περιοριστεί με την παρακολούθηση της κίνησης του κυρίαρχου τμήματος δέρματος –βλέπε Παράγραφο 2.2- όπως επιδεικνύεται στην Εικόνα B3.3(δ). Τμήματα δέρματος που αντιστοιχούν σε πραγματικά πρόσωπα εμφανίζουν μικρότερη κίνηση ανάμεσα σε διαδοχικά καρέ, ενώ αντίθετα, σε εξωτερικά πλάνα τα οποία ταξινομούνται λανθασμένα ως πλάνα προσώπων εμφανίζονται ταλαντώσεις στην κίνηση.

Η περίπτωση λανθασμένης ενοποίησης δύο πλάνων, εξαιτίας συμπτωματικής συμφωνίας ως προς τη θέση και το μέγεθος των προσώπων που αντιστοιχούν στα τελευταία καρέ του ενός και στα πρώτα του άλλου, γενικά δεν εμφανίζεται, δεδομένου ότι και τα δύο πλάνα θα έχουν ήδη ταξινομηθεί ως διαφορετικά κατά τη διάρκεια της διαδικασίας αρχικής κατάταξης της ακολουθίας. Σε κάθε πάντως περίπτωση, μια απλή τεχνική διαχωρισμού τους στηρίζεται στην ταύτιση ιστογράμματος όπως αυτή περιγράφεται στη Παράγραφο 2.1 και επιδεικνύεται στην Εικόνα B3.3(ε).

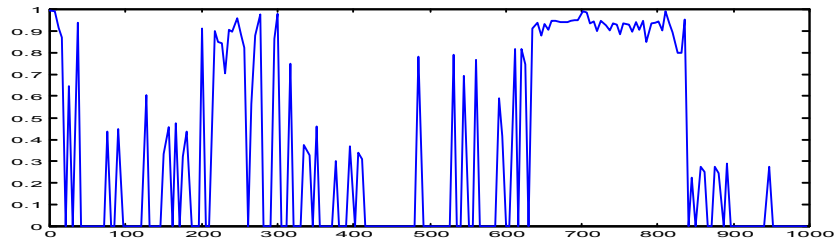


Εικόνα B3.2: Ένα τυπικό απόσπασμα, από την ακολουθία MEGA, συνολικής διάρκειας 100 δευτερόλεπτων, με 15 διακεκριμένες σκηνές μία από τις οποίες αντιστοιχεί στον παρουσιαστή.

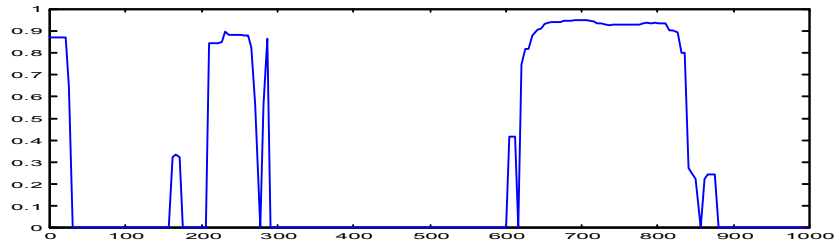
Η συνολική επίδοση του συστήματος αξιολογήθηκε με τις μετρήσεις *precision* και *recall* και παρουσιάζεται στον Πίνακα B3.1. Όλες οι ακολουθίες βίντεο διαιρέθηκαν σε πλάνα και ταξινομήθηκαν χειρονακτικά ώστε να υπάρχει κάποιο μέτρο σύγκρισης με τη διαδικασία αυτόματης ανίχνευσης και ταξινόμησης. Η μέτρηση *precision* [19] ορίζεται ως ο λόγος των σωστά ευθυγραμμισμένων «περιστατικών» προς το συνολικό αριθμό «περιστατικών» –στην ουσία είναι το αντίστροφο του *false alarm*- ενώ η μέτρηση *recall* ορίζεται ως ο λόγος των σωστά ευθυγραμμισμένων «περιστατικών» προς το συνολικό αριθμό πραγματικών «περιστατικών» και αντιστοιχεί στο αντίστροφο του *dismissal*. Περιστατικό ορίζεται κάθε μετάβαση πλάνου ανάμεσα σε διαφορετικές κατηγορίες και σωστή ευθυγράμμιση ορίζεται ως η ανίχνευση της αλλαγής με ακρίβεια ± 2 καρέ ως προς το πραγματικό περιστατικό.



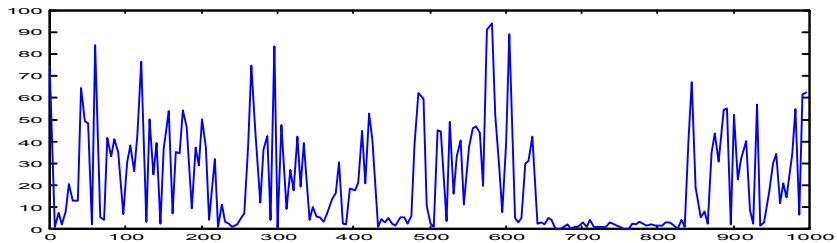
(α)



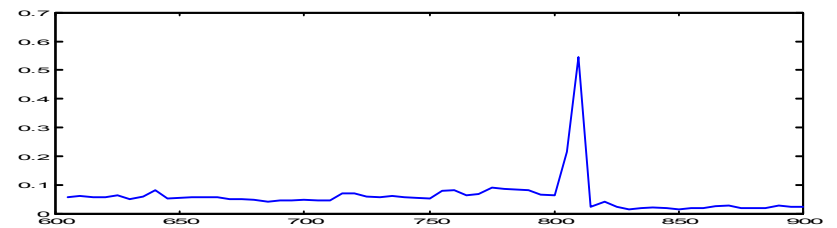
(β)



(γ)



(δ)



(ε)

Εικόνα Β3.3: (α) Η πιθανότητα προσώπου σε σχέση με τον αριθμό του καρέ (0-1000), (β) καμπύλη πιθανότητας για το κυρίαρχο τμήμα δέρματος (μετά την απομάκρυνση των μικρότερων τμημάτων δέρματος), (γ) η ίδια καμπύλη μετά από φιλτράρισμα, (δ) καμπύλη μετατόπισης για το κυρίαρχο τμήμα δέρματος (ε) Μεταβολή ιστογράμματος για τα καρέ 600-900.

Πείραμα	Παρουσιαστής		Ρεπόρτερ / Συνέντευξη		Στατικά πλάνα		Εξωτερικά	
	P	R	P	R	P	R	P	R
A5 (a)	0.92	0.94	0.65	0.82	N/A	N/A	0.75	0.85
A5 (b)	0.95	1.00	0.83	0.94	0.50	1.00	0.73	0.87
ET-1	1.00	1.00	0.71	0.88	0.66	1.00	0.81	0.93
MEGA (a)	0.93	0.93	0.76	0.86	0.75	0.75	0.67	0.86
MEGA (b)	0.96	1.00	0.84	0.91	N/A	N/A	0.74	0.81
ANT1	0.93	0.94	0.77	0.88	0.75	0.66	0.85	0.86
Συνολικό	0.95	0.97	0.76	0.88	0.67	0.85	0.76	0.86

Πίνακας Β3.1: Μετρήσεις **Precision** και **Recall** για την ταξινόμηση σκηνών για τις έξι ακολουθίες βίντεο.

Τα πλάνα με παρουσιαστές ειδήσεων -anchorperson shots- εμφανίζουν τα υψηλότερα ποσοστά ορθής ταξινόμησης. Το γεγονός αυτό είναι αναμενόμενο εξαιτίας της ομοιογένειας την οποία παρουσιάζουν και τις εξαιρετικές συνθήκες κάτω από τις οποίες καταγράφονται τα πρόσωπα. Η ταξινόμηση πλάνων με ρεπόρτερ εμφανίζει χαμηλότερα ποσοστά εξαιτίας των ανομοιογενών συνθηκών φωτισμού οι οποίες επηρεάζουν το μοντέλο χρώματος που χρησιμοποιείται για τον εντοπισμό του προσώπου. Οι σκηνές με ρεπόρτερ ταξινομούνται σε αρκετές περιπτώσεις ως εξωτερικές και αντίστροφα. Ο μικρός αριθμός στατικών σκηνών δεν μπορεί να μας οδηγήσει σε αξιόπιστα συμπεράσματα.

ΑΝΑΚΛΗΣΗ ΕΙΚΟΝΩΝ ΠΡΟΣΩΠΟΥ ΜΕ ΒΑΣΗ ΤΟ ΧΡΩΜΑ ΚΑΙ ΤΙΣ ΚΕΝΤΡΙΚΕΣ ΡΟΠΕΣ

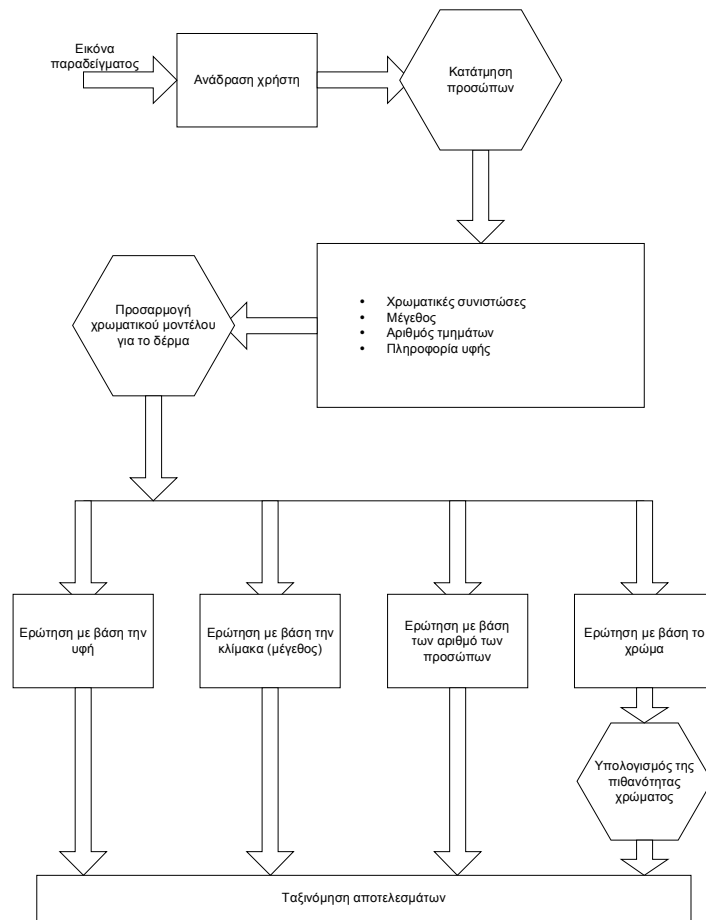
Η ανάκληση με βάση το περιεχόμενο σε βάσεις δεδομένων με εικόνες έχει προσελκύσει μεγάλο ενδιαφέρον τα τελευταία χρόνια. Από την άλλη πλευρά πολλές εικόνες περιέχουν με τον ένα ή τον άλλο τρόπο πρόσωπα. Όπως παρουσιάστηκε στα προηγούμενα κεφάλαια ο εντοπισμός προσώπων με βάση την κατανομή του χρώματός τους, σε συνδυασμό με κάποια απλά χαρακτηριστικά σχήματος, είναι αποτελεσματικός ιδιαίτερα σε εφαρμογές όπου η ακρίβεια δεν είναι το κρίσιμότερο ζήτημα. Στο κεφάλαιο αυτό παρουσιάζεται μια προσέγγιση της ανάκλησης εικόνων προσώπου από βάσεις εικόνων με βάση χρωματικά χαρακτηριστικά. Η μέθοδος σε καμία περίπτωση δεν αποτελεί σχήμα αναγνώρισης προσώπων, μπορεί απλά να απαντήσει σε ερωτήματα όπως «εμφάνισε τις εικόνες της βάσης οι οποίες περιέχουν ένα πρόσωπο το οποίο ομοιάζει είτε ως προς το χρώμα, είτε ως προς τις κεντρικές ροπές, είτε ως προς το μέγεθος με αυτό που εμφανίζεται στην φωτογραφία που σου παρουσιάζω». Μια ερώτηση όπως η προηγούμενη, με παράδειγμα μια εικόνα ενός παρουσιαστή δελτίου ειδήσεων, μπορεί να δώσει ως αποτελέσματα, εικόνες μερικές από τις οποίες θα περιέχουν τον ίδιο ή και άλλους παρουσιαστές ειδήσεων. Το σύστημα έχει μια αλληλοδραστική δομή έτσι ώστε να ικανοποιεί κατά το μέγιστο δυνατό βαθμό τις απαιτήσεις του χρήστη.

1. ΕΙΣΑΓΩΓΗ

Τα τελευταία χρόνια πολλά συστήματα ανάκλησης εικόνων από βάσεις δεδομένων με βάση το περιεχόμενο έχουν εμφανιστεί στη διεθνή βιβλιογραφία [23][28]. Μερικά από αυτά όπως τα VIRAGE, QBIC, Photobook και VisualSEEK, έχουν υλοποιηθεί και αποτελούν πλέον αντικείμενο εμπορικής εμετάλλευσης. Στις περισσότερες περιπτώσεις η ανάκληση με βάση το περιεχόμενο πραγματοποιείται σε τρία στάδια: (α) μοντελοποίηση των αντικείμενων της εικόνας ή της ακολουθίας βίντεο, (β) κατάτμηση των αντικείμενων αυτών και (γ) εξαγωγή χαρακτηριστικών χρώματος, κίνησης, σχήματος, υψής και από τα αντικείμενα καθώς και από τη χωροχρονική συσχέτιση τους [24][13].

Ένα από τα σημαντικότερα προβλήματα που αντιμετωπίζουν τα περισσότερα από τα υπάρχοντα σχήματα είναι η αδυναμία εξαγωγής πληροφορίας με νοηματικό περιεχόμενο. Αυτό οδηγεί σε αποτελέσματα ανακλήσεων που είναι λογικοφανή μόνο ως προς τα χαρακτηριστικά χαμηλού επιπέδου –όπως για παράδειγμα το χρώμα. Στην πραγματικότητα το γεγονός αυτό είναι ικανοποιητικό μόνο αν ο ανθρώπινος παράγοντας προσαρμοστεί ως προς τον τρόπο λειτουργίας του υπολογιστή. Με δεδομένο ότι η βασική αρχή της επικοινωνίας ανθρώπου μηχανής αποσκοπεί στο να αποκτήσουν οι υπολογιστές τρόπο «σκέψης» παρόμοιο με αυτό των ανθρώπων και όχι το αντίστροφο, οι ανακλήσεις εικόνων με βάση το περιεχόμενο πρέπει να έχουν κάποια νοηματική αξία. Για να επιτευχθεί αυτό απαιτείται η ύπαρξη πρότερης γνώσης και ένα συγκεκριμένο πλαίσιο για την εφαρμογή. Παραδείγματα τέτοιων εφαρμογών αποτελούν τα συστήματα ανάκλησης εικόνων με πρόσωπα τα οποία μπορούν να αποτελέσουν προστάδια περισσότερο εξελιγμένων σχημάτων όπως αυτά της αναγνώρισης προσώπων.

Στο κεφάλαιο αυτό παρουσιάζεται ένα σύστημα ανάκλησης εικόνων προσώπου το οποίο στηρίζεται στη διαδικασία εντοπισμού προσώπων που αναπτύχθηκε στο Κεφάλαιο Β2. Το σχήμα ανάκλησης παρουσιάζεται στο τμήμα αυτό της διατριβής –και όχι για παράδειγμα σε αυτό της αναγνώρισης προσώπων- για δύο λόγους: (α) χρησιμοποιεί την μοντελοποίηση χρώματος που αναπτύχθηκε στο Κεφάλαιο Β2 όχι μόνο για τον εντοπισμό των προσώπων σε μια εικόνα αλλά και ως συστατικό στοιχείο της διαδικασίας ανάκλησης και, (β) καταδεικνύει ότι στο πλαίσιο των πολυμεσιών εφαρμογών η ανίχνευση προσώπων είναι ένα σημαντικό εργαλείο και επομένως δεν πρέπει να θεωρείται μόνο ως ένα αναπόσπαστο τμήμα συστημάτων αναγνώρισης προσώπων ή σχημάτων κωδικοποίησης.



Σχήμα Β4.1: Μπλοκ διάγραμμα του σχήματος ανάκλησης

2. ΕΝΑ ΣΕΝΑΡΙΟ ΑΝΑΚΛΗΣΗΣ

Το σύστημα ανάκλησης εικόνων το οποίο παρουσιάζεται στο κεφάλαιο αυτό είναι ένα σχήμα «ερώτησης με βάση παράδειγμα». Οι αποθηκευμένες στη βάση εικόνες –προσώπων ή μη- έχουν διερευνηθεί ως προς την ύπαρξη τμημάτων προσώπου με χρήση της τεχνικής που παρουσιάστηκε στο Κεφάλαιο Β2. Για κάθε εικόνα της βάσης υπάρχει αποθηκευμένο ένα διάνυσμα χαρακτηριστικών στο οποίο αναφέρεται ο αριθμός

των πιθανών τμημάτων προσώπου και κάποια συνοπτική περιγραφή για κάθε τμήμα. Το διάνυσμα αυτό χρησιμοποιείται για δεικτοδότηση των εικόνων της βάσης και ταχεία ανάκληση.

Το σενάριο ανάκλησης συνοψίζεται στο Σχήμα Β4.1. Η εικόνα εισόδου αναλύεται ως προς την ύπαρξη προσώπου /προσώπων και στη συνέχεια κάποια χαρακτηριστικά όπως το μέσο χρώμα, η κλίμακα και οι κανονικοποιημένες κεντρικές ροπές εξάγονται για κάθε εντοπισμένο τμήμα προσώπου. Οι μέσες χρωματικές συνιστώσες χρησιμοποιούνται για την τροποποίηση του αρχικού μοντέλου για το χρώμα του δέρματος. Μετά την εξαγωγή των παραπάνω χαρακτηριστικών διάφορα σενάρια ανάκλησης είναι εφικτά: (α) ανάκληση με βάση το χρώμα, (β) ανάκληση με βάση τον αριθμό των εντοπισμένων προσώπων, (γ) ανάκληση με βάση την υφή του προσώπου του παραδείγματος, (δ) ανάκληση με βάση το μέγεθος του προσώπου και (ε) συνδυασμοί των προηγούμενων περιπτώσεων.

3. ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΠΕΡΙΓΡΑΦΗΣ ΤΜΗΜΑΤΩΝ ΠΡΟΣΩΠΟΥ

Για κάθε εικόνα που είναι αποθηκευμένη στη βάση υπάρχει μια αντίστοιχη εγγραφή στην οποία αποθηκεύονται τα επόμενα στοιχεία:

- *Αριθμός τμημάτων προσώπου.* Τα τμήματα προσώπου έχουν ανιχνευθεί με τη διαδικασία που περιγράφηκε στο Κεφάλαιο Β2. Ο αριθμός τους ποικίλει από 0 έως 4 (περισσότερα από 4 πρόσωπα με ικανοποιητική χωρική ανάλυση δεν μπορούν στη πράξη να χωρέσουν σε μια τυπική εικόνα)
- *Μέση τιμή χρωματικών συνιστωσών για κάθε ένα από τα τμήματα προσώπου.* Αν I_Y, I_{C_r}, I_{C_b} είναι οι συνιστώσες YC_b της εικόνας I , και S_i είναι ένα χωρικό τμήμα της –σύνολο από pixels- τότε $m_{C_r}(S_i) = \frac{1}{\|S_i\|} \sum_{p \in S_i} I_{C_r}(p)$ είναι η μέση τιμή της χρωματικής συνιστώσας C_r και $m_{C_b}(S_i) = \frac{1}{\|S_i\|} \sum_{p \in S_i} I_{C_b}(p)$ είναι η μέση τιμή της χρωματικής συνιστώσας C_b που αντιστοιχεί στο τμήμα S_i .
- *Το διάνυσμα των κανονικοποιημένων κεντρικών ροπών για κάθε ένα από τα τμήματα προσώπου υπολογισμένο με βάση τη συνιστώσα φωτεινότητας I_Y*
- *Το κέντρο μάζας $g(S_i)$ για κάθε ένα από τα τμήματα προσώπου*
- *Το ποσοστό $A(S_i)$ της εικόνας το οποίο καταλαμβάνεται από κάθε ένα από τα τμήματα προσώπου*

Τα παραπάνω χαρακτηριστικά αποτελούν στην ουσία ένα διάνυσμα περιγραφής κάποιας εικόνας ως προς το περιεχόμενο «πρόσωπα» και μπορεί να χρησιμοποιηθεί για σκοπούς δεικτοδότησης (indexing). Όπως θα δούμε στη συνέχεια με βάση το διάνυσμα αυτό μπορούν να αναπτυχθούν διάφορα σενάρια ανάκλησης εικόνων που περιέχουν πρόσωπα.

4. ΑΝΑΚΛΗΣΗ ΜΕ ΒΑΣΗ ΤΟ ΠΕΡΙΕΧΟΜΕΝΟ

Στην ανάκληση με βάση παραδείγματα μια εικόνα παρουσιάζεται στο σύστημα και ζητείται από αυτό να επιστρέψει ένα σύνολο από εικόνες οι οποίες ομοιάζουν με αυτήν με βάση κάποιο κριτήριο. Στην περίπτωση ανάκλησης εικόνων προσώπου το παράδειγμα είναι μια εικόνα που περιέχει κάποιο πρόσωπο –ή πρόσωπα– και το ζητούμενο από το σύστημα είναι: (α) να εντοπίσει το πρόσωπο στην εικόνα του παραδείγματος και (β) να ανακαλέσει από τη βάση πρόσωπα που ομοιάζουν με αυτό ως προς κάποια απλά κριτήρια όπως είναι το χρώμα, το μέγεθος και οι κανονικοποιημένες κεντρικές ροπές. Σε περίπτωση που στην εικόνα του παραδείγματος υπάρχουν περισσότερα του ενός πρόσωπα ο χρήστης μπορεί να υποδείξει στο σύστημα με βάση πιο από τα πρόσωπα επιθυμεί να γίνει η ανάκληση.

4.1 Ανάκληση με βάση τις μέσες χρωματικές συνιστώσες

Έστω I η εικόνα του παραδείγματος η οποία περιέχει ένα πρόσωπο το οποίο αντιστοιχεί στο χωρικό τμήμα F της εικόνας. Η κατάτμηση του τμήματος πραγματοποιείται με την μέθοδο που παρουσιάστηκε στο Κεφάλαιο Β2 και με χρήση κάποιου αρχικού μοντέλου χρώματος το οποίο περιγράφεται από την κατανομή:

$$P(\mathbf{x} | \boldsymbol{\mu}_0, \mathbf{C}) = \frac{\exp\left\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_0)^T \mathbf{C}^{-1}(\mathbf{x} - \boldsymbol{\mu}_0)\right\}}{(2\pi)^{\frac{k}{2}} \cdot |\mathbf{C}|^{\frac{1}{2}}} \quad (\text{B4.1})$$

Αν I_Y , I_{C_r} , I_{C_b} είναι οι συνιστώσες $YC_r C_b$ της εικόνας I τότε υπολογίζουμε τις τιμές των μέσων χρωματικών συνιστωσών της περιοχής F από τις σχέσεις:

$$m_{C_r}(F) = \frac{1}{\|F\|} \sum_{p \in F} I_{C_r}(p) \quad (\text{B4.2})$$

$$m_{C_b}(F) = \frac{1}{\|F\|} \sum_{p \in F} I_{C_b}(p) \quad (\text{B4.3})$$

και κατασκευάζουμε το διάνυσμα $\boldsymbol{\mu} = [m_{C_r}(F) \ m_{C_b}(F)]^T$. Αν τροποποιήσουμε το μοντέλο της σχέσης (B4.1) με βάση τη σχέση:

$$P(\mathbf{x} | \hat{\boldsymbol{\mu}}_0, \mathbf{C}) = \frac{\exp\left\{-\frac{1}{2}(\mathbf{x} - \hat{\boldsymbol{\mu}}_0)^T \mathbf{C}^{-1}(\mathbf{x} - \hat{\boldsymbol{\mu}}_0)\right\}}{(2\pi)^{\frac{k}{2}} \cdot |\mathbf{C}|^{\frac{1}{2}}} \quad (\text{B4.4})$$

$$\hat{\boldsymbol{\mu}}_0 = m \cdot \boldsymbol{\mu}_0 + (1-m) \cdot \boldsymbol{\mu} \quad (\text{B4.5})$$

Χρησιμοποιώντας μια σχετικά μικρή τιμή για τη σταθερά μνήμης m ($m=0.4$ είναι μια τυπική τιμή) τότε στην ουσία έχουμε προσαρμόσει το μοντέλο της κατανομής χρώματος στις χρωματικές συνιστώσες του προσώπου που απεικονίζεται στην ελάχιστη εικόνα του παραδείγματος.

Έστω $S_i^{(k)}$ το i -στο τμήμα προσώπου της k -στης εικόνας της βάσης, $x(S_i^{(k)}) = [m_{C_r}(S_i^{(k)}) \ m_{C_b}(S_i^{(k)})]^T$ το αντίστοιχο διάνυσμα των μέσων χρωματικών συνιστωσών και $c(S_i^{(k)})$ μια ποσότητα που περιγράφει το βαθμός προσαρμογής του $S_i^{(k)}$ στο τροποποιημένο μοντέλο της σχέσης (B4.4) και η οποία δίνεται από τη σχέση:

$$c(S_i^{(k)}) = \exp\left\{-\frac{1}{2}(x(S_i^{(k)}) - \hat{\mu}_0)^T \mathbf{C}^{-1}(x(S_i^{(k)}) - \hat{\mu}_0)\right\} \quad (\text{B4.6})$$

τότε η βέλτιστη ταύτιση με κάποια εικόνα της βάσης πραγματοποιείται με χρήση του κριτηρίου:

$$k_0 = \arg \max_k \{\max_i \{c(S_i^{(k)})\}\} \quad (\text{B4.7})$$

Το κριτήριο της σχέσης (B4.7) μπορεί να χρησιμοποιηθεί διαδοχικά για την εμφάνιση μιας σειράς από εικόνες που ομοιάζουν με το πρόσωπο της εικόνας εισόδου σύμφωνα με το κριτήριο χρώματος.

Πρέπει εδώ να τονιστεί ότι ανάκληση μπορούσε να πραγματοποιηθεί με βάση την απευθείας σύγκριση του διανύσματος των χρωματικών συνιστωσών. Αυτό όμως θα μπορούσε σε κάποιες οριακές περιπτώσεις να οδηγήσει στην ανάκληση εικόνων οι οποίες δεν περιέχουν πρόσωπα. Η έννοια της χρήσης του προσαρμοσμένου μοντέλου είναι να δώσει τμήματα τα οποία ομοιάζουν χρωματικά ως προς αυτό του παραδείγματος αλλά πάνω από όλα εξακολουθούν να έχουν χρώμα το οποίο ταιριάζει με το χρώμα του δέρματος.

4.2 Ανάκληση με βάση τις κανονικοποιημένες κεντρικές ροπές

Έστω F το χωρικό τμήμα της εικόνας του παραδείγματος με διάνυσμα κανονικοποιημένων κεντρικών ροπών $n(F) = [n_{00}(F) \ n_{20}(F) \ n_{02}(F) \ n_{30}(F) \ n_{03}(F) \ n_{12}(F) \ n_{21}(F)]^T$. Αν $S_i^{(k)}$ είναι το i -στο τμήμα προσώπου της k -στης εικόνας της βάσης με διάνυσμα κανονικοποιημένων κεντρικών ροπών $n(S_i^{(k)})$ τότε η βέλτιστη ταύτιση σε σχέση με το τμήμα εισόδου υπολογίζεται με χρήση του κριτηρίου:

$$k_0 = \arg \min_k \{\|n(S_{i_0}^{(k)}) - n(F)\|\} \quad (\text{B4.8})$$

$$\text{όπου } i_0 = \arg \min_i \{\|n(S_i^{(k)}) - n(F)\|\} \quad (\text{B4.9})$$

υπολογίζεται ξεχωριστά για κάθε εικόνα της βάσης.

Το κριτήριο της σχέσης (B4.8) μπορεί να χρησιμοποιηθεί διαδοχικά για την εμφάνιση μιας σειράς από εικόνες που ομοιάζουν με το πρόσωπο της εικόνας εισόδου σύμφωνα με τις κανονικοποιημένες κεντρικές ροπές. Πρέπει επίσης να διευκρινιστεί ότι οι ροπές υπολογίζονται μόνο με χρήση του καναλιού φωτεινότητας Y .

Η περιγραφή κάποιου τμήματος με τη χρήση των κανονικοποιημένων κεντρικών ροπών δίνει μια ένδειξη για την υφή και το περιεχόμενο του τμήματος αυτού. Η ανάκληση με βάση τις κανονικοποιημένες κεντρικές ροπές πλησιάζει κάπως την έννοια της αναγνώρισης προσώπου αλλά διαφέρει σημαντικά από

αυτήν δεδομένου ότι δεν υπάρχει καμιά υπόθεση για την θέση των προσώπων ως προς τους άξονες \mathbf{x} , \mathbf{y} και \mathbf{z} –βλέπε Σχήμα A4.1. Σε κάθε πάντως περίπτωση η ανάκληση δίνει πάντοτε κάποια λογικοφανή αποτελέσματα.

4.3 Ανάκληση με βάση την κλίμακα

Στην ανάκληση με βάση την κλίμακα το ζητούμενο από το σύστημα είναι η ανάκληση εικόνων στις οποίες κάποιο από τα τμήματα προσώπου καλύπτει μέρος της εικόνας ανάλογο με αυτό που καλύπτει το πρόσωπο της εικόνας του παραδείγματος.

Έστω F το χωρικό τμήμα της εικόνας του παραδείγματος και $A(F)$ το ποσοστό της εικόνας που καταλαμβάνεται από το τμήμα αυτό. Αν $S_i^{(k)}$ είναι το i -στο τμήμα προσώπου της k -στης εικόνας της βάσης και $A(S_i^{(k)})$ το ποσοστό της εικόνας που αυτό καταλαμβάνει τότε η βέλτιστη ταύτιση σε σχέση με το τμήμα εισόδου υπολογίζεται με χρήση του κριτηρίου:

$$k_0 = \arg \min_k \{ |A(S_{i_0}^{(k)}) - A(F)| \} \quad (\text{B4.10})$$

$$\text{όπου } i_0 = \arg \min_i \{ |A(S_i^{(k)}) - A(F)| \} \quad (\text{B4.11})$$

υπολογίζεται ξεχωριστά για κάθε εικόνα της βάσης

Το κριτήριο της σχέσης (B4.10) μπορεί να χρησιμοποιηθεί διαδοχικά για την εμφάνιση μιας σειράς από εικόνες που ομοιάζουν με το πρόσωπο της εικόνας εισόδου σύμφωνα με το κριτήριο μεγέθους.

4.4 Ανάκληση με βάση τον αριθμό των προσώπων

Στη περίπτωση αυτή η κατάταξη των εικόνων που ικανοποιούν τον ζητούμενο αριθμό τμημάτων προσώπου πραγματοποιείται με βάση την ομοιότητα ως προς το χρώμα. Επομένως για το υποσύνολο των εικόνων της βάσης που περιέχουν ένα συγκεκριμένο αριθμό τμημάτων προσώπου πραγματοποιείται η διαδικασία της Παραγράφου 4.1 με χρήση μιας σταθεράς μνήμης m με σχετικά μεγάλη τιμή.

5. ΠΕΙΡΑΜΑΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ

Στην παράγραφο αυτή παρουσιάζονται κάποια πειραματικά αποτελέσματα ως προς την ανάκληση εικόνων προσώπου με βάση τα κριτήρια της Παραγράφου 4. Η βάση των εικόνων κατασκευάστηκε από εικόνες προερχόμενες από τηλεοπτικά προγράμματα, ατομικές βιντεοκάμερες και ψηφιοποιημένες έγχρωμες φωτογραφίες. Χρησιμοποιήθηκαν συνολικά 200 εικόνες, 156 από τις οποίες περιέχουν τουλάχιστον ένα πρόσωπο. Δεδομένου ότι δεν πρόκειται για κάποια εφαρμογή αναγνώρισης προσώπου δεν δόθηκε έμφαση στην παρουσία περισσότερων από μιας αναπαραστάσεων ενός συγκεκριμένου προσώπου.



$m=0.4$



Εικόνα που παρουσιάζεται στο σύστημα

Κατάτμηση προσώπου



0.9992



0.9872



0.9735



0.9591

Εικόνα B4.1: Ανάλυση με βάση την ομοιότητα χρώματος



Εικόνα που παρουσιάζεται στο σύστημα



Κατάτμηση προσώπου



0.0873



0.0883



0.0969



0.0985

Εικόνα B4.2: Ανάλυση με βάση την ομοιότητα στην κλίμακα

Στην Εικόνα B4.1 παρουσιάζονται τα αποτελέσματα της ανάλυσης με βάση την ομοιότητα στις χρωματικές συνιστώσες. Κάτω από κάθε εικόνα εμφανίζεται ο βαθμός προσαρμογής της ως προς το τροποποιημένο μοντέλο χρώματος για το οποίο χρησιμοποιήθηκε σταθερά μνήμης $m=0.4$. Αξιοσημείωτο

είναι ότι στην Εικόνα B4.1(γ) παρότι το πρόσωπο εμφανίζεται ισχυρά φωτισμένο αυτό δεν το εμποδίζει να έχει τον υψηλότερο βαθμό προσαρμογής ως προς το μοντέλο χρώματος. Στην Εικόνα B4.2 παρουσιάζονται τα αποτελέσματα της ανάκλησης με βάση την ομοιότητα ως προς το μέγεθος. Είναι προφανές ότι τα αποτελέσματα είναι τουλάχιστον λογικοφανή. Κάποιες μικρές διακυμάνσεις προέρχονται από την παρουσία του λαιμού σε ορισμένες εικόνες. Στην Εικόνα B4.3 επιδεικνύονται τα αποτελέσματα της ανάκλησης εικόνων στις οποίες υπάρχουν δύο πρόσωπα –σε αποδεικτική χωρική ανάλυση. Η κατάταξη τους πραγματοποιείται με βάση την ομοιότητα χρώματος που παρουσιάζουν σε σχέση με το πρόσωπο του παραδείγματος.

Τέλος στην Εικόνα B4.4 έχουμε τα αποτελέσματα της ανάκλησης με βάση τις κανονικοποιημένες κεντρικές ροπές. Κάτω από κάθε εικόνα εμφανίζεται η απόσταση της αναπαράστασης του προσώπου της εν λόγω από το πρόσωπο της εικόνας του παραδείγματος.

6. ΣΥΜΠΕΡΑΣΜΑΤΑ

Η κατάτμηση προσώπων με βάση χρωματικά χαρακτηριστικά και χαρακτηριστικά σχήματος μπορεί να χρησιμοποιηθεί στο πλαίσιο πολυμεσιών βάσεων δεδομένων για ανάκληση εικόνων που περιέχουν πρόσωπα, σύμφωνα με κάποια απλά κριτήρια. Η εφαρμογή έχει την έννοια της δεικτοδότησης και σε καμία περίπτωση αυτήν της αναγνώρισης προσώπων δεδομένου ότι δεν τίθεται κανένας περιορισμός ως προς τα πρόσωπα που απεικονίζονται στις εικόνες. Επιπλέον τα χαρακτηριστικά που χρησιμοποιούνται για την περιγραφή δεν μπορούν να θεωρηθούν ως ικανά να διαχωρίσουν ανθρώπινα πρόσωπα μεταξύ τους αλλά πρόσωπα από άλλα αντικείμενα. Οι ανακλήσεις σε όλες σχεδόν τις περιπτώσεις είναι λογικοφανείς και ο χρόνος εκτέλεσης ουσιαστικά προσδιορίζεται από την διαδικασία κατάτμησης του προσώπου στην εικόνα εισόδου.



Εικόνα B4.3: Ανάκληση με βάση τον αριθμό των τμημάτων δέρματος



Εικόνα που παρουσιάζεται στο σύστημα



Κατάτμηση προσώπων



0.9998



0.9444



0.9165



0.8416

Εικόνα B4.4: Ανάκληση με βάση τις κεντρικές ροπές φωτεινότητας

BIBΛΙΟΓΡΑΦΙΑ

- [1]. H. Wang and S.-F. Chang, "A Highly Efficient System for Automatic Face Region Detection in MPEG Video," *IEEE Trans. on Circuits and Systems for Video Technology*, special issue on Multimedia Systems and Technologies, August 1997.
- [2]. A. Samal and P.A. Iyengar, "Automatic Recognition and Analysis of Human Faces and Facial Expressions: A Survey," *Pattern Recognition*, Vol. 25, No. 1, pp. 65-77, 1992.
- [3]. G. Yang and T.S. Huang, "Human Face Detection in Complex Background," *Pattern Recognition*, Vol. 27, No. 1, pp. 55-63, 1994.
- [4]. K.C. Yow and C. Cipolla, "Feature-based Human Face Detection in Complex Background," *Image and Vision Computing Recognition*, Vol. 15, pp. 713-735, 1997.
- [5]. C. Garcia and G. Tziritas, "Face Detection Using Quantized Skin Color Regions Merging and Wavelet Packet Analysis," *IEEE Trans. on Multimedia*, vol. 1, no. 3, pp.264-277, September 1999.
- [6]. Y. Avrithis, N. Tsapatsoulis and S. Kollias, "Color-Based Retrieval of Facial Images," *Proc. of EUSIPCO 2000*, Tampere, Finland, September 2000.
- [7]. Y. Avrithis, N. Tsapatsoulis and S. Kollias, "Broadcast News Parsing Using Visual Cues: A Robust Face Detection Approach," *Proc. of IEEE International Conference on Multimedia and Expo*, July 2000, New York City, USA
- [8]. N. Tsapatsoulis, N. Doulamis, A. Doulamis, and S. Kollias, "Face Extraction from Non-uniform Background and Recognition in Compressed Domain," *Proc. of ICASSP'98*, Seattle WA, May 1998.
- [9]. T. Rzeszewski, "A Novel Automatic Hue Control System," *IEEE Trans. Consumer Electron.*, vol. CE-21, pp. 155-162, May 1975.
- [10]. L. A. Harwood, "A Chrominance Demodulator IC with Dynamic Flesh Correction," *IEEE Trans. Consumer Electron.*, vol. CE-22, pp. 111-117, Feb 1976
- [11]. P. Maragos, "Morphological Signal and Image Processing," *Digital Signal Processing Handbook*, V. Madisetti and D. Williams, eds., IEEE Press.
- [12]. N. Tsapatsoulis, Y. Avrithis and S. Kollias "Face Detection for Multimedia Applications," *Proc. of ICIP 2000*, September 2000, Vancouver, BC, Canada
- [13]. Y. Avrithis, A. Doulamis, N. Doulamis and S. Kollias, "A Stochastic Network for Optimal Key Frame Extraction from MPEG Video Databases," *Computer Vision and Image Understanding* **75** (1/2), pp. 3-24, July 1999.
- [14]. P. Bouthemy, M. Gelgon and F. Ganancia, "A Unified Approach to Shot Change Detection and Camera Motion Characterization," *IEEE Trans. CSVT* **9** (7), pp.1030-1044, Oct. 1999.
- [15]. M. Brown, J. Foote, G. Jones, K. Sparck-Jones and S. Young, " Automatic Content-Based Retrieval of Broadcast News," *Proc. of ACM Multimedia Conference*, San Francisco, CA, Nov. 1995.
- [16]. S. Eickeler, A. Kosmala and G. Rigoll, "A New Approach to Content-Based Video Indexing Using Hidden Markov Models," *Proc. of WLAMIS*, Belgium, June 1997.
- [17]. ISO/IEC JTC1/SC29/WG11, "MPEG-7: Context and Objectives (v.5)," Doc. N1920, 1997.
- [18]. B. Merialdo, "Automatic Indexing of TV News," *Proc. of WLAMIS*, Belgium, June 1997.
- [19]. A. Merlino, D. Morey and M. Maybury, "Broadcast News Navigation Using Story Segments," *Proc. of ACM Multimedia Conference*, Seattle, WA, Nov. 1997.
- [20]. Y. Nakamura and T. Kanade, "Semantic Analysis for Video Contents Extraction – Spotting by Association in News Video," *Proc. of ACM Multimedia Conference*, Seattle, WA, Nov. 1997.
- [21]. S. Tsekeridou and I. Pitas, "Audio-Visual Content Analysis for Content-Based Video Indexing," *Proc. of Int. Conf. on Multimedia Computing and Systems*, Florence, Italy, June 1999.

- [22]. H.J. Zhang, S.Y. Tan, S. Smoliar and G. Yihong, "Automatic Parsing and Indexing of News Video," *Multimedia Systems* **2**, pp. 256-266, 1995.
- [23]. D. Androutsos, K. N. Plataniotis and A. N. Venetsanopoulos, "Extraction of Detailed Image Regions for Content-Based Image Retrieval," *Proceedings of IEEE ICASSP*, Seattle WA, USA, May 1998.
- [24]. A. Alatan, L. Onural, M. Wollborn, R. Mech, E. Tuncel and T. Sikora, "Image Sequence Analysis for Emerging Interactive Multimedia Services - The European Cost 211 Kαρέwork," *IEEE Trans. Circuits and Systems for Video Technology*, Vol. 8, No. 7, pp. 802- 813, Nov. 1998.
- [25]. O. J. Morris, M. J. Lee and A. G. Constantinides, "Graph Theory for Image Analysis: an Approach based on the Shortest Spanning Tree," *IEE Proceedings*, Vol. 133, pp.146-152, April 1986.
- [26]. P. J. Mulroy, "Video Content Extraction: Review of Current Automatic Segmentation Algorithms," *Proceedings of Workshop on Image Analysis and Multimedia Interactive Systems (WLAMIS)*, Louvain-la-Neuve, Belgium, June 1997.
- [27]. Y. Rui, T. S. Huang and S.-F. Chang, "Digital Image/Video Library and MPEG-7: Standardization and Research Issues," *Proc. of IEEE ICASSP*, pp. 3785-3788, Seattle, WA, May 1998.
- [28]. Special Issue on Segmentation, Description and Retrieval of Video Content, *IEEE Trans. Circuits and Systems for Video Technology*, Vol. 8, No. 5, 1998.
- [29]. Y. Avrithis, Y. Xirouhakis and S. Kollias "Affine-Invariant Curve Normalization for Shape-Based Retrieval," *Proc. of International Conference on Pattern Recognition (ICPR)*, Barcelona, Spain, September 2000.
- [30]. A. D. Bimbo and P. Pala, "Visual Image Retrieval by Elastic Matching of User Sketches," *IEEE Trans. PAMI* **19** (2), pp. 121-132, 1997.
- [31]. K. Karpouzis, G. Votsis, N. Tsapatsoulis and S. Kollias, "Compact 3D Model Generation based on 2D Views of Human Faces: Application to Face Recognition," in *Machine Graphics and Vision*, vol. 7, no.1-2, 1998.
- [32]. N. Tsapatsoulis, K. Karpouzis, G. Votsis and S. Kollias, "Analysis by synthesis of facial images based on frontal and profile views," in *Proceedings of the IWSNHC3DI'97*, Rhodes, Greece, September 1997.
- [33]. G. Votsis, N. Tsapatsoulis, K. Karpouzis and S. Kollias, "A Simplified Representation of 3D Human Faces adapted from 2D Images," in *Proceedings of the NMBLA'98*, Glasgow, UK, July 1998.

ΆΛΛΕΣ ΕΦΑΡΜΟΓΕΣ: ΕΝΤΟΠΙΣΜΟΣ ΚΥΤΤΑΡΩΝ ΣΕ ΙΣΤΟΠΑΘΟΛΟΓΙΚΕΣ ΕΞΕΤΑΣΕΙΣ

Στο κεφάλαιο παρουσιάζονται κάποιες άλλες εφαρμογές των τεχνικών εντοπισμού αντικείμενων με βάση το μέγεθος και με χρήση μορφολογικών τελεστών. Η ιδιαιτερότητα έγκειται στο γεγονός ότι οι εφαρμογές αναφέρονται σε εικόνες απόχρωσης του γκρι χωρίς να έχει προηγηθεί κάποια άλλη μέθοδος κατάτμησης. Η τεχνική που περιγράφεται στη συνέχεια αποσκοπεί στον επακριβή εντοπισμό των πυρήνων από κύτταρα σε ιστοπαθολογικές εξετάσεις. Αποτελεί τμήμα ενός γενικότερου συστήματος, που χρησιμοποιείται για την πρόβλεψη της εξέλιξης του καρκίνου του μαστού, που εμπλέκει ένα νευρωνικό ταξινομητή και ένα σχήμα πρόβλεψης προερχόμενο από ιατρικές μελέτες.

1. ΕΙΣΑΓΩΓΗ

Η ανίχνευση πυρήνων από κύτταρα, σε εικόνες βιοψίας, αποτελεί ένα εξαιρετικά σημαντικό τμήμα των σχημάτων πρόγνωσης όσον αφορά την εξέλιξη της πορείας του καρκίνου του μαστού γιατί επηρεάζει άμεσα τα αποτελέσματα της. Κατά τη διαδικασία πρόγνωσης τα κύτταρα εμποτίζονται με μονόκλωνα αντισώματα και οι στεροειδείς υποδοχείς εμφανίζονται με χρήση ανοσοκυτταρολογικών μεθόδων οι οποίες εφαρμόζονται σε slides ιστών των δειγμάτων της βιοψίας [1]. Τα slides αξιολογούνται από τους ειδικούς οι οποίοι αναθέτουν ένα δείκτη διάγνωσης ο οποίος στηρίζεται στο βαθμό εμπότισης και το ποσοστό των πυρήνων που αντιδρούν θετικά στα μονόκλωνα αντισώματα. Ο δείκτης διάγνωσης αποτελεί μια πρόγνωση για την εξέλιξη της ασθένειας και χρησιμοποιείται από τους ογκολόγους για την εκτίμηση της βέλτιστης μεθοδολογίας αντιμετώπισης με σκοπό την παράταση της ζωής των ασθενών. Το βασικό μειονέκτημα της ανωτέρω μεθοδολογίας είναι ότι οι ειδικοί καλούνται να ταξινομήσουν τους πυρήνες χειρονακτικά σε διάφορες κατηγορίες. Η διαδικασία αυτή είναι και κουραστική και χρονοβόρα με αποτέλεσμα οι ειδικοί να πραγματοποιούν μια χονδρική εκτίμηση η οποία είναι επιρρεπής σε σφάλματα. Επιπλέον υπάρχουν σημαντικές διακυμάνσεις όσον αφορά την εκτίμηση του ίδιου ειδικού σε μεταγενέστερο χρόνο με βάση την ίδια εικόνα αλλά και μεταξύ δύο ειδικών. Η ανάπτυξη ενός αυτοματοποιημένου συστήματος πρόγνωσης εγγυάται τουλάχιστον μια σταθερή πρόγνωση, με bias ή χωρίς, και μπορεί πάντοτε να χρησιμοποιηθεί παράλληλα με τους ειδικούς.

Η διαδικασία αυτόματου εντοπισμού των πυρήνων είναι εξαιρετικά σημαντική γιατί αφενός μας δίνει τον συνολικών αριθμό των πυρήνων και αφετέρου προσδιορίζει ουσιαστικά το αποτέλεσμα της πρόγνωσης – σωστός εντοπισμός οδηγεί σχεδόν πάντα και σε σωστή ταξινόμηση στο επόμενο στάδιο. Επιπλέον είναι το τμήμα το οποίο κουράζει περισσότερο τους ειδικούς και η αυτόματη υλοποίηση του οδηγεί σε περισσότερο αξιόπιστη πρόγνωση. Οι ειδικοί μπορούν σε κάθε περίπτωση να διατηρήσουν την επίβλεψη και να απορρίψουν λανθασμένα εντοπισμένους πυρήνες ή και να προσθέσουν κάποιους άλλους τους οποίους το σύστημα απέτυχε να εντοπίσει. Η διαδικασία εντοπισμού πρέπει να είναι αναισθητή από το βαθμό

εμπότισης των πυρήνων ο οποίος προκαλεί πυρήνες με διαφορετική ένταση φωτεινότητας. Επιπλέον το πάχος των slides δημιουργεί την αίσθηση επικαλυπτόμενων πυρήνων οι οποίοι πρέπει να μπορούν να διαχωρίζονται. Ο αλγόριθμος εντοπισμού δεν πρέπει να επηρεάζεται από την κλίμακα των πυρήνων και τις γενικότερες συνθήκες φωτισμού.

2. ΕΝΤΟΠΙΣΜΟΣ ΤΩΝ ΠΥΡΗΝΩΝ

Έστω $I \in \mathbb{R}^{n \times m}$ μια ψηφιακή εικόνα εισόδου ορισμένη πάνω σε ένα ορθογωνικό πλέγμα L , για την οποία ισχύει $l \leq I(i, j) \leq h, \forall i, j$. Έστω επίσης η ακολουθία $T(k), k = 1, \dots, N$ για την οποία ισχύει $l \leq T(k) \leq h$ με $T(1) = l, T(N) = h$. Ορίζουμε τα σύνολα $S^{(k)}, k = 1, \dots, N$, σύμφωνα με τη σχέση:

$$S^{(k)}(i, j) = \begin{cases} 1 & \text{όταν } I(i, j) \geq T(k) \\ 0 & \text{όταν } I(i, j) < T(k) \end{cases}$$

Προφανώς ισχύει $S^{(N)} \subseteq S^{(N-1)} \subseteq \dots \subseteq S^{(1)} = L$.

Ορίζουμε τον τελεστή *opening by reconstruction* στο σύνολο X με σύνολο markers το m ως: $Op^{(r)}(X, m)$. Με βάση τον ορισμό αυτό δημιουργούμε τα σύνολα $R^{(k)}$ σύμφωνα με τη σχέση:

$R^{(k)} = Op^{(r)}(S^{(k)}, R^{(k+1)})$ με $R^{(N)} = Op^{(r)}(S^{(N)}, M)$ και $M \in S^{(N)}$. Ανάλογα με την επιλογή του σημείου M δημιουργούμε εν γένει μια διαφορετική ακολουθία συνόλων $R^{(k)}$.

Ορίζουμε επίσης τις ποσότητες::

$$a^{(k)} = \|R^{(k)} - R^{(k+1)}\| \text{ με } a^{(N)} = 0 \quad (\text{B5.1})$$

$$c^{(k)} = \frac{4 \cdot \pi \cdot \|R^{(k)}\|}{\|P^{(k)}\|^2} \quad (\text{B5.2})$$

όπου $P^{(k)}$ αποτελείται από τα συνοριακά στοιχεία του συνόλου $R^{(k)}$.

Η σχέση (1) εκφράζει τον αριθμό των στοιχείων που έχουν προστεθεί στο σύνολο $R^{(k)}$ κατά την εφαρμογή του τελεστή $Op^{(r)}$ και η σχέση (2) τον βαθμό κυκλικότητας του συνόλου $R^{(k)}$.

Αν D και E είναι οι εικόνες που προήλθαν από την εικόνα I με εφαρμογή των τελεστών *dilation* και *erosion* αντίστοιχα και $G = D - E$ η εικόνα των διαφορών τους η παρακάτω ποσότητα δίνει μια ένδειξη για την ύπαρξη ακμής στην περιοχή της εικόνας I που ορίζεται από το σύνολο $R^{(k)} - R^{(k+1)}$:

$$b^{(k)} = \max G\{R^{(k)} - R^{(k+1)}\} \quad (\text{B5.3})$$

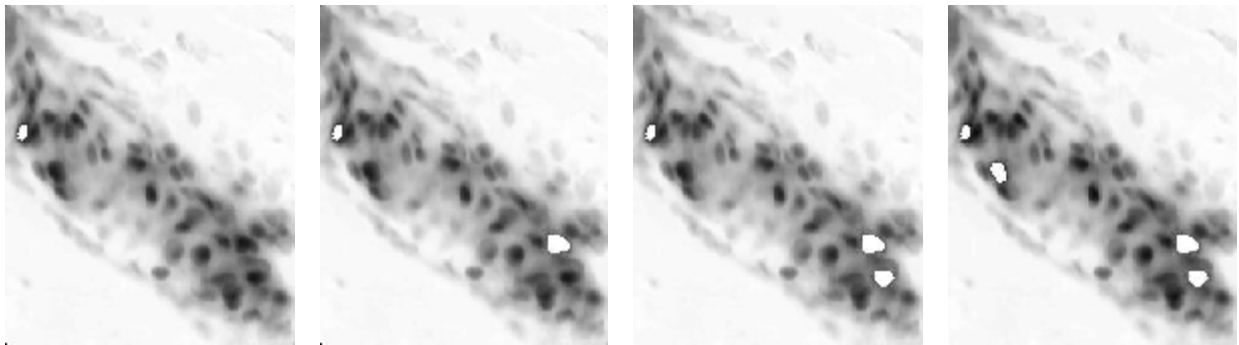
Η διαδικασία ανίχνευσης των κυττάρων προσδιορίζεται από τα παρακάτω βήματα:

(1) Επιλέγουμε ένα σημείο $M \in S^{(N)}$ και υπολογίζουμε το σύνολο $R^{(N)} = Op^{(r)}(S^{(N)}, M)$

- (2) Επαναλαμβάνουμε τη διαδικασία $R^{(k)} = Op^{(r)}(S^{(k)}, R^{(k+1)})$ για όλα τα k μέχρι έως ότου εντοπιστεί απότομη μεταβολή στη ποσότητα $a^{(k)}$ ταυτόχρονα με μείωση της ποσότητας $c^{(k)}$ η μεγάλη τιμή της ποσότητας $b^{(k)}$. Τα ανωτέρω υποδηλώνουν ενοποίηση δύο τμημάτων που ενδεχομένως ανήκουν σε γειτονικά κύτταρα ή η ότι επέκταση του συνόλου $R^{(k)}$ πέφτει σε ακμή.
- (3) Έστω $R^{(q)}$ το τελικό σύνολο που προέκυψε από την επανάληψη του βήματος (2). Θέτουμε $S^{(k)}(R^{(k)} \cap S^{(k)}) = 0, \forall k \geq q$
- (4) Επαναλαμβάνουμε τα βήματα (1)-(4). Αν $\|S^{(N)}\| = 0$ στο βήμα (1) επιλέγεται κάποιο $M \in S^{(N-1)}$ και αν $\|S^{(N-1)}\| = 0$ επιλέγεται $M \in S^{(N-2)}$ και οκ.

Τα σύνολα $R^{(q)}$ που προκύπτουν από την εφαρμογή της διαδικασίας αποτελούν τις περιοχές των εντοπισμένων κυττάρων. Προφανώς χαρακτηριστικά όπως το μέγεθος των κυττάρων, η κυκλικότητα και η μέση φωτεινότητα μπορούν να εξαχθούν άμεσα με τη βοήθεια των συνόλων $R^{(q)}$.

Η συνολική διαδικασία επιδεικνύεται στις παρακάτω εικόνες:

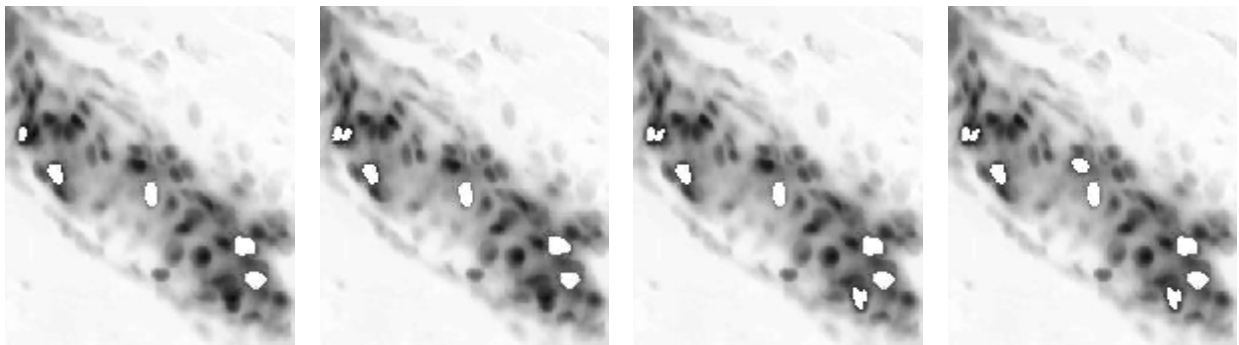


(α)

(β)

(γ)

(δ)

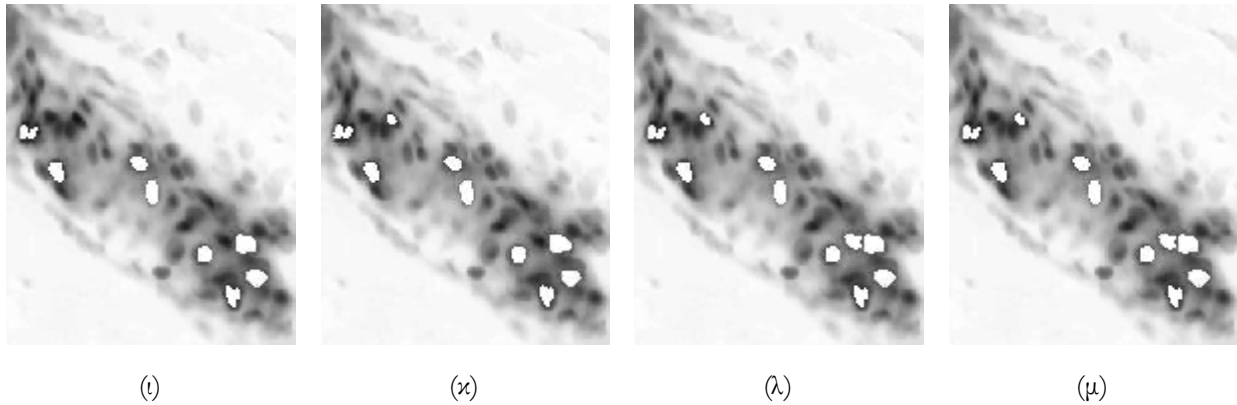


(ε)

(ζ)

(η)

(θ)



Εικόνα B5.1: Διαδοχικά στάδια εντοπισμού των πυρήνων των κυττάρων

Στην Εικόνα B5.1(α) ο πυρήνας που ανιχνεύθηκε γειτιιάζει άμεσα με κάποιον άλλο. Η ταυτόχρονη απότομη μεταβολή στο μέγεθος και την κυλιτικότητα του κυττάρου είναι η ένδειξη για τη διακοπή της διαδικασίας *opening by reconstruction* στο κρίσιμο μέγεθος. Στην Εικόνα B5.1(β) ο πυρήνας που ανιχνεύθηκε δεν γειτιιάζει άμεσα με κάποιον άλλο με αποτέλεσμα η διαδικασία να σταματά με το κριτήριο της πρόσκρουσης σε ακμή. Αυτό αναγκάζει τον πυρήνα να εξαντλεί πλήρως την περιοχή που καταλαμβάνει, σε αντίθεση με την περίπτωση της Εικόνας B5.1(α) στην οποία το μέγεθος του πυρήνα που ανιχνεύθηκε είναι σχετικά μικρό. Αντίστοιχες είναι οι περιπτώσεις των εικόνων B5.1(γ) και (ε) ενώ η περίπτωση της Εικόνας B5.1(δ) μοιάζει με αυτή της B5.1(α). Στην Εικόνα B5.1(ζ) ανιχνεύεται πυρήνας κοντά σε ήδη εντοπισμένο πυρήνα. Η διαδικασία τερματίζεται με το κριτήριο της πρόσκρουσης σε ακμή –εξαιτίας και της παρουσίας του ανιχνευμένου πυρήνα.

3. ΤΑΞΙΝΟΜΗΣΗ ΚΥΤΤΑΡΩΝ – ΣΥΣΤΗΜΑ ΠΡΟΓΝΩΣΗΣ

Όπως έχει ήδη αναφερθεί το στάδιο ανίχνευσης των πυρήνων είναι το πρώτο στάδιο της διαδικασίας πρόγνωσης της εξέλιξης της ασθένειας. Στην παράγραφο αυτή και για την πληρότητα της παρουσίασης συνοψίζεται ο τρόπος με τον οποίο πραγματοποιείται η πρόγνωση. Για περισσότερες πληροφορίες ο αναγνώστης παραπέμπεται στην εργασία [33].

Χαρακτηριστικά για την ταξινόμηση των πυρήνων

Για κάθε εντοπισμένο πυρήνα υπολογίζονται οι μέσες τιμές των χρωματικών συνιστωσών Y , C_r και C_b , καθώς και το χαρακτηριστικό υψός T_{xt} το οποίο υπολογίζεται από τη σχέση:

$$T_{xt} = 1 - \frac{1}{1 + \sigma^2(\mathbf{Y}_{probe})} \quad (B5.4)$$

όπου σ^2 είναι η διασπορά των τιμών του καναλιού Y στη περιοχή που καταλαμβάνεται από τον πυρήνα (\mathbf{Y}_{probe}).

Ταξινόμηση των πυρήνων

Με χρήση ενός νευρωνικού δικτύου RBF (Radial Basis Function) οι πυρήνες ταξινομούνται σε μια από τις επόμενες κατηγορίες ανάλογα με τον βαθμό εμπότισης: Μηδενικός, μικρός, ενδιάμεσος, υψηλός, πολύ υψηλός.

Υπολογισμός του δείκτη διάγνωσης

Ο δείκτης διάγνωσης εξάγεται με βάση το ποσοστό των πυρήνων που έχουν ταξινομηθεί σε κάθε κατηγορία και με χρήση του Πίνακα Β5.1. Για παράδειγμα αν σε μια εικόνα έχει βρεθεί θετικό ένα συνολικό ποσοστό 45% των πυρήνων με κατανομή στο βαθμό εμπότισης: 20% => Μικρός, 15% => Ενδιάμεσος, 10% => Υψηλός, τότε η συνολική επίδοση θα είναι $(1 \times 1) + (1 \times 2) + (1 \times 3) = 6$ η οποία αντιστοιχεί σε δείκτη διάγνωσης 2+.

Θετικά κύτταρα (%)	Επίδοση	Βαθμός Εμπότισης	Επίδοση	Συνολική Επίδοση	Δείκτης Διάγνωσης
0	0	Μηδενικός	0	0	0
0 – 25 %	1	Μικρός	1	1 – 4	1+
26 – 50 %	2	Ενδιάμεσος	2	5 – 8	2+
51 – 75 %	3	Υψηλός	3	9 – 12	3+
≥ 76 %	4	Πολύ Υψηλός	4	≥ 13	4+

Πίνακας Β5.1: Υπολογισμός του δείκτη διάγνωσης

4. ΔΙΑΔΙΚΑΣΙΑ ΕΠΑΛΗΘΕΥΣΗΣ – ΣΥΜΠΕΡΑΣΜΑΤΑ

Η επαλήθευση τη αποδοτικότητας ανίχνευσης των πυρήνων πραγματοποιήθηκε σύγκριση ανάμεσα στο προτεινόμενο σύστημα και δύο ειδικούς. Συγκεκριμένα δυο γιατροί σημείωσαν –σε χωριστές διαδικασίες– τους πυρήνες σε 28 εικόνες βιοψιών. Οι εικόνες ήταν έγχρωμες με βάθος χρώματος 24bit και χωρικής ανάλυσης 480×640. Η τάση των ειδικών είναι να μην σημειώνουν όλους τους πυρήνες στις εικόνες και αυτό συμβαίνει για δύο λόγους: (α) Λόγω της κούρασης που υπεισέρχεται στη διαδικασία προσημείωσης των πυρήνων επιλέγεται ένα κατά το δυνατό αντιπροσωπευτικό υποσύνολο. Η αντιμετώπιση αυτή είναι επιρρεπής σε κρίσιμα σφάλματα εξαιτίας του τρόπου πρόβλεψης που περιγράφηκε στην παράγραφο 3. (β) Ορισμένοι από τους εμφανιζόμενους πυρήνες δημιουργούνται εξαιτίας του πάχους των slides των βιοψιών ή δεν είναι πραγματικοί πυρήνες αλλά άλλα στοιχεία. Ο αλγόριθμος από τον ορισμό του δεν μπορεί να έχει πρότερη γνώση ώστε να μπορεί να διακρίνει τους πραγματικούς πυρήνες από τα άλλα στοιχεία αλλά εντοπίζει όλα τα διακεκριμένα –πλήρως ή μερικώς– αντικείμενα. Το γεγονός αυτό τον καθιστά χρησιμότερο σε παράλληλη με τους ειδικούς εργασία παρά αυτόνομη. Η ιδέα είναι πολύ απλή: Μετά τον εντοπισμό από το σύστημα όλων των αντικειμένων οι ειδικοί μπορούν να επιβεβαιώσουν ποιο από αυτά αντιστοιχούν σε πυρήνες. Δεδομένου ότι η πλειοψηφία των εμφανιζόμενων στις βιοψίες αντικειμένων, είναι πράγματι πυρήνες η διαδικασία απόρριψης είναι λιγότερο κουραστική από τη διαδικασία προσημείωσης.

Επιπλέον με τον τρόπο αυτό μειώνεται και η ασυμφωνία μεταξύ των ειδικών όσον αφορά τους πραγματικούς πυρήνες.

Για την αριθμητική αξιολόγηση της επίδοσης του αλγορίθμου ορίζουμε δύο μετρήσεις την *ευαισθησία* (*S* - Sensitivity) και *θετική πρόβλεψη* (*PPV* - Positive Predictive Value). *Ευαισθησία* καλείται η πιθανότητα να εντοπιστεί ένας πυρήνας στην περίπτωση που έχει προσημειωθεί από τους ειδικούς και ορίζεται από τη σχέση:

$$S = \frac{TP}{TP + FN} \quad (B5.5)$$

όπου *TP* (true positive) είναι ο αριθμός των πυρήνων που εντοπίστηκαν τόσο από τον αλγόριθμο όσο και από τους ειδικούς και *FN* (false negative) είναι ο αριθμός των πυρήνων που προσημειώθηκαν από τους ειδικούς χωρίς να εντοπιστούν από τον αλγόριθμο.

Θετική πρόβλεψη καλείται η πιθανότητα ο εντοπισμένος από τον αλγόριθμο πυρήνας να αντιστοιχεί σε πυρήνα που έχει προσημειωθεί από τους ειδικούς και δίνεται από τη σχέση:

$$PPV = \frac{TP}{TP + FP} \quad (B5.6)$$

όπου *FP* (false positive) είναι ο αριθμός των πυρήνων που εντοπίστηκαν από τον αλγόριθμο αλλά δεν προσημειώθηκαν από τους ειδικούς.

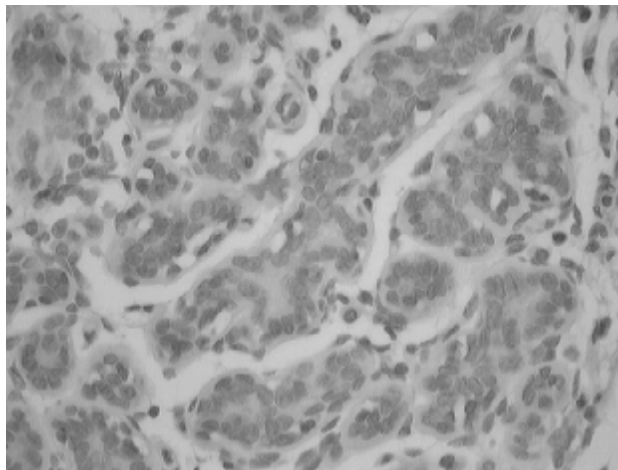
Στο Πίνακα B5.1 παρουσιάζονται τα συνολικά αποτελέσματα της διαδικασίας αξιολόγησης. Ο αλγόριθμος ελέγχθηκε σε σχέση με κάθε ένα από τους ειδικούς αλλά και σε σχέση με τον συνδυασμό τους –λογικό OR. Για σκοπούς σύγκρισης παρουσιάζεται και η επίδοση του ενός ειδικού σε σχέση με τον άλλο. Το πρώτο σημαντικό στοιχείο που εμφανίζεται στον Πίνακα B5.2 είναι η ασυμφωνία μεταξύ των ειδικών που εκφράζεται κυρίως από την υψηλή τιμή της τυπικής απόκλισης. Το γεγονός αυτό δηλώνει ότι δεν υπάρχει πλήρης βεβαιότητα ούτε στους ειδικούς για το ποια από τα απεικονιζόμενα στις βιοψίες στοιχεία είναι πυρήνες και ποια όχι. Η συνεισφορά του συστήματος προς την κατεύθυνση αυτή είναι εξαιρετικά σημαντική. Η δεύτερη παρατήρηση είναι η αδυναμία του προτεινόμενου αλγορίθμου να εντοπίσει πλήρως όλους τους πυρήνες που προσημειώθηκαν από τους ειδικούς. Παρά την υψηλή τιμή της, η *ευαισθησία* δεν φθάνει το 100% όπως θα περίμενε κανείς με βάση τον τρόπο λειτουργίας του αλγορίθμου. Αυτό οφείλεται κυρίως στην αποτυχία διαχωρισμού πολύ γειτονικών ή και επικαλυπτόμενων πυρήνων σε εικόνες με χαμηλό contrast –βλέπε Εικόνα B5.2- τους οποίους οι ειδικοί είναι ικανοί να διακρίνουν. Η τιμή της *θετικής πρόβλεψης* είναι σχετικά χαμηλή, όταν υπολογίζεται σε σχέση με κάποιον από τους ειδικούς ξεχωριστά, εξαιτίας της πολύ υψηλής τιμής της παραμέτρου *FP*. Όταν η σύγκριση πραγματοποιείται και με τους δύο ειδικούς ταυτόχρονα η τιμή της *θετικής πρόβλεψης* ανεβαίνει σημαντικά.

Το συνολικό συμπέρασμα είναι ότι από τον τρόπο σχεδιασμού του ο αλγόριθμος έχει υψηλά ποσοστά ευαισθησίας και σχετικά χαμηλότερη επίδοση όσον αφορά τη θετική πρόβλεψη. Με βάση το γεγονός αυτό μπορεί να χρησιμοποιηθεί σε συνδυασμό με τους ειδικούς παρά αυτόνομα. Αυτό συμφωνεί και με την

υπάρχουσα θέση των ειδικών –ιατρών- οι οποίοι αποφεύγουν να εμπιστευθούν απόλυτα τέτοια υπολογιστικά συστήματα αλλά επιθυμούν να έχουν σε κάθε περίπτωση την τελική επίβλεψη.

	S +/- SD	PPV +/- SD
Exp1 - Exp2	72.6 +/- 15.1	74.5 +/- 14.3
System - Exp1	84.6 +/- 8.7	61.7 +/- 15.2
System - Exp2	86.8 +/- 7.4	65.8 +/- 14.7
System – OREx	83.3 +/- 9.6	75.6 +/- 13.1

Πίνακας B5.2: Επίδοση της ανίχνευσης κυττάρων



Εικόνα B5.2: Εικόνα βιοψίας με χαμηλό contrast και πολλούς επικαλυπτόμενους πυρογήνες

Η ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΣΩΠΩΝ: ΜΑΘΗΜΑΤΙΚΟ ΥΠΟΒΑΘΡΟ

Στο κεφάλαιο αυτό παρουσιάζεται η γενική δομή των συστημάτων αναγνώρισης προσώπων, δίνονται κάποιοι ορισμοί και εισαγωγικές έννοιες και παρουσιάζονται αναλυτικά κάποια σχετικά δημοφιλή σχήματα.

1. ΕΙΣΑΓΩΓΗ

Οι μεγάλες διακυμάνσεις στα ανθρώπινα πρόσωπα, όπως μεταβολές στις εκφράσεις, στις συνθήκες φωτισμού, τη θέση παρατήρησης, την ηλικία και η παρουσία αμφιέσης ή η επίδραση θορύβου καθιστούν την αναγνώριση προσώπων ένα πραγματικά πολύ δύσκολο πεδίο έρευνας δεδομένου ότι ένα τέτοιο σύστημα πρέπει να είναι ικανό να πραγματοποιεί επιτυχημένες ταυτοποιήσεις ανεξάρτητα από διακυμάνσεις αυτές. Η πλειονότητα των συστημάτων αναγνώρισης προσώπου που έχουν προταθεί σε ερευνητικές εργασίες θέτουν αυστηρούς περιορισμούς όπως ομοιόμορφο φόντο, οι φωτογραφίες προσώπων να είναι φωτογραφίες εμπρόσθιας όψης, τα πρόσωπα στις εικόνες να βρίσκονται στο κέντρο και υπό μικρή κλίση και περιστροφή κλπ [1]. Οι μέθοδοι αυτοί επικεντρώνεται στην εξαγωγή χαρακτηριστικών, είτε γεωμετρικών είτε μέσω γραμμικών μετασχηματισμών [3][4][6]. Μια από τις δημοφιλέστερες προσεγγίσεις, με ισχυρό θεωρητικό υπόβαθρο, είναι η τεχνική των ιδιοδιανυσμάτων η οποία αναπαριστά τα πρόσωπα σε ένα χώρο μικρότερης διάστασης [6]. Ο μετασχηματισμός Karhunen-Loeve και η Principal Component Analysis αποτελούν την καρδιά της αναπαράστασης των ιδιοδιανυσμάτων και χρησιμοποιούνται τόσο για εντοπισμό, όσο και για αναγνώριση προσώπων [4][6], με μάλλον αποδοτικότερη την διαδικασία εντοπισμού.

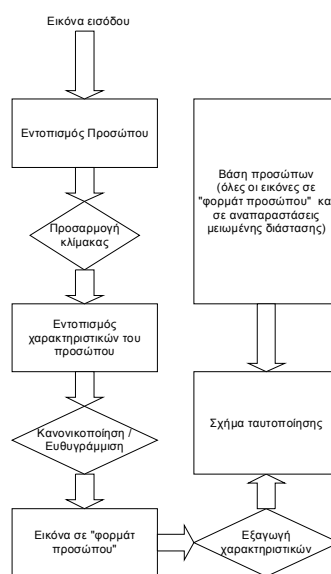
Η εξαγωγή χαρακτηριστικών μπορεί να υπερκεράσει προβλήματα που σχετίζονται με τη μεγάλη διάσταση του χώρου των εικόνων όπως η υπολογιστική πολυπλοκότητα και η ικανότητα γενίκευσης των ταξινομητών που χρησιμοποιούνται για την κατηγοριοποίηση προσώπων [15] [16]. Τα χαρακτηριστικά που επιλέγονται πρέπει να οδηγούν στην περιγραφή κάθε προσώπου με μοναδικό τρόπο και να επιτρέπουν την ταυτοποίηση του μέσω ενός χώρου μικρότερης διάστασης. Επιπλέον τα χαρακτηριστικά αυτά πρέπει να μπορούν να εξάγονται αυτόματα από το υπολογιστικό σύστημα που χρησιμοποιείται για την ταυτοποίηση. Η εύρεση τέτοιων χαρακτηριστικών είναι το βασικότερο πρόβλημα των συστημάτων αναγνώρισης προσώπων.

Μετά την επιλογή της αναπαράστασης των προσώπων μέσω ενός συνόλου χαρακτηριστικών απαιτείται εκτεταμένη επεξεργασία έτσι ώστε να απαλειφθούν εκείνοι οι παράγοντες οι οποίοι αυξάνουν τις εντός κλάσεων διακυμάνσεις. Η επιλογή του διάνυσματος χαρακτηριστικών είναι ένας πραγματικός γρίφος όταν πρέπει να ληφθούν υπόψη τα προβλήματα που δημιουργούνται από τις μεταβολές στην οριοθέτηση, κλίμακα, φωτισμό και υφή των προσώπων. Με άλλα λόγια το διάνυσμα χαρακτηριστικών πρέπει να υποστεί εκτεταμένη επεξεργασία ώστε να είναι ανεπηρέαστο από τους παραπάνω παράγοντες αλλά ταυτόχρονα να διατηρεί τις ιδιότητες εκείνες που το καθιστούν ικανό να οδηγήσει σε επιτυχείς ταυτοποιήσεις. Οι ανωτέρω

προσεγγίσεις απαιτούν τα πρόσωπα στις βάσεις να είναι σε κατάλληλο «φορμάτ» και να χαρακτηρίζονται από μια χονδρική περιγραφή για σκοπούς δεικτοδότησης.

2. ΣΥΣΤΗΜΑΤΑ ΑΝΑΓΝΩΡΙΣΗΣ ΠΡΟΣΩΠΩΝ

Για την επίτευξη υψηλότερων ποσοστών αναγνώρισης οι περισσότεροι αλγόριθμοι απαιτούν την μετατροπή των φωτογραφιών προσώπων σε ένα κατάλληλο φορμάτ το οποίο καλείται «φορμάτ προσώπου- head format». Σε αυτό τα απεικονιζόμενα πρόσωπα παρουσιάζονται σε συγκεκριμένη κλίμακα και οριοθέτηση και κάποια βασικά χαρακτηριστικά προσώπου –συνήθως τα μάτια- είναι ευθυγραμμισμένα. Ένα γενικό διάγραμμα που απεικονίζει την διαδικασία αναγνώρισης προσώπων φαίνεται στο Σχήμα Γ1.1.



Σχήμα Γ1.1: Η αρχιτεκτονική ενός γενικού συστήματος αναγνώρισης προσώπων.

Το βασικό τμήμα ενός συστήματος αναγνώρισης προσώπων είναι το τμήμα εξαγωγής χαρακτηριστικών με βάση τα οποία θα αναπαριστάται το πρόσωπο. Η διάσταση του διανύσματος χαρακτηριστικών πρέπει να είναι σημαντικά μικρότερη από την διάσταση των εικόνων που περιέχουν το πρόσωπο. Αναγνώριση στο πεδίο της φωτεινότητας απαιτεί συσχέτιση στο χώρο των εικόνων η οποία είναι αναποτελεσματική και υπολογιστικά πολύπλοκη, καθώς απαιτεί τη σύγκριση του προσώπου, του οποίου η ταυτότητα διερευνάται, με όλα τα πρόσωπα τα οποία βρίσκονται αποθηκευμένα στη βάση. Η αποθήκευση ολόκληρων των φωτογραφιών σε μια βάση δεδομένων είναι σε κάθε περίπτωση επιθυμητή αλλά τα διανύσματα χαρακτηριστικών επιτρέπουν πολύ σημαντικές διαδικασίες όπως η δεικτοδότηση.

3. ΜΑΘΗΜΑΤΙΚΗ ΠΕΡΙΓΡΑΦΗ ΤΗΣ ΑΝΑΓΝΩΡΙΣΗΣ ΠΡΟΣΩΠΩΝ

Για όλους σχεδόν τους αλγόριθμους αναγνώρισης προσώπων η επιτυχία βασίζεται στην αποδοτικότερη λύση δύο επιμέρους προβλημάτων: της αναπαράστασης και της ταύτισης.

Στις τεχνικές αναγνώρισης προσώπων που βασίζονται σε πρότυπα και όχι σε γεωμετρικά χαρακτηριστικά, το πρόσωπο θεωρείται ένας διδιάστατος πίνακας τα στοιχεία του οποίου αντιστοιχούν στις τιμές φωτεινότητας των pixels. Μαθηματικά αυτό γράφεται ως:

$$x = \{x_i, i \in S\} \quad (\Gamma 1.1)$$

όπου S είναι ένα τετραγωνικό πλέγμα. Πολλές φορές εντούτοις είναι καταλληλότερο να εκφράζουμε το x ως ένα μονοδιάστατο διάνυσμα στήλη με λεξιλογραφική διάταξη των pixels των γραμμών της εικόνας, δηλαδή:

$$x = [x_1, x_2, \dots, x_n]^T \quad (\Gamma 1.2)$$

όπου $n = \|S\|$ είναι ο συνολικός αριθμός των pixels της εικόνας. Επομένως θα ισχύει $x \in \mathbb{R}^n$ όπου \mathbb{R}^n ο n -διάστατος Ευκλείδειος χώρος. Με δεδομένη μια αναπαράσταση προσώπων σε κάποιο χώρο δύο πράγματα είναι ιδιαίτερα σημαντικά: η *διαχωριστικότητα* και η *αποδοτικότητα*. Η *διαχωριστικότητα* αναφέρεται στο πόσο μακριά είναι οι αναπαραστάσεις διαφορετικών προσώπων στο χώρο ενώ η *αποδοτικότητα* αναφέρεται στο πόσο συμπαγής είναι χώρος δηλαδή πόσο ικανοποιητικά μπορεί να αναπαρασταθούν τα πρόσωπα στο χώρο αυτό.

Παρόλο που με βάση τις ικανότητες των σύγχρονων υπολογιστικών συστημάτων, οι αναπαραστάσεις στο αρχικό χώρο της εικόνας όπως, περιγράφονται από τις σχέσεις (Γ1.1) και (Γ1.2), είναι εφικτές, οι περισσότερες από τις τεχνικές αναγνώρισης προσώπων χρησιμοποιούν την αναπαράσταση των προσώπων μέσω κάποιου *διανύσματος χαρακτηριστικών*.

$$f(x) = [f_1(x), f_2(x), \dots, f_m(x)]^T \quad (\Gamma 1.3)$$

όπου $f_1(\cdot), f_2(\cdot), \dots, f_m(\cdot)$ είναι κάποιες γραμμικές ή μη συναρτήσεις. Δεδομένου ότι στη πράξη ισχύει $m \ll n$ οι αναπαραστάσεις μέσω διανύσματος χαρακτηριστικών είναι αποδοτικότερες. Ένας απλός τρόπος για την αύξηση της αποδοτικότητας των αναπαραστάσεων είναι η χρήση εναλλακτικών ορθοκανονικών βάσεων για το χώρο \mathbb{R}^n . Συγκεκριμένα έστω ότι τα διανύσματα $e_1, e_2, e_3, \dots, e_n$ αποτελούν μια ορθοκανονική βάση τότε το διάνυσμα x μπορεί να εκφραστεί από τη σχέση:

$$x = \sum_{i=1}^n \hat{x}_i \cdot e_i \quad (\Gamma 1.4)$$

όπου τα $\hat{x}_i = \langle x \cdot e_i \rangle$ λαμβάνονται από το εσωτερικό γινόμενο των διανυσμάτων x και e_i και το διάνυσμα x μπορεί να αναπαρασταθεί ως $x = [\hat{x}_1, \hat{x}_2, \dots, \hat{x}_n]^T$. Δυο κλασικές ορθοκανονικές βάσεις είναι η φυσική η οποία χρησιμοποιείται στη σχέση (Γ1.2) και για την οποία είναι $e_i = [0, 0, \dots, 0, 1, 0, \dots, 1]^T$ με τη μονάδα στην i -στή θέση και η βάση Fourier με $e_i = \frac{1}{\sqrt{n}} [1, e^{j2\pi \frac{i}{n}}, e^{j2\pi \cdot 2 \frac{i}{n}}, \dots, e^{j2\pi \cdot (n-1) \frac{i}{n}}]^T$. Αν για κάποια δεδομένη ορθοκανονική βάση ισχύει $\hat{x}_i \approx 0$ όταν $i \geq m$ τότε το διάνυσμα x μπορεί να αναπαρασταθεί αποδοτικά από ένα το m -διάστατο διάνυσμα:

$$\hat{x} = [\hat{x}_1, \hat{x}_2, \dots, \hat{x}_m]^T \quad (\Gamma 1.5)$$

Πρέπει πάντως να τονίσουμε ότι μια αποδοτική αναπαράσταση δεν έχει υποχρεωτικά και υψηλή διαχωριστικότητα [21].

Στο πρόβλημα της ταύτισης ένα νέο πρόσωπο εισόδου αναγνωρίζεται μετά την ταυτοποίηση του με κάποιο αποθηκευμένο στη βάση. Έστω για παράδειγμα ότι το πρόσωπο εισόδου συμβολίζεται από το διάνυσμα x και τα K αποθηκευμένα στη βάση πρόσωπα συμβολίζονται με $c_k, k = 1, 2, \dots, K$, τότε το x ταυτοποιείται με το c_{k_0} αν ικανοποιείται η σχέση:

$$k_0 = \arg \min_k d(x, c_k) \quad (\Gamma 1.6)$$

όπου $d(x, c_k)$ η απόσταση των διανυσμάτων x και c_k με βάση κάποια μετρική. Αν χρησιμοποιηθεί ως μετρική η Ευκλείδεια απόσταση το χώρο \mathfrak{R}^n ($\|\cdot\|$) η σχέση (Γ1.6) γίνεται:

$$k_0 = \arg \min_k \|x - c_k\| \quad (\Gamma 1.7)$$

και αν τα διανύσματα c_k είναι κανονικοποιημένα ώστε $\|c_k\| = c$ για κάθε k τότε η ελαχιστοποίηση της σχέσης (Γ1.7) ταυτίζεται με την μεγιστοποίηση της συσχέτισης των διανυσμάτων x και c_k :

$$k_0 = \arg \max_k \langle x, c_k \rangle \quad (\Gamma 1.8)$$

Δεδομένου ότι η απόσταση και το εσωτερικό γινόμενο παραμένουν αναλλοίωτα στην αλλαγή της ορθοκανονικής βάσης οι σχέσεις (Γ1.7) και (Γ1.8) μπορούν να χρησιμοποιηθούν και για τις προβολές των διανυσμάτων x και c_k σε κάθε ορθοκανονική βάση.

Δεδομένων των διακυμάνσεων ως προς την οριοθέτηση, τις γωνίες κλίσης και περιστροφής, το φωτισμό και τη παρουσία θορύβου, εικόνες προσώπων του ίδιου ατόμου παρουσιάζουν τυχαίες μεταβολές και είναι προτιμότερο να θεωρούνται ως στοχαστικές ανελίξεις. Υπό αυτή τη θεώρηση η χρήση του κριτηρίου της μεγίστης πιθανοφάνειας (ML-maximum likelihood) είναι εύλογη:

$$k_0 = \arg \min_k \log p(x | c_k) \quad (\Gamma 1.9)$$

όπου $p(x | c_k)$ είναι η πιθανότητα του x με δεδομένο ότι είναι το k -στό πρόσωπο. Το κριτήριο ML ελαχιστοποιεί το σφάλμα ταύτισης όταν το πρόσωπο εισόδου έχει την ίδια πιθανότητα να αντιστοιχεί σε οποιοδήποτε από τα αποθηκευμένα στη βάση πρόσωπα.

3.1 Ο μετασχηματισμός Karhunen-Loeve

Όπως αναφέρθηκε νωρίτερα η χρήση εναλλακτικών ορθοκανονικών βάσεων μπορεί να χρησιμοποιηθεί για τη δημιουργία συμπαγούς αναπαράστασης των διανυσμάτων των προσώπων. Μια τέτοια ορθοκανονική βάση προκύπτει από τον διακριτό μετασχηματισμό Karhunen-Loeve. Σε αντίθεση με τους κλασικούς μετασχηματισμούς διανυσμάτων όπως ο Fourier και ο DCT, ο μετασχηματισμός Karhunen-Loeve

κατασκευάζεται με βάση τις στατιστικές ιδιότητες των ιδίων των διανυσμάτων. Έστω ότι τα διανύσματα προσώπων αναπαριστώνται από τη στοχαστική ανέλιξη x . Το μέσο διάνυσμα του πληθυσμού δίνεται από τη σχέση:

$$m_x = E\{x\} \quad (\Gamma 1.10)$$

όπου $E\{\cdot\}$ συμβολίζει την αναμενόμενη τιμή. Ο πίνακας συμμεταβλητότητας του πληθυσμού δίνεται από τη σχέση:

$$C = E\{(x - m_x) \cdot (x - m_x)^T\} \quad (\Gamma 1.11)$$

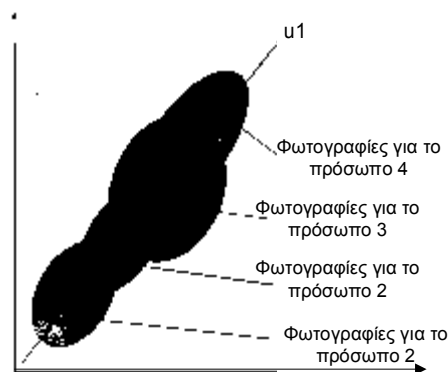
Η ορθοκανονική βάση του KL αποτελείται από τα ιδιοδιανύσματα του πίνακα C —τα οποία καλούνται και eigenfaces εξαιτίας της μορφής τους. Αν τα ιδιοδιανύσματα του C είναι u_1, u_2, \dots, u_n και αντιστοιχούν στις ιδιοτιμές $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ τότε το x εκφράζεται ως:

$$x = \sum_{i=1}^n \hat{x}_i \cdot u_i \quad (\Gamma 1.12)$$

και μια συμπαγής αναπαράσταση του—υπό την προϋπόθεση ότι λ_i είναι πολύ μικρό για $i > m$ —είναι:

$$x \approx \sum_{i=1}^m \hat{x}_i \cdot u_i, \quad \hat{x} = [\hat{x}_1, \hat{x}_2, \dots, \hat{x}_m]^T \quad (\Gamma 1.13)$$

Η αναπαράσταση που δίνεται από τη σχέση (Γ1.13) είναι η αποδοτικότερη δυνατή για όλα τα $m < n$ γιατί δίνει το μικρότερο μέσο τετραγωνικό σφάλμα για όλες τις αναπαραστάσεις του x με διανύσματα m στοιχείων. Δεδομένου ότι η ιδιοτιμή λ_i εκφράζει τη διασπορά του πληθυσμού κατά μήκος του άξονα u_i και ότι η εντροπία—ποσό πληροφορίας—αυξάνει με τη διασπορά, είναι προφανές ότι η προβολή στους m βασικούς άξονες διατηρεί το μεγαλύτερο ποσοστό της πληροφορίας του πληθυσμού. Τα ανωτέρω βέβαια δεν σημαίνουν σε καμία περίπτωση ότι ο μετασχηματισμός KL είναι βέλτιστος και ως προς την διαχωρισιμότητα, η οποία εξαρτάται από την διασπορά ανάμεσα στις αναπαραστάσεις διαφορετικών προσώπων και όχι από τη συνολική διασπορά του πληθυσμού—βλέπε Εικόνα Γ1.1.



Εικόνα Γ1.1: Η συνολική διασπορά του πληθυσμού δεν σημαίνει αναγκαστικά και διασπορά των επιμέρους κλάσεων

Η ταύτιση ενός προσώπου εισόδου με κάποιο από τη αποθηκευμένα πραγματοποιείται με τη χρήση του κριτηρίου:

$$k_0 = \arg \min_k \|\hat{x} - \hat{c}_k\| \quad (\Gamma 1.14)$$

Για την χρήση του μετασχηματισμού KL στη πράξη χρειάζεται η εκτίμηση του μέσου διανύσματος m_x και του πίνακα C . Για το σκοπό αυτό χρησιμοποιείται ένα σύνολο από N εικόνες προσώπων $\{I_1, I_2, \dots, I_N\}$ όπου $I_i \in \mathcal{R}^{l \times m}$ ως σύνολο μάθησης. Μετά από λεξιλογραφική διάταξη των pixels κάθε εικόνας I_i δημιουργούμε ένα σύνολο από διανύσματα x_1, x_2, \dots, x_N όπου $x_i \in \mathcal{R}^n$ και $n = l \cdot m$. Το μέσο διάνυσμα m_x και ο πίνακας C εκτιμώνται από τις σχέσεις:

$$m_x = \sum_{i=1}^N x_i \quad (\Gamma 1.15)$$

$$C = \frac{1}{N} \sum_{i=1}^N (x_i - m_x) \cdot (x_i - m_x)^T = \frac{1}{N} X X^T \quad (\Gamma 1.16)$$

όπου $X = [x_1 - m_x, x_2 - m_x, \dots, x_N - m_x]$. Οι συναρτήσεις βάσης του KLT υπολογίζονται από τη λύση του προβλήματος ιδιοτιμών:

$$\Lambda = \Phi^T \cdot C \cdot \Phi \quad (\Gamma 1.17)$$

όπου Φ είναι ο πίνακας των ιδιοδιανυσμάτων του C , και Λ είναι ο αντίστοιχος διαγώνιος πίνακας των ιδιοτιμών.

Όπως μπορεί κάποιος να παρατηρήσει η διάσταση του πίνακα συμμεταβλητότητας είναι υπερβολικά μεγάλη, συγκεκριμένα $C \in \mathcal{R}^{n \times n}$ και επομένως η διαγωνιοποίηση του είναι υπολογιστικά πολύ απαιτητική. Για παράδειγμα για εικόνες με χωρική ανάλυση 90×100 , ο $C \in \mathcal{R}^{9000 \times 9000}$. Επιπλέον για να αποφευχθεί η μη ομαλότητα (singularity) του πίνακα C ο αριθμός των εικόνων που απαιτούνται στο σύνολο εκπαίδευσης πρέπει να είναι $N > n^2$.

Η λύση του προβλήματος διαγωνιοποίησης του πίνακα C μπορεί να πραγματοποιηθεί μέσω της αποσύνθεσης SVD του πίνακα X . Συγκεκριμένα τα ιδιοδιανύσματα του πίνακα $X X^T$ μπορούν να υπολογιστούν από τα ιδιοδιανύσματα του πίνακα $X^T X \in \mathcal{R}^{N \times N}$. Αυτό είναι πολύ ευκολότερο γιατί συνήθως ισχύει $N \ll n$ -δεν υπάρχει απεριορίστος αριθμός δειγμάτων στο σύνολο μάθησης. Αν η τάξη του πίνακα X είναι r , με $r \leq N$ τότε σύμφωνα με τη θεωρία SVD ο X εκφράζεται ως:

$$X = \sum_{k=1}^r \sqrt{\lambda_k} \cdot u_k \cdot v_k^T \quad (\Gamma 1.18)$$

όπου u_k και v_k είναι τα ιδιάζοντα διανύσματα που αντιστοιχούν στις μη μηδενικές ιδιοτιμές λ_k των πινάκων XX^T , $X^T X$ αντίστοιχα. Πολλαπλασιάζοντας τα δύο μέρη της σχέσης (Γ1.18) από δεξιά με v_k βρίσκουμε ένα εναλλακτικό τρόπο υπολογισμού των ιδιοδιανυσμάτων u_k :

$$\frac{1}{\sqrt{\lambda_k}} X \cdot v_k = u_k \quad (\Gamma 1.19)$$

3.2 Fisher Linear Discriminant –FLD

Μια άλλη μέθοδος επιλογής διανύσματος χαρακτηριστικών είναι η μέθοδος FLD. Έστω οι κλάσεις εικόνων $\omega_1, \omega_2, \dots, \omega_L$ με πληθικούς αριθμούς N_1, N_2, \dots, N_L αντίστοιχα. Έστω επίσης M_1, M_2, \dots, M_L οι μέσες εικόνες των κλάσεων αυτών και M η ολική μέση εικόνα. Ορίζουμε τους εντός κλάσης και ανάμεσα στις κλάσεις πίνακες διασποράς $\Sigma_\omega \in \mathfrak{R}^{n \times n}$ και $\Sigma_b \in \mathfrak{R}^{n \times n}$ ως:

$$\Sigma_\omega = \sum_{i=1}^L p(\omega_i) \cdot C_i = \sum_{i=1}^L p(\omega_i) \cdot E\{[x - M_i] \cdot [x - M_i]^T | \omega_i\} \quad (\Gamma 1.20)$$

$$\Sigma_b = \sum_{i=1}^L p(\omega_i) \cdot (M_i - M) \cdot (M_i - M)^T \quad (\Gamma 1.21)$$

όπου $p(\omega_i)$ η εκ των προτέρων (a priori) πιθανότητα και C_i ο πίνακας συμμεταβλητότητας της κλάσης ω_i .

Η μέθοδος FLD δημιουργεί ένα πίνακα Ψ , τα διανύσματα του οποίου αποτελούν μια βάση –όχι ορθοκανονική– του χώρου \mathfrak{R}^n , και ο οποίος μεγιστοποιεί τον λόγο $\frac{|\Psi^T \cdot \Sigma_b \cdot \Psi|}{|\Psi^T \cdot \Sigma_\omega \cdot \Psi|}$. Για τη μεγιστοποίηση

του λόγου αυτού ο πίνακας Ψ πρέπει να αποτελείται από τα ιδιοδιανύσματα του πίνακα $\Sigma_\omega^{-1} \cdot \Sigma_b$. Το σχετικό πρόβλημα ιδιοτιμών δίνεται από τη σχέση:

$$\Sigma_\omega^{-1} \cdot \Sigma_b \cdot \Psi = \Psi \cdot \Delta \quad (\Gamma 1.22)$$

όπου οι πίνακες $\Psi, \Delta \in \mathfrak{R}^{n \times n}$ είναι οι πίνακες ιδιοδιανυσμάτων και ιδιοτιμών του πίνακα $\Sigma_\omega^{-1} \cdot \Sigma_b$. Παρόλο που αμφότεροι οι πίνακες Σ_ω και Σ_b είναι συμμετρικοί ο πίνακας $\Sigma_\omega^{-1} \cdot \Sigma_b$ δεν είναι υποχρεωτικά συμμετρικός. Σε κάθε όμως περίπτωση οι πίνακες Ψ και Δ μπορούν να υπολογιστούν ως αποτέλεσμα ταυτόχρονης διαγωνιοποίησης των πινάκων Σ_ω και Σ_b [21].

Η μέθοδος FLD αποφεύγει ένα από τα μειονεκτήματα της PCA και συγκεκριμένα αυτό της ικανότητας διαχωρισμού των κλάσεων. Αυτό επιτυγχάνεται με την ζήτηση των αξόνων εκείνων στους οποίους μεγιστοποιείται ταυτόχρονα η ολική διασπορά του πληθυσμού αλλά και των επιμέρους κλάσεων όπως αυτές εκφράζονται από το κέντρο τους. Για την ικανοποίηση της απαίτησης αυτής καταλήγουμε σε μη ορθογώνιους άξονες.

Στην περίπτωση της αναγνώρισης προσώπων η μέθοδος FLD παρουσιάζει δύο σημαντικά προβλήματα: (α) ο μεγάλος αριθμός κλάσεων –ατόμων– με λίγες εικόνες (φωτογραφίες του ίδιου προσώπου) σε κάθε κλάση δημιουργεί δυσκολίες στην ορθή εκτίμηση των πινάκων C_i και κατά συνέπεια του πίνακα Σ_b , (β) δυσκολία στην διαγωνιοποίηση του πίνακα $\Sigma_w^{-1} \cdot \Sigma_b$ εξαιτίας του πολύ μεγάλου μεγέθους του –παρόμοιο πρόβλημα με την PCA ανάλυση μόνο που στην περίπτωση αυτή δεν υπάρχει εναλλακτικός τρόπος υπολογισμού.

Στην πράξη αυτό που γίνεται είναι να πραγματοποιείται PCA ανάλυση στο αρχικό σύνολο μάθησης για αναπαράσταση των εικόνων σε ένα χώρο μειωμένης διάστασης και στη συνέχεια εφαρμογή της μεθόδου FLD για την επιλογή των αξόνων που μεγιστοποιούν τη διασπορά των κλάσεων.

ΠΡΟΣΕΓΓΙΣΗ ΤΟΥ ΜΕΤΑΣΧΗΜΑΤΙΣΜΟΥ KL ΜΕ ΧΡΗΣΗ ΔΙΑΝΥΣΜΑΤΩΝ

Στο κεφάλαιο αυτό παρουσιάζεται μια μέθοδος προσέγγισης του μετασχηματισμού KL με βάση τα διανύσματα – στήλες των εικόνων (VKLT). Με τον τρόπο αυτό απαλείφεται η μεγάλη απαίτηση σε μνήμη και η μη ομαλότητα (singularity) του πίνακα συμμεταβλητότητας. Η αποδοτικότητα του μετασχηματισμού VKLT όσον αφορά τη συσσώρευση ενέργειας στους βασικούς άξονες προσεγγίζει την αντίστοιχη του KLT. Επιπλέον είναι ένας μετασχηματισμός ο οποίος εξαρτάται –όπως και ο KLT– από τα δεδομένα, ιδιότητα πολύ σημαντική για συστήματα ανίχνευσης προσώπων όπου απαιτείται η απόφαση κατά πόσο ένα πρότυπο είναι πρόσωπο ή όχι. Το βασικό πλεονέκτημα του VKLT σε σχέση με τον KLT είναι η μεγαλύτερη αναισθησία σε εντός κλάσης διακυμάνσεις, γεγονός που αυξάνει την ευστάθεια του συστήματος αναγνώρισης.

1. ΕΙΣΑΓΩΓΗ

Οι περισσότεροι από τους αλγόριθμους αναγνώρισης προσώπων επιτυγχάνουν μεγάλα ποσοστά αναγνώρισης μόνο κάτω από πολύ μικρές διακυμάνσεις της φωτεινότητας, των εκφράσεων του προσώπου, της περιστροφής -στροφή περί τον άξονα y - και της κλίσης -στροφή περί τον άξονα z - [3][4]. Η ανεπάρκεια των αλγορίθμων αυτών κάτω από μεγάλες αλλαγές στους ανωτέρω παράγοντες είναι λογική. Πρόσφατες μελέτες έδειξαν ότι τα προτεινόμενα σχήματα αναγνώρισης δημιουργούν μεγαλύτερες διακυμάνσεις στο ίδιο πρόσωπο υπό διαφορετικές συνθήκες φωτισμού, εκφράσεων και οριοθέτησης παρά τα διαφορετικά πρόσωπα όταν οι παραπάνω παράγοντες διατηρούνται σταθεροί [2]. Με απλά λόγια οι διακυμάνσεις εντός κλάσεως –ιδίου προσώπου– είναι περισσότερες από τις αντίστοιχες εκτός κλάσεως –διαφορετικότητα των προσώπων.

Ένα από τα πιο σημαντικά συστήματα αναγνώρισης προσώπων βασίζεται στον μετασχηματισμό Karhunen Loeve (KLT- Karhunen Loeve Transform) και είναι γνωστό με το όνομα «Eigenfaces». Όπως το όνομα της μεθόδου μαρτυρά, στηρίζεται στην Ανάλυση σε Βασικούς Άξονες (PCA- Principal Component Analysis) και πραγματοποιεί μια συμπαγή αναπαράσταση των προσώπων μέσω των συντελεστών προβολής στους βασικούς άξονες –βλέπε Κεφάλαιο Γ1. Παρόλο που η μέθοδος είναι πολύ καλά τεκμηριωμένη θεωρητικά παρουσιάζει δύο τουλάχιστον σοβαρά μειονεκτήματα: (α) Η προβολή στους βασικούς άξονες είναι βέλτιστη όσον αφορά την συσσώρευση ενέργειας και την ανακατασκευή των προσώπων αλλά δεν είναι σε καμία περίπτωση βέλτιστη όσον αφορά στην ικανότητα διαχωρισμού προσώπων –πράγμα ιδιαίτερα επιθυμητό σε συστήματα αναγνώρισης. Βεβαίως σε εφαρμογές όπως η ανίχνευση προσώπων η ικανότητα διαχωρισμού πολύ λίγη σημασία έχει, (β) Η εφαρμογή της μεθόδου απαιτεί την εκτίμηση και διαγωνοποίηση του πίνακα αυτοσυσχέτισης, ο οποίος με δεδομένη τη μεγάλη διάσταση του χώρου των εικόνων και το σχετικά μικρό πλήθος των δειγμάτων εκπαίδευσης, είναι σχεδόν πάντοτε singular.

Στην παρούσα εργασία προτείνουμε μια μέθοδο προσέγγισης του μετασχηματισμού KL με βάση τα διανύσματα –στήλες των εικόνων (VKLT). Με τον τρόπο αυτό απαλείφεται η μεγάλη απαίτηση σε μνήμη και η μη ομαλότητα (singularity) του πίνακα συμμεταβλητότητας. Η αποδοτικότητα του μετασχηματισμού VKLT όσον αφορά τη συσσώρευση ενέργειας στους βασικούς άξονες προσεγγίζει την αντίστοιχη του KLT. Επιπλέον είναι ένας μετασχηματισμός ο οποίος εξαρτάται –όπως και ο KLT- από τα δεδομένα ιδιότητα πολύ σημαντική για συστήματα ανίχνευσης προσώπων όπου απαιτείται η απόφαση κατά πόσο ένα πρότυπο είναι πρόσωπο ή όχι. Το βασικό πλεονέκτημα του VKLT σε σχέση με τον KLT είναι η μεγαλύτερη αναισθησία σε εντός κλάσης διακυμάνσεις γεγονός που αυξάνει την ευστάθεια του συστήματος αναγνώρισης.

2. Η ΠΡΟΣΕΓΓΙΣΗ ΤΟΥ KLT ΜΕ ΔΙΑΝΥΣΜΑΤΑ

Η μέθοδος των eigenfaces χρησιμοποιεί τον προβολή στο χώρο του μετασχηματισμού KL για τη δημιουργία της συμπαγούς αναπαράστασης των προσώπων μέσω διανυσμάτων χαρακτηριστικών [4][6].

Στην τρέχον Κεφάλαιο προσεγγίζουμε τον KLT από μια διαφορετική σκοπιά. Θεωρούμε ότι κάθε μια από τις N εικόνες του συνόλου εκπαίδευσης $\{I_1, I_2, \dots, I_N\}$ με $I_i \in \mathbb{R}^{n \times m}$ αποτελείται από διανύσματα γραμμών και στηλών. Υπολογίζουμε τους δύο πίνακες συμμεταβλητότητας, τον πίνακα συμμεταβλητότητας των γραμμών:

$$R_o = \sum_{j=1}^N (I_j - M) \cdot (I_j - M)^T \quad (\Gamma 2.1)$$

και των στηλών:

$$C_o = \sum_{j=1}^N (I_j - M)^T \cdot (I_j - M) \quad (\Gamma 2.2)$$

όπου $M = \sum_{j=1}^N I_j$ είναι η μέση εικόνα προσώπου. Πρέπει να τονιστεί ότι $R_o \in \mathbb{R}^{n \times n}$, $C_o \in \mathbb{R}^{m \times m}$ και επομένως οι διαστάσεις τους είναι πολύ μικρότερες από τις αντίστοιχες του C –βλέπε σχέση (Γ1.16) στο Κεφάλαιο Γ1. Επιλύοντας τις εξισώσεις:

$$\Lambda_1 = \Phi_1^T R_o \Phi_1 \quad (\Gamma 2.3)$$

$$\Lambda_2 = \Phi_2^T C_o \Phi_2 \quad (\Gamma 2.4)$$

λαμβάνουμε τους πίνακες Φ_1 και Φ_2 οι οποίοι καθορίζουν μετασχηματισμούς συντεταγμένων οι οποίοι αποσυσχετίζουν τις γραμμές και τις στήλες, των εικόνων του συνόλου μάθησης, αντίστοιχα. Σημειώνεται ότι η αποσυσχέτιση των στηλών έχει μεταξύ άλλων την προφανή λογική της εκμετάλλευσης της συμμετρίας που εμφανίζουν τα πρόσωπα.

Διατηρώντας τα g ιδιοδιανύσματα των Φ_1 και Φ_2 που αντιστοιχούν στις μεγαλύτερες ιδιοτιμές των πινάκων Λ_1 και Λ_2 λαμβάνουμε τον πίνακα χαρακτηριστικών Y_i για κάθε εικόνα:

$$Y_i = \Phi_{1q}^T \cdot \hat{I}_i \cdot \Phi_{2q} \quad (\Gamma 2.5)$$

όπου $\hat{I}_i = I_i - M$

Στον VKLT, ο KLT προσεγγίζεται υπό την έννοια της αποσυσχέτισης των δεδομένων της εικόνας σε δύο στάδια: πρώτα αποσυσχετίζουμε τις γραμμές και μετά τις στήλες.

3. ΔΙΑΔΙΚΑΣΙΑ ΕΠΙΛΗΘΕΥΣΗΣ

Εφαρμόσαμε διάφορους ελέγχους για να αξιολογήσουμε την αποδοτικότητα του VKLT. Θεωρούμε ένα δεδομένο σύνολο από εικόνες μάθησης $S = \{I_1, I_2, \dots, I_N\}$ των οποίων οι αντίστοιχες λεξικογραφικές εκδοχές είναι $V = \{v_1, v_2, \dots, v_N\}$, και τους πίνακες Φ_k , Φ_{1q} και Φ_{2l} οι οποίοι αντιστοιχούν στα τμήματα εκείνα των πινάκων Φ -βλέπε σχέση (Γ1.17), Φ_1 και Φ_2 με τα k, q και l βασικότερα ιδιοδιανύσματα αντίστοιχα (ιδιοδιανύσματα που αντιστοιχούν στις μεγαλύτερες ιδιοτιμές).

Μέσο Σφάλμα Ανακατασκευής στο Σύνολο Μάθησης

Σε αυτή τη διαδικασία συγκρίνουμε το Μέσο Σφάλμα Ανακατασκευής στο σύνολο μάθησης για τους μετασχηματισμούς KLT και VKLT, για διάφορες τιμές των k, q και l . Για να έχουμε μια τίμια σύγκριση διατηρούμε το ίδιο μέγεθος αναπαράστασης των εικόνων του συνόλου μάθησης, συγκεκριμένα πραγματοποιούμε συγκρίσεις μόνο για τις περιπτώσεις $k = l \cdot q$.

Ορίζουμε τα παρακάτω σφάλματα ανακατασκευής:

$$e_{KLT}(k, j) = (\hat{v}_j - \tilde{v}_j)^T \cdot (\hat{v}_j - \tilde{v}_j) \quad (\Gamma 2.6)$$

όπου $\tilde{v}_j = \Phi_k \cdot y_j$, $y_j = \Phi_k^T \cdot \hat{v}_j$ και

$$e_{VKLT}(q, l, j) = \text{vec}(\hat{I}_j - \tilde{I}_j)^T \cdot \text{vec}(\hat{I}_j - \tilde{I}_j) \quad (\Gamma 2.7)$$

όπου $\text{vec}(I)$ είναι ένα διάνυσμα που αντιστοιχεί στη λεξικογραφική διάταξη της εικόνας I , $Y_i = \Phi_{1q}^T \cdot \hat{I}_i \cdot \Phi_{2l}$, και $\tilde{I}_i = \Phi_{1q} \cdot Y_i \cdot \Phi_{2l}^T$.

Με βάση τις εξισώσεις (Γ2.6) και (Γ2.7) ορίζουμε το μέσο σφάλμα ανακατασκευής για τους μετασχηματισμούς KLT και VKLT αντίστοιχα ως:

$$me_{KLT}(k) = \frac{1}{N} \sum_{j=1}^N e_{KLT}(k, j) \quad (\Gamma 2.8)$$

$$me_{VKLT}(q, l) = \frac{1}{N} \sum_{j=1}^N e_{VKLT}(q, l, j) \quad (\Gamma 2.9)$$

Διακρίμανση Εντός Κλάσης

Σε συστήματα αναγνώρισης προσώπων, διακυμάνσεις στις συνθήκες φωτισμού, στη μετατόπιση και τις εκφράσεις των προσώπων πρέπει να επηρεάζουν όσο το δυνατό λιγότερο τη διαδικασία της ταυτοποίησης. Με απλά λόγια οι εντός κλάσης διακυμάνσεις απαιτείται να ελαχιστοποιούνται στο χώρο αναπαράστασης των εικόνων –δηλαδή στο χώρο των χαρακτηριστικών. Στη παράγραφο αυτή πραγματοποιείται σύγκριση των μετασχηματισμών KLT και VKLT με βάση τις αποστάσεις ανάμεσα στην αναπαράσταση μιας εικόνας στο χώρο χαρακτηριστικών και τις αντίστοιχες των μετατοπισμένων, θορυβωδών και υπό διαφορετικές συνθήκες φωτισμού εκδοχών της.

Έστω η εικόνα I_i και το αντίστοιχο διάνυσμα v_i . Ορίζουμε την εικόνα T_i που αντιστοιχεί στη μετατοπισμένη εκδοχή της I_i και την εικόνα L_i που αντιστοιχεί στην εικόνα I_i υπό διαφορετικές συνθήκες φωτισμού. Ορίζουμε επίσης τα αντίστοιχα διανύσματα t_i και l_i . Οι αποστάσεις στο χώρο του KLT, που δίνονται από τις σχέσεις:

$$te_{KLT}(k) = \left\| \Phi_k^T \cdot \hat{v}_i - \Phi_k^T \cdot t_i \right\| \quad (Γ2.10)$$

$$le_{KLT}(k) = \left\| \Phi_k^T \cdot \hat{v}_i - \Phi_k^T \cdot l_i \right\| \quad (Γ2.11)$$

ανάμεσα στα διανύσματα v_i και t_i ή l_i πρέπει να είναι η μικρότερη δυνατή. Το ίδιο ισχύει και για τις αναπαράστασεις των I_i και T_i ή L_i στο χώρο του VKLT:

$$te_{VKLT}(q, l) = \left\| \text{vec}(\Phi_{1q}^T \cdot \hat{I}_i \cdot \Phi_{2l}) - \text{vec}(\Phi_{1q}^T \cdot T_i \cdot \Phi_{2l}) \right\| \quad (Γ2.12)$$

$$le_{VKLT}(q, l) = \left\| \text{vec}(\Phi_{1q}^T \cdot \hat{I}_i \cdot \Phi_{2l}) - \text{vec}(\Phi_{1q}^T \cdot L_i \cdot \Phi_{2l}) \right\| \quad (Γ2.13)$$

Οι μέσες αποστάσεις ανάμεσα στις ανωτέρω αναπαράστασεις σε όλη τη βάση από πρόσωπα υπολογίζεται με χρήση των εξισώσεων (Γ2.10) έως (Γ2.13) και ορισμών αντιστοιχών με αυτούς των εξισώσεων (Γ2.8) και (Γ2.9).

Ποσοστά Αναγνώρισης

Ο πιο σημαντικός έλεγχος για τους αλγορίθμους αναγνώρισης προσώπων είναι η αποδοτικότητα όσον αφορά την ορθή ταυτοποίηση. Η διαδικασία συνοψίζεται ως εξής: Βρες σε πιο αποθηκευμένο πρόσωπο στη βάση αντιστοιχεί η εικόνα η οποία δίνεται στο σύστημα για σκοπούς ταυτοποίησης. Στη γενική περίπτωση η προς ταυτοποίηση εικόνα διαφέρει από την αντίστοιχη αποθηκευμένη σε ένα η περισσότερους από τους παρακάτω παράγοντες: χρόνο λήψης, κλίμακα προσώπου, οριοθέτηση προσώπου, συνθήκες φωτισμού, εκφράσεις προσώπου κ.ο.κ. Το κριτήριο ταυτοποίησης είναι συνήθως η αρχή του πλησιέστερου γείτονα και οι συγκρίσεις πραγματοποιούνται στο χώρο αναπαράστασης. Πρέπει επίσης να σημειωθεί ότι προσοχή χρειάζεται στην διερεύνηση του κατά πόσο η προς ταυτοποίηση εικόνα απεικονίζει πράγματι πρόσωπο. Για τον έλεγχο αυτό είναι ιδιαίτερα σημαντικοί οι μετασχηματισμοί οι οποίοι εξαρτώνται από τα δεδομένα (απαιτούν κάποιο σύνολο μάθησης).

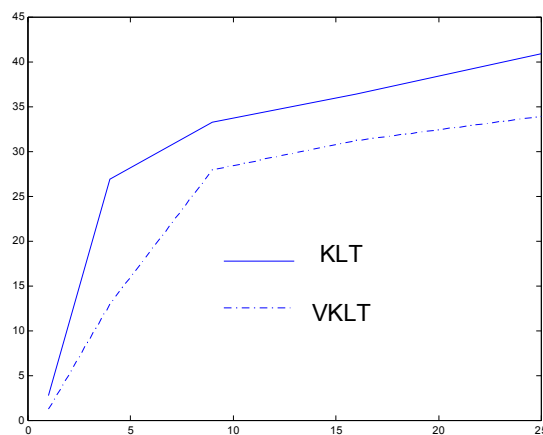
Στις προσομοιώσεις που ακολουθούν τα ποσοστά αναγνώρισης υπολογίζονται για διάφορα μεγέθη του διανύσματος αναπαράστασης στους χώρους του KLT και VKLT και διατηρώντας πάντοτε $k = l \cdot q$.

4. ΠΕΙΡΑΜΑΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ

Τα πειραματικά αποτελέσματα για όλες τις προσομοιώσεις ελήφθησαν χρησιμοποιώντας τη βάση προσώπων ORL. Σε όλες τις περιπτώσεις οι εικόνες προσώπων μετασχηματίστηκαν σε «φορμάτ προσώπου». Εξαιτίας περιορισμών στην απαιτούμενη μνήμη και στους υπολογισμούς για την υλοποίηση του KLT περιοριστήκαμε σε χωρική ανάλυση 64x72.



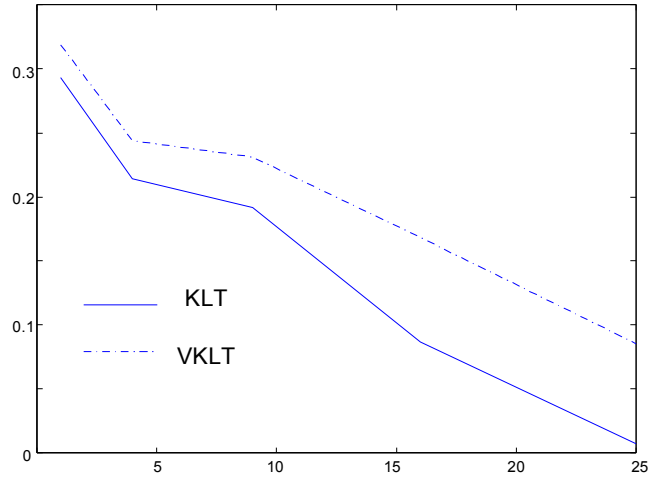
Εικόνα Γ2.1: (α) Αρχική εικόνα σε «φορμάτ προσώπου» (β) Υπό διαφορετικές συνθήκες φωτισμού (γ) Μετατοπισμένη (δ) Θορυβώδης



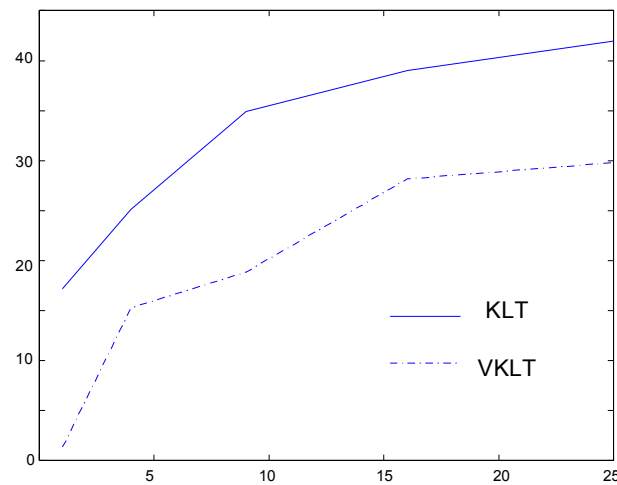
Σχήμα Γ2.1: Μέσες αποστάσεις στους χώρους του KLT και VKLT ανάμεσα στις αρχικές εικόνες και στις τυχαία μετατοπισμένες εκδοχές τους σε σχέση με το μέγεθος του διανύσματος αναπαράστασης.

Τα ποσοστά επιτυχούς ταυτοποίησης επιδεικνύονται στον Πίνακα Γ2.1 για αμφότερους τους μετασχηματισμούς KL και VKL. Η υπεροχή του VKLT είναι εμφανής σε όλες τις περιπτώσεις, σε σχέση με το μέγεθος του διανύσματος αναπαράστασης. Επιπλέον όπως επιδεικνύεται στα Σχήματα Γ2.3, Γ2.4 και Γ2.5, ο VKLT υπερέρχει του KLT, και σε όλους τους υπόλοιπους ελέγχους με εξαίρεση το μέσο σφάλμα ανακατασκευής στο οποίο KLT είναι εξ' ορισμού βέλτιστος. Συγκεκριμένα ο VKLT παρουσιάζει

μικρότερο μέσο σφάλμα μετατόπισης και φωτισμού και επομένως είναι περισσότερο ευσταθής σε εντός κλάσης διακυμάνσεις.



Σχήμα Γ2.2: Μέσο σφάλμα ανακατασκευής των μετασχηματισμών KL και VKL σε σχέση με το μέγεθος του διανύσματος αναπαράστασης.



Σχήμα Γ2.3: Μέσες αποστάσεις στους χώρους του KLT και VKLT ανάμεσα στις αρχικές εικόνες και στις αντίστοιχες υπό διαφορετικές συνθήκες φωτισμού σε σχέση με το μέγεθος του διανύσματος αναπαράστασης.

Συντελεστές Αναπαράστασης	16	25	36	49
Ποσοστά Αναγνώρισης (%)				
KLT	54	63	75	82
VKLT	62	69	81	85

Πίνακας Γ2.1: Ποσοστά αναγνώρισης του KLT και VKLT

ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΣΩΠΩΝ ΣΤΟ ΧΩΡΟ ΤΟΥ ΜΕΤΑΣΧΗΜΑΤΙΣΜΟΥ DCT

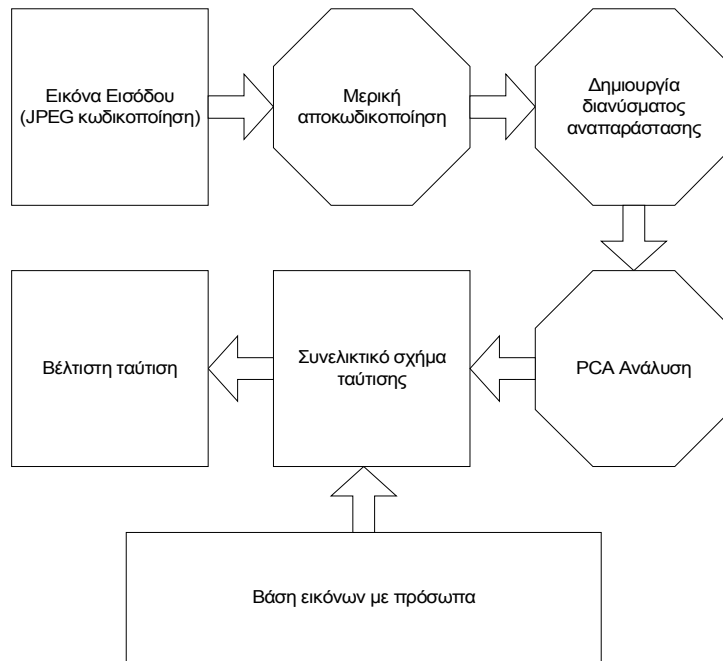
Στο κεφάλαιο αυτό διερευνάται η χρήση των συντελεστών του μετασχηματισμού DCT όπως αυτός εφαρμόζεται στο πρότυπο JPEG –εφαρμογή σε μπλοκ διάστασης 8×8- για την αναγνώριση προσώπων. Ταυτοποίηση με βάση τους συντελεστές DCT απαιτεί μερική μόνο αποκωδικοποίηση των εικόνων και είναι εξαιρετικά χρήσιμη για εφαρμογές δεικτοδότησης σε γενικού σκοπού βάσεις εικόνων. Η αντίθεση με τους υπάρχοντες αλγορίθμους είναι ότι δεν απαιτεί επεξεργασία όλων των εικόνων της βάσης και τη δημιουργία κάποιου επιπλέον διανύσματος αναπαράστασης για σκοπούς δεικτοδότησης.

1. ΕΙΣΑΓΩΓΗ

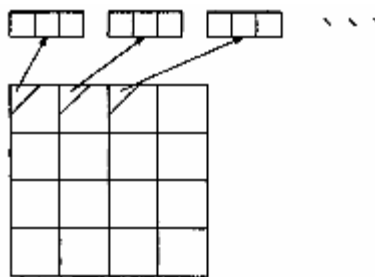
Στόχος του Κεφαλαίου αυτού είναι να διερευνήσει την καταλληλότητα της χρήσης των συντελεστών του μετασχηματισμού DCT, όπως εφαρμόζεται στο πρότυπο JPEG, για την περιγραφή προσώπων. Υπό αυτό το πρίσμα δεν παρουσιάζεται ένα πλήρες σύστημα αναγνώρισης –παραλείπονται τα στάδια προεπεξεργασίας για την μετατροπή του προσώπου σε «φορμάτ προσώπου»- αλλά γίνεται η υπόθεση ότι τα πρόσωπα που αποτελούν τη βάση βρίσκονται σε κατάλληλη μορφή για εφαρμογή άμεσα σε αυτά της διαδικασίας ταυτοποίησης. Η τεχνική, ακόμα και χωρίς το στάδιο της προεπεξεργασίας, μπορεί να εφαρμοστεί σε mug shot φωτογραφίες, σε περιβάλλον επικοινωνίας ανθρώπου μηχανής και σε εφαρμογές δεικτοδότησης με βάση το περιεχόμενο σε γενικού σκοπού βάσεις από εικόνες.

2. ΧΟΝΔΡΙΚΗ ΠΕΡΙΓΡΑΦΗ ΤΗΣ ΠΡΟΤΕΙΝΟΜΕΝΗΣ ΤΕΧΝΙΚΗΣ

Το μπλοκ διάγραμμα του προτεινόμενου σχήματος επιδεικνύεται στο Σχήμα Γ3.1. Η εικόνα εισόδου οδηγείται σε ένα αποκωδικοποιητή JPEG για μερική αποκωδικοποίηση –εξαγωγή των συντελεστών των μπλοκ- και στη συνέχεια δημιουργείται το διάνυσμα αναπαράστασης. Για περαιτέρω μείωση της διάστασης του διανύσματος εφαρμόζεται PCA ανάλυση. Στη συνέχεια τα διανύσματα οδηγούνται στο σχήμα ταύτισης όπου συγκρίνονται με τις αντίστοιχες αναπαραστάσεις των εικόνων της βάσης. Για κάθε πρόσωπο της βάσης μπορεί να είναι προϋπολογισμένο το διάνυσμα αναπαράστασης –κλασσικές εφαρμογές αναγνώρισης προσώπων- είτε να υπολογίζεται κατά τη διάρκεια της εκτέλεσης της εφαρμογής –δεικτοδότηση σε γενικού σκοπού βάσεις δεδομένων από εικόνες. Στην έξοδο του σχήματος μπορεί να επιλεγεί η εικόνα με τη βέλτιστη ταύτιση ή ένα σύνολο από πιθανές εικόνες.



Σχήμα Γ3.1: Η προτεινόμενη τεχνική αναγνώρισης προσώπων.



Σχήμα Γ3.2: Δημιουργία του διανύσματος αναπαράστασης από τους συντελεστές DCT των μπλοκ της εικόνας

3. ΔΗΜΙΟΥΡΓΙΑ ΤΟΥ ΔΙΑΝΥΣΜΑΤΟΣ ΑΝΑΠΑΡΑΣΤΑΣΗΣ

Η δημιουργία του διανύσματος αναπαράστασης διευκρινίζεται στο Σχήμα Γ3.2. Από κάθε μπλοκ επιλέγονται με ζιγκ-ζαγκ σάρωση οι συντελεστές του μετασχηματισμού DCT με τη μεγαλύτερη ενέργεια, που αντιστοιχεί στη συντριπτική πλειονότητα των περιπτώσεων, αφορούν στις χαμηλές χωρικές συχνότητες. Το πρόβλημα είναι ότι αφενός δεν υπάρχει σαφής ένδειξη για το αν αυτοί οι συντελεστές έχουν την απαιτούμενη διαχωριστικότητα και αφετέρου το διάνυσμα αναπαράστασης είναι πολύ μεγάλο σε μέγεθος – η αναπαράσταση δεν είναι ιδιαίτερα συμπαγής. Όσον αφορά το πρώτο πρόβλημα, το γεγονός ότι οι συντελεστές λαμβάνονται σε επίπεδο μπλοκ δίνει την απαιτούμενη λεπτομέρεια στην περιγραφή έστω και αν κρατάμε απλά τους χαμηλοσυχνοτικούς συντελεστές. Το πρόβλημα της συμπαγούς αναπαράστασης –

διατηρώντας 5 συντελεστές για κάθε μπλοκ 8×8 σε μια εικόνα 512×512 δημιουργούμε διάνυσμα αναπαράστασης με 20480 στοιχεία- λύνεται με εφαρμογή PCA ανάλυσης. Εναλλακτικά θα μπορούσε να εφαρμοστεί ανάλυση FLD αλλά αυτό απαιτεί την ύπαρξη πολλών αναπαραστάσεων του ίδιου προσώπου οι οποίες δεν είναι πάντα διαθέσιμες. Περισσότερα για την ανάλυση PCA και FLD δίνονται στο Κεφάλαιο Γ1. Η ζιγκ-ζάγκ σάρωση είναι σύμφωνη με το τρόπο που γίνεται χειρισμός των συντελεστών DCT στο πρότυπο JPEG και συνάδει με την μερική αποκωδικοποίηση η οποία εφαρμόζεται στον προτεινόμενο αλγόριθμο.

4. ΤΟ ΣΤΑΔΙΟ ΤΑΥΤΟΠΟΙΗΣΗΣ

Το στάδιο ταυτοποίησης εφαρμόζεται κατευθείαν στο χώρο συμπίεσης και απαιτεί την ελάχιστη δυνατή αποκωδικοποίηση των εικόνων. Αποτέλεσμα της ιδιότητας αυτή είναι η δυνατότητα της τεχνικής να εκτελείται σε πραγματικό χρόνο και σε τυπικά υπολογιστικά συστήματα. Ο αλγόριθμος επικεντρώνεται σε επίπεδο μπλοκ μια και τα μπλοκ είναι η ελάχιστη μονάδα κωδικοποίησης του προτύπου JPEG.

Ως κριτήριο ταύτισης χρησιμοποιήθηκε μια προσαρμοσμένη μετρική βασισμένη στη συνέλιξη. Έστω x και y τα διανύσματα αναπαράστασης δύο προσώπων, $z = x \otimes y$ το διάνυσμα της συνέλιξης τους και $i = \max(z)$ η μέγιστη τιμή του διανύσματος αυτού. Η εύρεση του προσώπου u το οποίο πλησιάζει περισσότερο από κάθε άλλο το πρόσωπο εισόδου x επιλέγεται με βάση το επόμενο κριτήριο:

$$u = y_i, \quad i = \arg \max_k \left\{ \frac{\max(x \otimes y_k)}{\|x\| \cdot \|y_k\|} \right\}$$

Η χρήση της συνέλιξης επιβάλλεται για σκοπούς ευστάθειας, δεδομένου ότι η τεχνική ταυτοποίησης στηρίζεται σε DCT συντελεστές των μπλοκ. Για παράδειγμα μια μικρή αποτυχία στην ορθή οριοθέτηση του προσώπου μπορεί να μετακινήσει τα μη επικαλυπτόμενα μπλοκ και να οδηγήσει μετρικές όπως ο συντελεστής συσχέτισης σε πλήρη αποτυχία. Τα ποσοστά αναγνώρισης αυξάνονται με τη αύξηση του αριθμού των συντελεστών οι οποίοι χρησιμοποιούνται δημιουργώντας προφανώς μια επιβάρυνση στους υπολογισμούς –προφανώς σε κάθε περίπτωση μπορεί να ζητηθεί από το σχήμα της PCA ανάλυσης να διατηρεί σταθερό μήκος του διανύσματος αναπαράστασης.

5. ΠΕΙΡΑΜΑΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ

Οι προσομοιώσεις πραγματοποιήθηκαν με χρήση της βάσης προσώπων ORL η οποία περιέχει πρόσωπα σε μορφή mug shot φωτογραφιών και το στάδιο της προεπεξεργασίας μπορεί να παραληφθεί. Για σκοπούς σύγκρισης με την τεχνική των ιδιοπροσώπων (KL) αναγκαστήκαμε να μειώσουμε την χωρική διάσταση των εικόνων σε 64×72 . Με δεδομένο το μικρό αριθμό των μπλοκ που προκύπτει από τη χωρική ανάλυση αυτή, παραλείψαμε το στάδιο της PCA ανάλυσης των DCT συντελεστών. Με τον τρόπο αυτό διερευνήθηκε άμεσα η διαχωριστικότητα που προκύπτει από τη χρήση τους. Οι εντός κλάσης διακυμάνσεις που εμφανίζονται στα πρόσωπα της βάσης ORL αφορούν κυρίως μικρές αλλαγές στη κλίμακα, στο φωτισμό και την στροφή περί τους άξονες \mathbf{x} , \mathbf{y} και \mathbf{z} . Εκτός από τη σύγκριση με την τεχνική των eigenfaces,

πραγματοποιήθηκε επίσης σύγκριση με χρήση ως χαρακτηριστικών των συντελεστών του μετασχηματισμού DCT ολόκληρης της εικόνας.

Τα συγκριτικά αποτελέσματα επιδεικνύονται στον Πίνακα Γ3.1. Πρέπει να σημειώσουμε ότι το ελάχιστο μήκος του διανύσματος αναπαράστασης στην τεχνική BDCT (Block DCT) ήταν ίσο με 72 (όσο ο αριθμός των μη επικαλυπτόμενων 8x8 μπλοκ μιας εικόνας με χωρική ανάλυση 64x72) και επομένως η υπεροχή της τεχνικής σε σχέση με τις άλλες δύο είναι μάλλον παραπλανητική. Σε κάθε όμως περίπτωση η τεχνική, BDCT έχει το πλεονέκτημα της εφαρμογής στο χώρο συμπίεσης το οποίο την καθιστά άμεσα εκτελέσιμη σε εφαρμογές ανάκλησης με βάση το περιεχόμενο και πολυμέσων γενικότερα.

Ποσοστά Αναγνώρισης (%) (Αριθμός διατηρούμενων συντελεστών)				
KLT	44 (16)	54 (25)	65 (36)	77 (49)
DCT	52 (16)	59 (25)	73 (36)	84 (49)
BDCT	81 (72)	81 (144)	85 (216)	88 (288)

Πίνακας Γ3.1: Ποσοστά αναγνώρισης για τρία διαφορετικά σχήματα.

6. ΣΥΜΠΕΡΑΣΜΑΤΑ

Στο κεφάλαιο αυτό περιγράφηκε μια διαδικασία αναγνώρισης η οποία εκμεταλλεύεται την πληροφορία στο χώρο συμπίεσης χρησιμοποιώντας τους συντελεστές του μετασχηματισμού DCT. Από τα πειραματικά αποτελέσματα προκύπτει ότι η τεχνική μπορεί να εφαρμοστεί για αναγνώριση προσώπων από mug shot φωτογραφίες, σε περιβάλλον επικοινωνίας ανθρώπου μηχανής και δεικτοδότησης με βάση το περιεχόμενο σε βάσεις εικόνων. Το σημαντικότερο μειονέκτημα της τεχνικής είναι το μεγάλο μήκος του διανύσματος αναπαράστασης που προκύπτει. Το πρόβλημα αυτό μπορεί να λυθεί με εφαρμογή τεχνικών όπως η PCA και η FLD σύμφωνα με την περιγραφή του Κεφαλαίου Γ1.

ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΣΩΠΩΝ ΑΝΕΞΑΡΤΗΤΑ ΑΠΟ ΤΙΣ ΔΙΑΚΥΜΑΝΣΕΙΣ ΦΩΤΙΣΜΟΥ

Στο κεφάλαιο αυτό προτείνεται ένα σύστημα ταυτοποίησης προσώπων το οποίο διατηρεί την ουσιαστική πληροφορία που παρέχει το πρόσωπο ανεξάρτητα από τις συνθήκες φωτισμού. Η πληροφορία αυτή αποθηκεύεται στη βάση υπό τη μορφή ιδιζόντων διανυσμάτων (singular vectors) ενώ οι συνθήκες φωτισμού προσαρμόζονται με βάση την προς ταυτοποίηση εικόνα. Με απλά λόγια το σύστημα εφαρμόζει την ερώτηση: «Πως θα ήταν τα αποθηκευμένα στη βάση πρόσωπα υπό τις συνθήκες φωτισμού της συγκεκριμένης προς ταυτοποίηση φωτογραφίας», πριν τη διαδικασία ταυτοποίησης. Οι συνθήκες φωτισμού της υπό ταυτοποίηση φωτογραφίας προσεγγίζονται από τις ιδιάζουσες τιμές της (singular values). Επειδή η παραπάνω διαδικασία δεν είναι ιδιαίτερα γρήγορη για να εφαρμοστεί σε πραγματικό χρόνο, προηγείται μια διαδικασία δεικτοδότησης με βάση τους συντελεστές του μετασχηματισμού DCT οι οποίοι έχουν υπολογιστεί με μερική αποκωδικοποίηση των JPEG εικόνων –βλέπε Κεφάλαιο Γ3. Τα πειραματικά αποτελέσματα επιδεικνύουν ότι η προτεινόμενη τεχνική όχι μόνο επιτυγχάνει υψηλότερα ποσοστά αναγνώρισης αλλά αυξάνει και την ευστάθεια της διαδικασίας ταυτοποίησης.

1. ΕΙΣΑΓΩΓΗ

Το προτεινόμενο σχήμα αναγνώρισης προσπαθεί να υπερκεράσει τις διακυμάνσεις στις συνθήκες φωτισμού ανάμεσα στο προς ταυτοποίηση πρόσωπο και στην αποθηκευμένη στη βάση εκδοχή του. Η τεχνική στηρίζεται στην αποσύνθεση ιδιζουσών τιμών τόσο της προς ταυτοποίηση όσο και των αποθηκευμένων στη βάση εικόνων. Η πληροφορία υψής λαμβάνεται από τις ιδιάζουσες τιμές ενώ το σχήμα τους και το περιεχόμενο υψηλών συχνοτήτων εξάγεται από τα ιδιάζοντα διανύσματα. Στη συνέχεια πραγματοποιείται ανακατασκευή των αποθηκευμένων στη βάση εικόνων, με χρήση των ιδιζουσών τιμών της προς ταυτοποίηση εικόνας και τα ιδιάζοντα διανύσματα των αποθηκευμένων εικόνων. Με τον τρόπο αυτό, το περιεχόμενο σχήματος και υψηλών συχνοτήτων των προσώπων της βάσης προσαρμόζεται στις συνθήκες φωτισμού και την υφή της προς ταυτοποίηση φωτογραφίας -δημιουργία των “face versions”. Η ταυτοποίηση πραγματοποιείται ανάμεσα στις τροποποιημένες φωτογραφίες της βάσης και στην υπό έλεγχο φωτογραφία. Με τον τρόπο αυτό το σφάλμα το οποίο οφείλεται στις συνθήκες φωτισμού μειώνεται δίνοντας υψηλότερα ποσοστά αναγνώρισης. Η υπολογιστική πολυπλοκότητα του αλγορίθμου περιορίζεται με δεικτοδότηση των πιθανών προσώπων με βάση τους συντελεστές του μετασχηματισμού DCT των μπλοκ της εικόνας [7].

2. ΑΝΑΛΥΣΗ ΙΔΙΑΖΟΥΣΩΝ ΤΙΜΩΝ

Κάθε εικόνα μπορεί να θεωρηθεί ως ένας ορθογώνιος πίνακας. Είναι επίσης γνωστό ότι κάθε πραγματικός και συμμετρικός πίνακας μπορεί να μετασχηματιστεί σε ένα διαγώνιο πίνακα είτε μέσω ενός ορθογώνιου μετασχηματισμού είτε μέσω της ανάλυσης ιδιζουσών τιμών (Singular Value Decomposition -SVD), ιδιότητα που ισχύει για κάθε ορθογώνιο πίνακα A με διαστάσεις $m \times n$. Στη συνέχεια αναφερόμαστε σε

μερικές βασικές ιδιότητες της ανάλυσης ιδιζουσών τιμών οι οποίες μας βοηθούν στη κατανόηση της προτεινόμενης τεχνικής.

Έστω \mathbf{A} ένας πραγματικός ορθογώνιος πίνακας με διαστάσεις $m \times n$ (υποθέτουμε $m > n$, χωρίς απώλεια της γενικότητας) και $\text{rank}(\mathbf{A}) = k$. Τότε υπάρχουν δύο ορθομοναδιαίοι πίνακες $\mathbf{U}_{m \times m}$, $\mathbf{V}_{n \times n}$ και ένας διαγώνιος πίνακας $\mathbf{\Sigma}_{m \times n}$ οι οποίοι ικανοποιούν την εξίσωση:

$$\mathbf{A} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^T \quad (\Gamma 4.1)$$

όπου $\mathbf{\Sigma} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_k, 0, \dots, 0)$ and $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k$.

Κάθε λ_i^2 , είναι μια από τις ιδιοτιμές των πινάκων $\mathbf{A}^T \mathbf{A}$ και $\mathbf{A} \mathbf{A}^T$, και οι λ_i καλούνται ιδιζουσες τιμές του πίνακα \mathbf{A} . Οι $\mathbf{U} = (\mathbf{u}_1, \dots, \mathbf{u}_k, \mathbf{u}_{k+1}, \dots, \mathbf{u}_m)$ και $\mathbf{V} = (\mathbf{v}_1, \dots, \mathbf{v}_k, \mathbf{v}_{k+1}, \dots, \mathbf{v}_n)$ είναι οι πίνακες με στήλες τα ιδιζόντα διανύσματα \mathbf{u}_i και \mathbf{v}_i που αντιστοιχούν στις ιδιοτιμές λ_i^2 των πινάκων $\mathbf{A}^T \mathbf{A}$, $\mathbf{A} \mathbf{A}^T$ αντίστοιχα.

Ορίζουμε το διάνυσμα των ιδιζουσών τιμών $\mathbf{x}_{n \times 1}$ της εικόνας \mathbf{A} ως:

$$\mathbf{x}_{n \times 1} = [\lambda_1 \lambda_2 \dots \lambda_k \ 0 \dots 0]^T \quad (\Gamma 4.2)$$

Για κάθε πραγματικό ορθογώνιο πίνακα \mathbf{A} και υπό τον περιορισμό $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k$, το διάνυσμα $\mathbf{x}_{n \times 1}$ είναι μοναδικό. Η μοναδικότητα και άλλες σημαντικές ιδιότητες όπως η αναισθησία στη περιστροφή και στη κλίμακα, του διανύσματος $\mathbf{x}_{n \times 1}$ το έχουν καταστήσει κατάλληλη επιλογή για την περιγραφή εικόνων από πρόσωπα [5]. Επιπλέον το διάνυσμα ιδιζουσών τιμών αναπαριστά εσωτερικές και όχι οπτικά παρατηρήσιμες ιδιότητες της εικόνας. Παρόλα αυτά, όπως φαίνεται στην επόμενη παράγραφο, το πρόβλημα των μεταβολών στο φωτισμό δεν μπορεί να επιλυθεί με το διάνυσμα ιδιζουσών τιμών, το οποίο παρουσιάζει αναισθησία μόνο ως προς αναλογικές μεταβολές στη φωτεινότητα.

2.1 Αναισθησία του διανύσματος ιδιζουσών τιμών σε αναλογικές μεταβολές στη φωτεινότητα

Έστω ότι οι διαφορετικές αποχρώσεις του γκρι σε μια εικόνα αναπαρίστανται από τον πίνακα \mathbf{A} . Μία αναλογική μεταβολή στη φωτεινότητα ισοδυναμεί με τον πολλαπλασιασμό των τιμών του πίνακα \mathbf{A} με μια μη μηδενική σταθερά α . Επομένως η προκύπτουσα εικόνα αναπαρίσταται από τον πίνακα $\alpha \mathbf{A}$.

Έστω ότι οι ιδιζουσες τιμές των πινάκων \mathbf{A} και $\alpha \mathbf{A}$ είναι $\lambda_1, \lambda_2, \dots, \lambda_k$ και $\lambda_1^*, \lambda_2^*, \dots, \lambda_k^*$ αντίστοιχα. Σύμφωνα με την εξίσωση ιδιοτιμών του πίνακα $(\alpha \mathbf{A})(\alpha \mathbf{A})^T$ έχουμε:

$$\left| (\alpha \mathbf{A})(\alpha \mathbf{A})^T - \lambda^{*2} I \right| = 0 \quad (\Gamma 4.3)$$

το οποίο είναι ισοδύναμο με $\left| \mathbf{A} \mathbf{A}^T - \left(\frac{1}{\alpha^2} \lambda^{*2} I \right) \right| = 0$

Η σχέση ανάμεσα στα διανύσματα ιδιζουσών τιμών των πινάκων $\alpha \mathbf{A}$ και \mathbf{A} είναι:

$$\left(\lambda_1^*, \lambda_2^*, \dots, \lambda_k^*, 0, \dots, 0 \right)^T = |\alpha| \left(\lambda_1, \lambda_2, \dots, \lambda_k, 0, \dots, 0 \right)^T \quad (\Gamma 4.4)$$

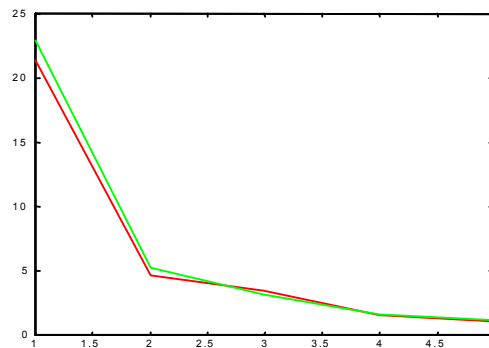
Επομένως το διάνυσμα ιδιζουσών τιμών είναι πρακτικά αναισθητο σε αναλογικές μεταβολές των τιμών φωτεινότητας της εικόνας ιδιότητα που είναι χρήσιμη για την αναγνώριση και περιγραφή σκηνών. Παρόλα αυτά οι μεταβολές στις συνθήκες φωτισμού σε καμία περίπτωση δεν μπορούν να θεωρηθούν ως αναλογική μεταβολή στις τιμές της εικόνας λόγω της κατευθυντικότητας της πηγής φωτισμού. Στην Εικόνα Γ4.1 επιδεικνύεται το ίδιο πρόσωπο υπό διαφορετικές συνθήκες φωτισμού. Συγκρίνοντας τις δύο εικόνες παρατηρούμε μια πλήρη ανακατανομή του γενικού φωτισμού –δεδομένου ότι η ανακλαστικότητα του προσώπου παραμένει η ίδια. Στην Εικόνα Γ4.1(γ) παρατηρούμε τα διαγράμματα των ιδιζουσών τιμών στις ανωτέρω περιπτώσεις. Τα σημεία τομής των διαγραμμάτων δεν επιβεβαιώνουν την αναλογική μεταβολή των τιμών της φωτεινότητας εξαιτίας των συνθηκών φωτισμού. Πρακτικά οι μεταβολές στις συνθήκες φωτισμού δημιουργούν μια αναδιανομή των τιμών φωτεινότητας που επηρεάζει κατά βάση την υφή της εικόνας.



(α)



(β)



(γ)

Εικόνα Γ4.1: Πρόσωπα κάτω από διαφορετικές συνθήκες φωτισμού και τα αντίστοιχα διανύσματα ιδιζουσών τιμών.

Στη παρούσα εργασία ισχυριζόμαστε ότι τα ιδιάζοντα διανύσματα διατηρούν την πληροφορία περιεχομένου της εικόνας δεδομένου ότι αποτελούν τους βασικούς άξονες της. Για την υποστήριξη του ισχυρισμού αυτού ανακατασκευάσαμε πολλές εικόνες διατηρώντας τους πίνακες των ιδιάζοντων τιμών και αντικαθιστώντας τις ιδιάζουσες τιμές με τυχαίες τιμές. Για σκοπούς επίδειξης θεωρούμε τρεις διαφορετικές προσεγγίσεις της κατανομής των ιδιάζουσών τιμών.

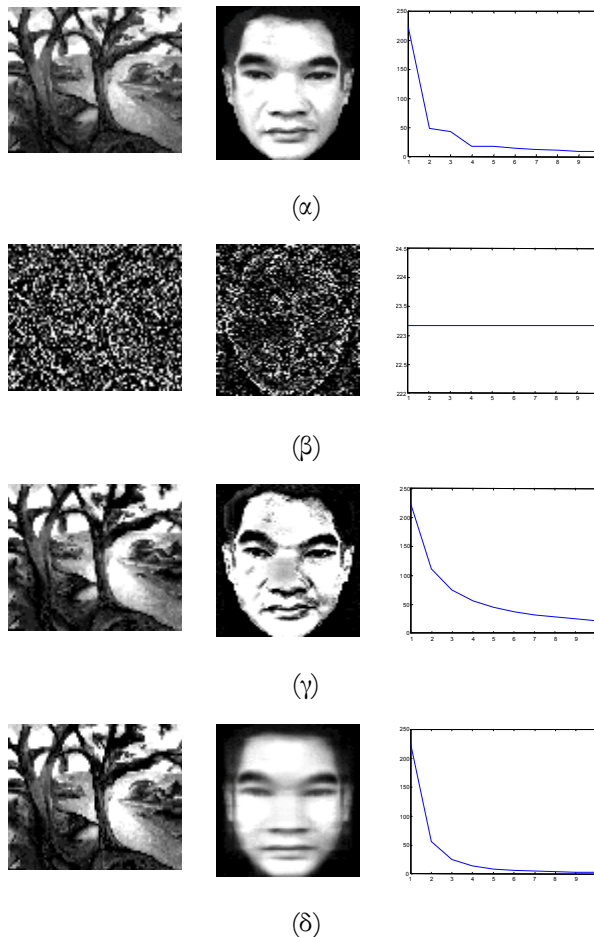
Έστω το επεκτεταμένο διάνυσμα ιδιζουσών τιμών $\mathbf{x}_{nx1} = [\lambda_1 \lambda_2 \dots \lambda_k \ 0 \dots 0]^T$. Θεωρούμε τα παρακάτω διανύσματα:

$$\mathbf{a}_{nx1} = \{\lambda_i = \lambda_1, i = 1..k \mid \lambda_i = 0, i = k + 1..n\} \quad (\Gamma 4.5)$$

$$\mathbf{b}_{nx1} = \{\lambda_i = \frac{\lambda_1}{i}, i = 1..k \mid \lambda_i = 0, i = k + 1..n\} \quad (\Gamma 4.6)$$

$$\mathbf{c}_{nx1} = \{\lambda_i = \frac{\lambda_1}{i^2}, i = 1..k \mid \lambda_i = 0, i = k + 1..n\} \quad (\Gamma 4.7)$$

Στην Εικόνα Γ4.2 οι ανακατασκευασμένες εικόνες και πρόσωπα αναδεικνύουν το γεγονός ότι τα ιδιάζοντα διανύσματα διατηρούν το ουσιαστικό περιεχόμενο των εικόνων ενώ οι ιδιάζουσες τιμές διατηρούν την πληροφορία υψής. Σε όλες τις περιπτώσεις οι κατανομές των ιδιζουσών τιμών έχουν πολλαπλασιαστεί με την μεγαλύτερη ιδιάζουσα τιμή της αρχικής εικόνας, όπως περιγράφεται στις εξισώσεις (Γ4.5)-(Γ4.7), ώστε να διατηρηθεί η μέση φωτεινότητα της εικόνας.



Εικόνα Γ4.2: (α) Η αρχική εικόνα και το αντίστοιχο διάνυσμα ιδιζουσών τιμών. (β),(γ),(δ) Ανακατασκευή της αρχικής εικόνας χρησιμοποιώντας τους πίνακες ιδιάζόντων διανυσμάτων και τις ιδιάζουσες τιμές που περιγράφονται από τις εξισώσεις (Γ4.5),(Γ4.6) και (Γ4.7) αντίστοιχα.

3. ΠΡΟΣΑΡΜΟΓΗ ΤΩΝ ΕΙΚΟΝΩΝ ΠΡΟΣΩΠΟΥ ΣΕ ΑΛΛΕΣ ΣΥΝΘΗΚΕΣ ΦΩΤΙΣΜΟΥ

Έστω οι ορθογώνιοι πίνακες \mathbf{I} και \mathbf{S} οι οποίοι αναπαριστούν την εικόνα εισόδου, δηλαδή το προς ταυτοποίηση πρόσωπο, και μια αποθηκευμένη στη βάση εικόνα αντίστοιχα. Για να συγκρίνουμε τις εικόνες ανεξάρτητα από τις μεταβολές στο φωτισμό εφαρμόζουμε την αποσύνθεση ιδιαζουσών τιμών στις εικόνες:

$$\mathbf{I} = \mathbf{U}_I \mathbf{\Sigma}_I \mathbf{V}_I^T = \sum_{j=1}^k \lambda_{Ij} \mathbf{u}_{Ij} \mathbf{v}_{Ij}^T \quad (\Gamma 4.8)$$

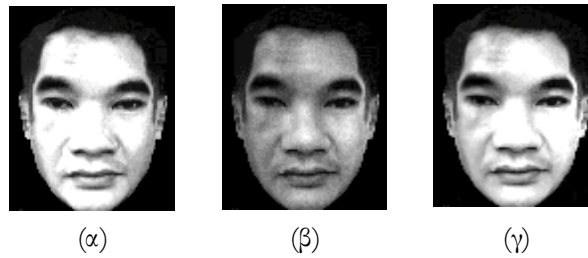
$$\mathbf{S} = \mathbf{U}_S \mathbf{\Sigma}_S \mathbf{V}_S^T = \sum_{j=1}^k \lambda_{Sj} \mathbf{u}_{Sj} \mathbf{v}_{Sj}^T \quad (\Gamma 4.9)$$

Με βάση την προτεινόμενη τεχνική αντί να συγκρίνουμε άμεσα τους πίνακες \mathbf{I} και \mathbf{S} δημιουργούμε μια προσέγγιση του πίνακα $\hat{\mathbf{S}}$, την οποία ονομάζουμε “face version” του \mathbf{S} , σύμφωνα με την εξίσωση (Γ4.10):

$$\hat{\mathbf{S}} = \mathbf{U}_S \mathbf{\Sigma}_I \mathbf{V}_S^T = \sum_{j=1}^k \lambda_{Ij} \mathbf{u}_{Sj} \mathbf{v}_{Sj}^T \quad (\Gamma 4.10)$$

Όπως φαίνεται από τις εξισώσεις (Γ4.8), (Γ4.9), (Γ4.10) για τη δημιουργία του $\hat{\mathbf{S}}$ χρησιμοποιούμε τους πίνακες ιδιαζόντων διανυσμάτων $\mathbf{U}_S, \mathbf{V}_S$ σε συνδυασμό με τον πίνακα ιδιαζουσών τιμών $\mathbf{\Sigma}_I$ της εικόνας εισόδου \mathbf{I} . Η επιτυχημένη προσαρμογή των συνθηκών φωτισμού της εικόνας \mathbf{S} στην εικόνα \mathbf{I} μέσω της ανωτέρω διαδικασίας επιδεικνύεται στην Εικόνα Γ4.3.

Το επόμενο βήμα είναι η σύγκριση των ποσοστών αναγνώρισης όταν στη διαδικασία ταυτοποίησης αντί της εικόνας \mathbf{S} θεωρούμε την “face version” $\hat{\mathbf{S}}$.



Εικόνα Γ4.3: (α) Πρόσωπο αποθηκευμένο στη βάση (β) Πρόσωπο εισόδου (γ) Προσαρμογή του αποθηκευμένου στη βάση προσώπου με βάση τη κατανομή φωτεινότητας του προσώπου εισόδου -“Face version”.

4. ΔΙΑΔΙΚΑΣΙΑ ΤΑΥΤΟΠΟΙΗΣΗΣ

Η ταυτοποίηση εφαρμόζεται στο χώρο των εικόνων και συγκρίνεται το πρόσωπο εισόδου με τις “face versions” των προσώπων της βάσης. Για να μειωθεί η υπολογιστική πολυπλοκότητα της παραπάνω διαδικασίας η τεχνική δεν εφαρμόζεται σε όλο το πλήθος των αποθηκευμένων στη βάση προσώπων αλλά προηγείται μια διαδικασία δεικτοδότησης ενός ποσοστού από αυτά με κριτήριο τους συντελεστές DCT των μπλοκ των εικόνων. Η διαδικασία της δεικτοδότησης επιλέγει με αποτελεσματικό τρόπο ένα ποσοστό 10%

των αποθηκευμένων προσώπων στα οποία περιέχεται και το πρόσωπο στόχος. Για την αποτελεσματικότητα της ομαδοποίησης σύμφωνα με την παραπάνω διαδικασία βλέπε Κεφάλαιο Γ3 και εργασία [7]. Επιπλέον πρέπει να διευκρινιστεί ότι οι εικόνες της βάσης έχουν ήδη αναλυθεί σύμφωνα με την τεχνική της ανάλυσης ιδιαζουσών τιμών και ένας περιορισμένος αριθμός ιδιαζόντων διανυσμάτων αποθηκεύεται ως διάνυμα χαρακτηριστικών για την κάθε εικόνα -βλέπε Κεφάλαιο Γ5. Η αποδοτικότητα του αλγορίθμου είναι πάντως πολύ υψηλή ακόμα και όταν χρησιμοποιηθεί πολύ μικρός αριθμός ιδιαζόντων διανυσμάτων -2 έως 4. Η διαδικασία δεικτοδότησης των πιθανότερων προσώπων απαιτεί την δημιουργία μικρού μόνο αριθμού “face versions” από τα αποθηκευμένα πρόσωπα.

Στον Πίνακα Γ4.1 εμφανίζονται τα αποτελέσματα ορθής ομαδοποίησης με χρήση διαφορετικού αριθμού συντελεστών για κάθε μπλοκ 8×8. Η ορθή ομαδοποίηση αναφέρεται στην περίπτωση κατά την οποία το ζητούμενο αποθηκευμένο στη βάση πρόσωπο, περιλαμβάνεται στο 10% των πιθανών προσώπων που προκύπτουν από τη δεικτοδότηση μέσω των συντελεστών του μετασχηματισμού DCT. Ως μέτρο ταύτισης που χρησιμοποιήθηκε το συνελικτικό σχήμα του προηγούμενου Κεφαλαίου. Παρατηρούμε ότι με χρήση 6 συντελεστών ανά μπλοκ επιτυγχάνεται ορθή ομαδοποίηση της τάξης του 99%. Τα πειράματα εκτελέστηκαν στη βάση ORL.

Αριθμός συντελεστών DCT (ανά μπλοκ 8×8)		
1	3	6
Ορθή Ομαδοποίηση (%)		
81	95	99

Πίνακας Γ4.1: Ορθή ομαδοποίηση αφορά την περίπτωση κατά την οποία το πρόσωπο στόχος περιλαμβάνεται στο σύνολο των υποψήφιων - πιθανών προσώπων.

Μετά τη διαδικασία ομαδοποίησης πραγματοποιείται η ταυτοποίηση ανάμεσα στο πρόσωπο εισόδου και τις “face versions” των προσώπων της βάσης που προέκυψαν από ην ομαδοποίηση.

Το αποθηκευμένο πρόσωπο \mathbf{S}_m το οποίο ταιριάζει σε μεγαλύτερο βαθμό με το πρόσωπο εισόδου λαμβάνεται από την παρακάτω σχέση η οποία απαιτεί την ελαχιστοποίηση της νόρμας Frobenious (l_2) των διαφορών των πινάκων $\hat{\mathbf{S}}_k$ και \mathbf{I} :

$$\mathbf{S}_m = \arg \min_k \left\| \hat{\mathbf{S}}_k - \mathbf{I} \right\|_F = \arg \min_k \left\| \mathbf{U}_{S_k} \boldsymbol{\Sigma}_1 \mathbf{V}_{S_k}^T - \mathbf{U}_I \boldsymbol{\Sigma}_I \mathbf{V}_I^T \right\|_F \quad (\Gamma 4.11)$$

Ο ευθύς υπολογισμός της ανωτέρω σχέσης απαιτεί πολύ χρόνο και δεν είναι αποδοτικός. Στην συνέχεια παρουσιάζεται ένας απλούστερος προσεγγιστικός υπολογισμός της εξίσωσης (Γ4.11).

Το τετράγωνο της νόρμας Frobenious δίνεται από τη σχέση:

$$\begin{aligned} \left\| \mathbf{I} - \hat{\mathbf{S}} \right\|_F^2 &= \text{tr} \left\{ (\mathbf{I} - \hat{\mathbf{S}})^T (\mathbf{I} - \hat{\mathbf{S}}) \right\} \\ &= \text{tr} \left\{ (\mathbf{I}^T \mathbf{I}) + (\hat{\mathbf{S}}^T \hat{\mathbf{S}}) + (\hat{\mathbf{S}}^T \mathbf{I}) + (\mathbf{I}^T \hat{\mathbf{S}}) \right\} \end{aligned}$$

$$\begin{aligned}
&= tr(I^T I) + tr(\hat{\mathbf{S}}^T \hat{\mathbf{S}}) - tr(\hat{\mathbf{S}}^T I) - tr(I^T \hat{\mathbf{S}}) \\
&= tr(I^T I) + tr(\hat{\mathbf{S}}^T \hat{\mathbf{S}}) - 2 \cdot tr(I^T \hat{\mathbf{S}})
\end{aligned} \tag{Γ4.12}$$

όπου $tr(A)$ είναι το άθροισμα των διαγωνίων στοιχείων του τετραγωνικού πίνακα A και $\|A\|_F$ η Frobenious νόρμα του.

Οι τρεις παράγοντες της σχέσης (Γ4.12) αναλύονται περαιτέρω όπως παρακάτω:

$$\begin{aligned}
tr(I^T I) &= tr\{(\mathbf{U}_1 \boldsymbol{\Sigma}_1 \mathbf{V}_1^T)^T (\mathbf{U}_1 \boldsymbol{\Sigma}_1 \mathbf{V}_1^T)\} \\
&= tr(\mathbf{V}_1 \boldsymbol{\Sigma}_1^T \mathbf{U}_1^T \mathbf{U}_1 \boldsymbol{\Sigma}_1 \mathbf{V}_1^T) \\
&= tr(\mathbf{V}_1 \boldsymbol{\Sigma}_1^T \boldsymbol{\Sigma}_1 \mathbf{V}_1^T) \\
&= tr\left(\sum_{i=1}^k \lambda_i^2 \mathbf{v}_{\mathbf{I}i} \cdot \mathbf{v}_{\mathbf{I}i}^T\right) \\
&= \sum_{i=1}^k \lambda_i^2 \cdot \|\mathbf{v}_{\mathbf{I}i}\|^2
\end{aligned} \tag{Γ4.13}$$

όπου $\mathbf{v}_{\mathbf{I}i}$ το i -οστό ιδιάζων διάνυσμα του πίνακα $I^T I$ και k οι μη μηδενικές ιδιοτιμές του ίδιου πίνακα. Δεδομένου ότι $\|\mathbf{v}_{\mathbf{I}i}\|_2 = 1$ αφού οι πίνακες $\mathbf{U}_1, \mathbf{V}_1$ είναι ορθοκανονικοί η σχέση (Γ5.5.2) καταλήγει στη σχέση:

$$tr(I^T I) = \sum_{i=1}^k \lambda_i^2 \tag{Γ4.14}$$

ομοίως $tr(\hat{\mathbf{S}}^T \hat{\mathbf{S}}) = \sum_{i=1}^k \lambda_i^2$ (αφού από τον ορισμό του ο πίνακας $\hat{\mathbf{S}}$ έχει τις ίδιες ιδιοτιμές με τον πίνακα I) οπότε:

$$\|I - \hat{\mathbf{S}}\|_F^2 = 2 \sum_{i=1}^k \lambda_i^2 - 2 \cdot tr(I^T \hat{\mathbf{S}}) \tag{Γ4.15}$$

ελαχιστοποίηση της ανωτέρω σχέσης απαιτεί την μεγιστοποίηση της ποσότητας $tr(I^T \hat{\mathbf{S}})$ η οποία είναι φραγμένη άνω από την ποσότητα $\sum_{i=1}^k \lambda_i^2$.

Το γινόμενο $I^T \hat{\mathbf{S}}$ μπορεί να γραφεί και ως:

$$I^T \hat{\mathbf{S}} = \mathbf{V}_1 \boldsymbol{\Sigma}_1^T \mathbf{U}_1^T \mathbf{U}_s \boldsymbol{\Sigma}_s \mathbf{V}_s^T = \mathbf{V}_1 \mathbf{R} \mathbf{V}_s^T \tag{Γ4.16}$$

$$\text{όπου } \mathbf{R} = \mathbf{\Sigma}_1^T \mathbf{U}_1^T \mathbf{U}_s \mathbf{\Sigma}_1 = \begin{pmatrix} \lambda_1^2 \langle \mathbf{u}_{11}, \mathbf{u}_{s1} \rangle & \lambda_1 \cdot \lambda_2 \langle \mathbf{u}_{11}, \mathbf{u}_{s2} \rangle & \dots \\ \lambda_1 \cdot \lambda_2 \langle \mathbf{u}_{12}, \mathbf{u}_{s1} \rangle & \lambda_2^2 \langle \mathbf{u}_{12}, \mathbf{u}_{s2} \rangle & \dots \\ \dots & \dots & \dots \end{pmatrix} \quad (\Gamma 4.17)$$

και $\langle \mathbf{u}_{1i}, \mathbf{u}_{sj} \rangle$ είναι το εσωτερικό γινόμενο των διανυσμάτων \mathbf{u}_{1i} και \mathbf{u}_{sj} . Προσεγγίζοντας τον πίνακα \mathbf{R} με τον διαγώνιο πίνακα

$$\hat{\mathbf{R}} = \begin{pmatrix} \lambda_1^2 \langle \mathbf{u}_{11}, \mathbf{u}_{s1} \rangle & 0 & \dots \\ 0 & \lambda_2^2 \langle \mathbf{u}_{12}, \mathbf{u}_{s2} \rangle & \dots \\ \dots & \dots & \dots \end{pmatrix} \quad (\Gamma 4.18)$$

η ποσότητα $tr(I^T \hat{\mathbf{S}})$ προσεγγίζεται από τη σχέση:

$$\begin{aligned} tr(I^T \hat{\mathbf{S}}) &\approx tr(\mathbf{V}_1 \hat{\mathbf{R}} \mathbf{V}_s^T) = tr\left(\sum_{i=1}^k \lambda_i^2 \cdot \langle \mathbf{u}_{1i}, \mathbf{u}_{si} \rangle \cdot \mathbf{v}_{1i} \cdot \mathbf{v}_{si}^T\right) \\ &= \sum_{i=1}^k \lambda_i^2 \cdot \langle \mathbf{u}_{1i}, \mathbf{u}_{si} \rangle \cdot \langle \mathbf{v}_{1i}, \mathbf{v}_{si} \rangle \end{aligned} \quad (\Gamma 4.19)$$

Η ποσότητα $\sum_{i=1}^k \lambda_i^2 \cdot \langle \mathbf{u}_{1i}, \mathbf{u}_{si} \rangle \cdot \langle \mathbf{v}_{1i}, \mathbf{v}_{si} \rangle$ μεγιστοποιείται όταν ισχύουν οι σχέσεις

$$\langle \mathbf{u}_{1i}, \mathbf{u}_{sj} \rangle = \begin{cases} 1 & \text{όταν } i = j \\ 0 & \text{όταν } i \neq j \end{cases} \quad \text{και} \quad \langle \mathbf{v}_{1i}, \mathbf{v}_{sj} \rangle = \begin{cases} 1 & \text{όταν } i = j \\ 0 & \text{όταν } i \neq j \end{cases}$$

οι ανωτέρω σχέσεις είναι αναγκαίες και ικανές για την μεγιστοποίηση και της ποσότητας $tr(I^T \hat{\mathbf{S}})$ επομένως μπορούμε να προσεγγίσουμε τη νόρμα $\|I - \hat{\mathbf{S}}\|_F$ με τη σχέση:

$$\|I - \hat{\mathbf{S}}\|_F = \sqrt{2\left(\sum_{i=1}^k \lambda_i^2 - \sum_{i=1}^k \lambda_i^2 \cdot \langle \mathbf{u}_{1i}, \mathbf{u}_{si} \rangle \cdot \langle \mathbf{v}_{1i}, \mathbf{v}_{si} \rangle\right)} \quad (\Gamma 4.20)$$

Δεδομένου ότι ο παράγοντας $\sum_{i=1}^k \lambda_i^2$ εξαρτάται αποκλειστικά από την εικόνα εισόδου η σχέση

$$\mathbf{S}_m = \arg \min_k \|\hat{\mathbf{S}}_k - I\|_F = \arg \min_k \|\mathbf{U}_{s_k} \mathbf{\Sigma}_1 \mathbf{V}_{s_k}^T - \mathbf{U}_1 \mathbf{\Sigma}_1 \mathbf{V}_1^T\|_F \quad (\Gamma 4.21)$$

είναι ισοδύναμη με την σχέση

$$\mathbf{S}_m = \arg \max_L \sum_{i=1}^k \lambda_i^2 \cdot \langle \mathbf{u}_{1i}, \mathbf{u}_{s_{Li}} \rangle \cdot \langle \mathbf{v}_{1i}, \mathbf{v}_{s_{Li}} \rangle \quad (\Gamma 4.22)$$

Στην πράξη για κάθε αποθηκευμένο στη βάση πρόσωπο κρατάμε τα q πρώτα ιδιοδιανύσματα με $q \ll k$ και με τυπική τιμή $q=5$. Η εξίσωση (Γ4.22) μετατρέπεται τότε στην εξίσωση:

$$\mathbf{S}_m = \arg \max_L \sum_{i=1}^q \lambda_i^2 \cdot \langle \mathbf{u}_{I_i}, \mathbf{u}_{S_{L,i}} \rangle \cdot \langle \mathbf{v}_{I_i}, \mathbf{v}_{S_{L,i}} \rangle \quad (\Gamma 4.23)$$

η οποία είναι υπολογιστικά πολύ απλούστερη από τη σχέση (Γ4.11)

5. ΠΕΙΡΑΜΑΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ

Τα πειραματικά αποτελέσματα ελήφθησαν με χρήση της βάσης προσώπων ORL. Για να διευκρινιστεί η αξία της προσαρμογής φωτεινότητας θεωρούμε τα κριτήρια ταύτισης \mathbf{s} , \mathbf{s}^* τα οποία αντιστοιχούν σε συγκρίσεις χωρίς και με προσαρμογή φωτισμού αντίστοιχα:

$$\mathbf{s}^* \Rightarrow \mathbf{S}_m = \arg \min_k \|\hat{\mathbf{S}}_k - I\|_F = \arg \min_k \|\mathbf{U}_{S_k} \Sigma_I \mathbf{V}_{S_k}^T - \mathbf{U}_I \Sigma_I \mathbf{V}_I^T\|_F \quad (\Gamma 4.24)$$

$$\mathbf{s} \Rightarrow \mathbf{S}_m = \arg \min_k \|\mathbf{S}_k - I\|_F = \arg \min_k \|\mathbf{U}_{S_k} \Sigma_{S_k} \mathbf{V}_{S_k}^T - \mathbf{U}_I \Sigma_I \mathbf{V}_I^T\|_F \quad (\Gamma 4.25)$$

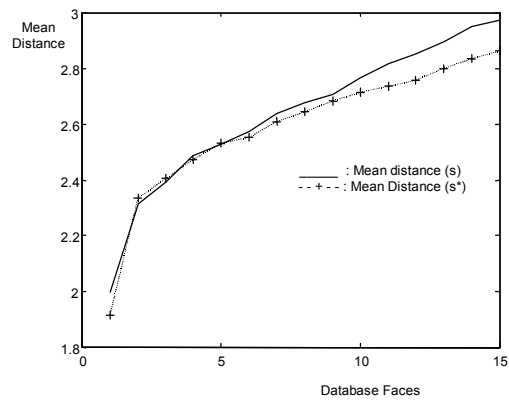
Προφανώς η σχέση (Γ4.25) δεν προσεγγίζεται από κάποια απλούστερη όπως γίνεται με την (Γ4.24). Οι συγκρίσεις πραγματοποιούνται υποχρεωτικά στο χώρο των τιμών φωτεινότητας της εικόνας. Τα αποτελέσματα της σύγκρισης επιδεικνύονται στον Πίνακα Γ4.2.

Διατηρούμενα ιδιάζοντα διανύσματα			
	1	3	5
Ποσοστά αναγνώρισης (%)			
Κριτήριο \mathbf{s}	76	79	83
Κριτήριο \mathbf{s}^*	77	85	91

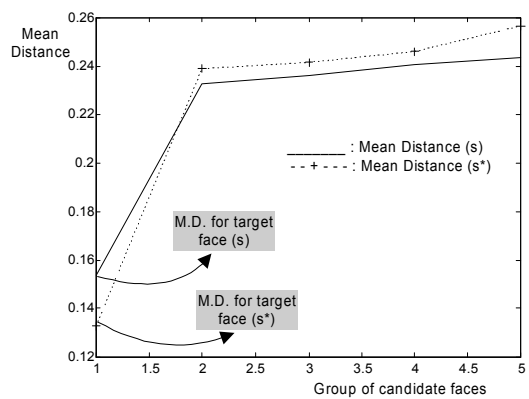
Πίνακας Γ4.2: Ποσοστά αναγνώρισης, για τα κριτήρια \mathbf{s} και \mathbf{s}^* σε σχέση με τον αριθμό των ιδιάζοντων διανυσμάτων που χρησιμοποιούνται για τη δημιουργία των "face version".

Επιπλέον χρειάζεται να διερευνηθεί η επίδραση της προσαρμογής φωτεινότητας σε σχέση με τον βαθμό ομοιότητας του προς ταυτοποίηση πρόσωπου με τα αποθηκευμένα στη βάση πρόσωπα. Ο στόχος είναι να επιβεβαιωθεί ότι η προσαρμογή φωτεινότητας δεν μειώνει ισοδύναμα την απόσταση ανάμεσα το προς ταυτοποίηση πρόσωπο και όλα τα πρόσωπα της βάσης αλλά μειώνει σε μεγαλύτερο βαθμό την απόσταση ανάμεσα στο πρόσωπο εισόδου και την αποθηκευμένη στη βάση εκδοχή του –πρόσωπο στόχος. Για να καταλήξουμε σε ένα αξιόπιστο συμπέρασμα υπολογίσαμε και σχεδιάσαμε τη μέση απόσταση ανάμεσα στα πρόσωπα εισόδου και τα αποθηκευμένα στη βάση. Το Σχήμα Γ4.1 επιδεικνύει ότι η μείωση της απόστασης ανάμεσα στα πρόσωπα εισόδου και τα αποθηκευμένα στη βάση, εξαιτίας της προσαρμογής φωτεινότητας, δεν έχει καμιά επίπτωση στην ευστάθεια της ταυτοποίησης δεδομένου ότι παραμένει σημαντικά μεγαλύτερη από ότι η απόσταση ανάμεσα στα πρόσωπα εισόδου και στα πρόσωπα στόχο. Επιπλέον αναλύοντας την απόσταση ανάμεσα στα πρόσωπα εισόδου και τα πρόσωπα της βάσης συμπεραίνουμε – βλέπε Σχήμα Γ4.2- ότι η προσαρμογή φωτεινότητας όχι μόνο βελτιώνει τα ποσοστά αναγνώρισης αλλά

οδηγεί και σε πιο ευσταθείς ταυτοποιήσεις δεδομένου ότι η διαφορά ανάμεσα στις αποστάσεις πρόσωπα εισόδου-πρόσωπα στόχος και πρόσωπα εισόδου-σύνολο προσώπων της βάσης, αυξάνεται.



Σχήμα Γ4.1: Μέση απόσταση ανάμεσα στα πρόσωπα εισόδου και τα πρόσωπα της βάσης σε σχέση με τα κριτήρια s και s^* .



Σχήμα Γ4.2: Μέση απόσταση ανάμεσα στα πρόσωπα στόχο και τα πρόσωπα που προέκυψαν από τη διαδικασία ομαδοποίησης, σε σχέση με τα κριτήρια s και s^* .

6. ΣΥΜΠΕΡΑΣΜΑΤΑ

Στο κεφάλαιο αυτό παρουσιάστηκε μια μέθοδος για προσαρμογή των συνθηκών γενικού φωτισμού και την εξάλειψη της επίδρασης τους. Τα πειραματικά αποτελέσματα επιδεικνύουν την αποδοτικότητα της προτεινόμενης μεθόδου να διατηρεί το σχηματικό περιεχόμενο της εικόνας υπό διαφορετικές συνθήκες φωτισμού. Η προσαρμογή της φωτεινότητας οδηγεί σε αυξημένα ποσοστά αναγνώρισης καθώς και σε μεγαλύτερη ευστάθεια και αξιοπιστία του συνολικού σχήματος ταυτοποίησης.

ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΣΩΠΩΝ ΜΕ ΧΡΗΣΗ ΠΟΛΛΑΠΛΩΝ ΑΝΑΠΑΡΑΣΤΑΣΕΩΝ

Το προτεινόμενο στο κεφάλαιο αυτό σχήμα στοχεύει στην αναγνώριση προσώπων ανεξάρτητα από διακυμάνσεις στη θέση, οριοθέτηση, κλίμακα, φωτισμό και υφή μέσα από ένα διαφορετικό πρίσμα. Συγκεκριμένα αντί να προσπαθήσουμε να βρούμε ένα γενικό σύνολο από χαρακτηριστικά το οποίο να είναι ανεπηρέαστο από όλους τους παραπάνω παράγοντες, πράγμα εξαιρετικά δύσκολο, επιλέγουμε πολλαπλές αλληλοσυμπληρούμενες αναπαραστάσεις οι οποίες επιλύουν ένα συγκεκριμένο πρόβλημα η καθεμιά. Οι πολλαπλές συμπληρωματικές αναπαραστάσεις υιοθετούνται με στόχο τον κατακερματισμό του χώρου των σφαλμάτων, ο οποίος υποτίθεται ότι αποτελείται από τη σύνθεση προβλημάτων που οφείλονται σε διακυμάνσεις στη οριοθέτηση, κλίμακα, φωτισμό και υφή, και οδηγούν σε επιτυχείς μερικές λύσεις οι οποίες εκτελούνται παράλληλα. Όπως είναι φυσικό, κάθε αναπαράσταση χρησιμοποιεί τον δικό της αλγόριθμο ταυτοποίησης ο οποίος είναι προσαρμοσμένος σε αυτήν. Σε ένα επόμενο τελικό στάδιο μια μηχανή λήψης απόφασης βασισμένη σε κανόνες επιλέγει την πιο αξιόπιστη λύση εκτιμώντας ποιο ήταν το πρόβλημα που προκάλεσε την εντός κλάσης διακύμανση, αν υπάρχει, και λαμβάνοντας υπόψη τα αποτελέσματα των μερικών ταυτοποιήσεων.

1. ΕΝΤΟΠΙΣΜΟΣ ΚΑΙ ΚΑΤΑΤΜΗΣΗ ΤΟΥ ΠΡΟΣΩΠΟΥ

Το στάδιο προεπεξεργασίας καθιστά εύρωστη ολόκληρη την διαδικασία. Σκοπός του είναι ο εντοπισμός του προσώπου και η κατάτμηση της περιοχής που αυτό καλύπτει στην φωτογραφία εισόδου. Η διαδικασία αυτή πραγματοποιείται με βάση τον αλγόριθμο που περιγράφηκε στο Κεφάλαιο Β1. Τα χαρακτηριστικά περιγραφής του προσώπου λαμβάνονται από την περιοχή της έλλειψης η οποία αναπαριστά το κατατμημένο τμήμα του προσώπου. Εναλλακτικά θα μπορούσαν να εφαρμοστούν *ενεργά περιγράμματα* με αρχική καμπύλη την περιφέρεια της έλλειψης για λεπτομερέστερη προσαρμογή στο περιγράμμα του προσώπου. Στα επόμενα ο όρος *κρίσιμη επιφάνεια* αναφέρεται στην επιφάνεια της έλλειψης η οποία έχει κατατμηθεί από το φόντο.

2. ΣΥΝΟΛΑ ΠΕΡΙΓΡΑΦΗΣ

Για την αναπαράσταση των προσώπων χρησιμοποιούμε τρία διαφορετικά σύνολα χαρακτηριστικών δύο από τα οποία προκύπτουν από την ανάλυση ιδιάζουσών τιμών και το τρίτο βασίζεται στις γεωμετρικές ροπές.

2.1 Αναπαράσταση ανεξάρτητη κλίμακας

Ένας μετασηματισμός που δίνει εσωτερική πληροφορία για ένα αντικείμενο είναι οι Ροπές (Moments) [18]. Με δεδομένη την εξαγωγή της κρίσιμης επιφάνειας από το προηγούμενο στάδιο υπολογίζονται ένα σύνολο από ροπές που λαμβάνουν υπόψη τις τιμές των pixels και την θέση που κατέχουν στην περιοχή του

σχήματος και με αυτό τον τρόπο δημιουργείται ένας χώρος μικρής διάστασης που περιέχει πληροφορία για τη δομή του. Χρησιμοποιώντας τις ροπές μέχρι τρίτης τάξης για ένα διδιάστατο σήμα καταλήγουμε σε ένα χώρο με επτά διαστάσεις. Στον χώρο αυτό κάθε σχήμα αναπαρίσταται με ένα σημείο (ένα διάνυσμα) και έτσι είναι δυνατή η αποθήκευση πολλών διαφορετικών σχημάτων σε μια Βάση Δεδομένων με βάση το σημείο που αυτά ορίζουν.

Για μια μονόχρωμη ψηφιακή εικόνα με συνάρτηση φωτεινότητας $g(x,y)$ οι ροπές μέχρι τρίτης τάξης ορίζονται ως εξής:

$$m_{uv} = \sum_x \sum_y g(x,y) x^u y^v \quad u, v=0,1,2,3 \quad (\Gamma 5.1)$$

Η άθροιση γίνεται στην υποπεριοχή που καταλαμβάνει η *κρίσιμη επιφάνεια*. Από τον ορισμό τους οι ροπές δίνουν την συσχέτιση που θα έχουν οι διαφορετικές περιοχές υψής του σχήματος σε σχέση με την ένταση τους. Οι ροπές είναι ανεξάρτητες από την κλίμακα και την περιστροφή αλλά εξαρτώνται από την απόλυτη θέση του σχήματος μέσα στην εικόνα. Παρόλο που το ζήτημα αυτό υποτίθεται ότι έχει λυθεί κατά τη διαδικασία της κατάτμησης, για σκοπούς μεγαλύτερης ευστάθειας θεωρούμε τις Κεντρικές Ροπές (central moments) μ_{uv} οι οποίες παίρνουν ως σημείο αναφοράς το Κέντρο Βάρους του σχήματος πράγμα που τις καθιστά ανεξάρτητες από την απόλυτη θέση του σχήματος μέσα στην εικόνα.

$$\mu_{uv} = \sum_x \sum_y g(x,y) (x-\bar{x})^u (y-\bar{y})^v \quad \text{where } \bar{x} = \frac{m_{10}}{m_{00}} \quad \text{and } \bar{y} = \frac{m_{01}}{m_{00}} \quad (\Gamma 5.2)$$

Το σημείο (\bar{x}, \bar{y}) αποτελεί το Κέντρο Βάρους του εξεταζόμενου σχήματος. Στην περίπτωση δυαδικών εικόνων αυτό συμπίπτει με το γεωμετρικό Κέντρο Βάρους (Centroid).

Οι δέκα πρώτες κεντρικές ροπές ορίζονται ως εξής:

$$\begin{aligned} \mu_{00} &= m_{00}, & \mu_{10} &= \mu_{01} = 0, & \mu_{20} &= m_{20} - \bar{x} \cdot m_{10}, & \mu_{02} &= m_{02} - \bar{y} \cdot m_{01}, \\ \mu_{11} &= m_{11} - \bar{y} \cdot m_{10}, & \mu_{30} &= m_{30} - 3 \cdot \bar{x} \cdot m_{20} + 2 \cdot \bar{x}^2 \cdot m_{10}, & \mu_{03} &= m_{03} - 3 \cdot \bar{y} \cdot m_{02} + 2 \cdot \bar{y}^2 \cdot m_{01}, \\ \mu_{12} &= m_{12} - 2 \cdot \bar{y} \cdot m_{11} - \bar{x} \cdot m_{02} + 2 \cdot \bar{y}^2 \cdot m_{10}, & \mu_{21} &= m_{21} - 2 \cdot \bar{x} \cdot m_{11} - \bar{y} \cdot m_{20} + 2 \cdot \bar{x}^2 \cdot m_{01}. \end{aligned}$$

Οι κεντρικές αυτές ροπές μπορούν να κανονικοποιηθούν σύμφωνα με τους τύπους:

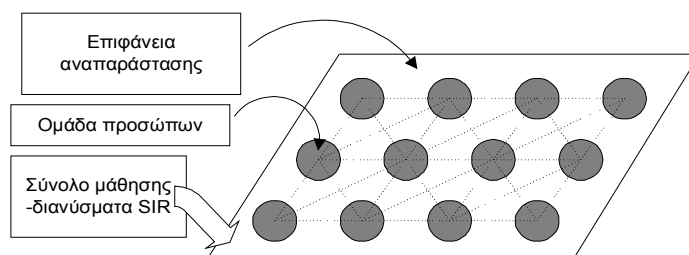
$$\begin{aligned} n_{00} &= \mu_{00} \\ n_{ij} &= \frac{\mu_{ij}}{(\mu_{00})^k}, \quad k = \frac{i+j}{2} + 1, \quad j + i \neq 0 \end{aligned} \quad (\Gamma 5.3)$$

Με αυτό τον τρόπο έχουμε δημιουργήσει ένα χώρο με επτά διαστάσεις –οι ροπές $n_{10} = n_{01} = 0$ και $n_{11} \rightarrow 0$ δεν υπολογίζονται. Σε αυτό το χώρο κάθε πρόσωπο παριστάνεται με ένα σημείο. Έτσι το πρόβλημα αναγνώρισης μετατρέπεται σε ένα πρόβλημα ταύτισης στον μικρότερο αυτό χώρο. Όπως προκύπτει από την εξέλιξη των υπολογισμών οι Κανονικοποιημένες Κεντρικές Ροπές δίνουν πληροφορία

υψής για το σχήμα και είναι ανεξάρτητες από την τοποθέτηση, κλίμακα και το μέγεθος του σχήματος σε σχέση με την εικόνα. Το διάνυσμα που περιγράφει κάθε πρόσωπο στο συγκεκριμένο χώρο (Scale Invariant Representation-SIR) είναι: $SIR = [n_{00}, n_{20}, n_{02}, n_{30}, n_{03}, n_{12}, n_{21}]^T$

Έχει προταθεί σε πολλές εργασίες ότι η ομαδοποίηση των διανυσμάτων χαρακτηριστικών οδηγεί σε περισσότερο αξιόπιστα αποτελέσματα σε προβλήματα όπως αυτό της ταυτοποίησης προσώπων [15]. Στην προτεινόμενη τεχνική η ομαδοποίηση αυτή στηρίζεται στην θεωρία των νευρωνικών δικτύων και ειδικότερα των Αυτοδιαρθρωμένων Χάρτων (Self Organized Maps - SOM) οι οποίοι προτάθηκαν πρώτη φορά για σκοπούς ταξινόμησης από τον Kohonen [19]. Οι Αυτοδιαρθρωμένοι Χάρτες είναι μια μη επιβλεπόμενη διαδικασία μάθησης η οποία προσεγγίζει την κατανομή ενός συνόλου από πρότυπα χωρίς να της παρέχεται κάποια πληροφορία σχετικά με τον αριθμό ή τη δομή των κλάσεων. Τα πρότυπα προβάλλονται από το χώρο εισόδου σε μια θέση στο Χάρτη, πληροφορία που κωδικοποιείται από τη θέση ενός ενεργού νευρώνα. Οι επιλογή των SOM ως σχήματος ομαδοποίησης έγινε επειδή αυτά παρέχουν μια τοπολογική διάταξη των κλάσεων. Η ομοιότητα στο χώρο εισόδου διατηρείται στην χώρο εξόδου καθιστώντας τα SOM ιδιαίτερα χρήσιμα για την κατηγοριοποίηση δεδομένων τα οποία περιέχουν μεγάλο και άγνωστο αριθμό από κλάσεις.

Το σύνολο μάθησης των SOM αποτελείται από τα SIR διανύσματα των προσώπων. Η ομαδοποίηση πραγματοποιείται με την αναγνώριση της συσχέτισης και των κρυμμένων ομοιοτήτων ανάμεσα στα διανύσματα εισόδου. Με τον τρόπο αυτό κάθε διάνυσμα προσώπου κατέχει μια θέση σε κάποια συγκεκριμένη κλάση –βλέπε Σχήμα Γ5.1. Το SOM εκπαιδεύτηκε με τον αλγόριθμο του Kohonen και επομένως διατηρεί την τοπολογία του χώρου εισόδου με τρόπο παρόμοιο με αυτό που λειτουργεί η λογική του ανθρώπου σε αντίστοιχες περιπτώσεις.



Σχήμα Γ5.1: Η αρχιτεκτονική του SOM

2.2 Αναπαράσταση με βάση την υφή και το φωτισμό

Όπως φάνηκε στο Κεφάλαιο Γ4, η χρήση του διανύσματος των ιδιαιδισμών τιμών παρέχει πληροφορία σχετικά με την κατανομή της υψής και του φωτισμού σε μια εικόνα ενώ δεν λαμβάνει υπόψη τους άξονες της. Όπως ήδη αναφέρθηκε, για κάθε πραγματικό ορθογώνιο πίνακα A και υπό τον περιορισμό $\lambda_1 > \lambda_2 > \dots > \lambda_k$, το διάνυσμα των ιδιαιδισμών τιμών $\mathbf{x}_{n \times 1}$ είναι μοναδικό. Η μοναδικότητα και άλλες σημαντικές ιδιότητες όπως η αναισθησία στη περιστροφή και στη κλίμακα [8], του διανύσματος $\mathbf{x}_{n \times 1}$

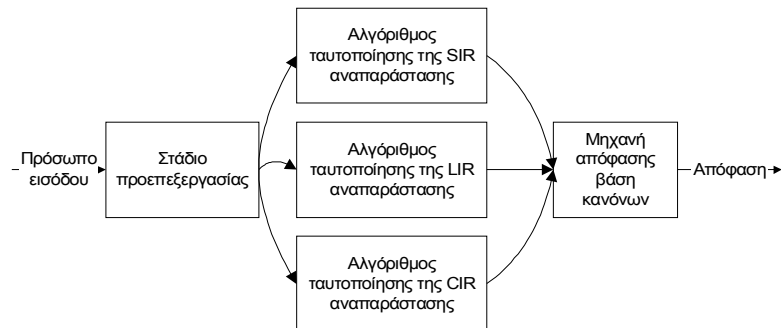
καθιστούν κατάλληλη επιλογή για την αναπαράσταση CIR -‘Content’ Invariant Representation-δηλαδή $CIR = [\lambda_1 \lambda_2, \dots, \lambda_k]$.

2.3 Αναπαράσταση ανεξάρτητη από το γενικό φωτισμό - Luminance Invariant Representation

Η αναπαράσταση αυτή προσπαθεί να επιλύσει το πρόβλημα των διακυμάνσεων φωτεινότητας ανάμεσα στο πρόσωπο εισόδου και στην αποθηκευμένη στη βάση εκδοχή του. Η μέθοδος στηρίζεται στην αποσύνθεση ιδιαζουσών τιμών αμφοτέρων των εικόνων εισόδου και βάσης. Στο Κεφάλαιο Γ4 –βλέπε και εργασία [20]- έχει επιδειχθεί ότι το ουσιαστικό περιεχόμενο μιας εικόνας –άξονες- διατηρείται μέσω των ιδιάζοντων διανυσμάτων. Με βάση αυτή την παραδοχή τα ιδιάζοντα διανύσματα μπορούν να αναπαραστήσουν το περιεχόμενο των εικόνων προσώπου ανεξάρτητα από τις γενικότερες συνθήκες φωτισμού. Η αναπαράσταση LIR –Luminance Invariant Representation- αποτελείται από τα τρία σημαντικότερα ιδιάζοντα διανύσματα των πινάκων U_s, V_s -βλέπε Κεφάλαιο Γ4.

3. ΑΛΓΟΡΙΘΜΟΣ ΤΑΥΤΟΠΟΙΗΣΗΣ

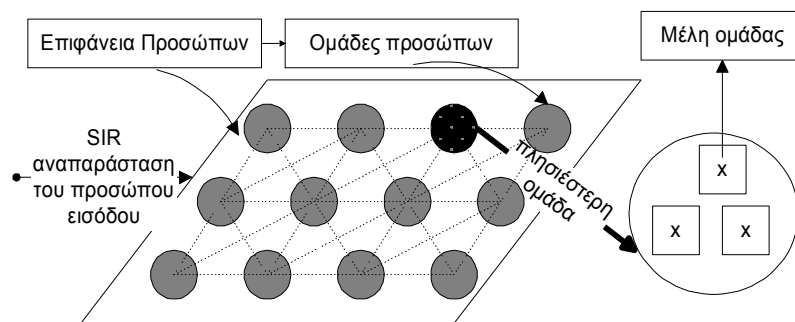
Η αρχιτεκτονική του αλγόριθμου αναγνώρισης επιδεικνύεται στο διάγραμμα του Σχήματος Γ5.2: Για κάθε πρόσωπο εισόδου πραγματοποιούνται οι διαδικασίες εντοπισμού και κατάτμησης και στη συνέχεια υπολογίζονται οι αναπαραστάσεις CIR, LIR και SIR. Για κάθε αναπαράσταση χρησιμοποιείται διαφορετικός αλγόριθμος ταυτοποίησης και τα επιμέρους αποτελέσματα συνδυάζονται με τη χρήση των κανόνων της μηχανής λήψης απόφαση οδηγώντας στην τελική ταυτοποίηση.



Σχήμα Γ5.2: Η αρχιτεκτονική του αλγόριθμου ταυτοποίησης

3.1 Ταυτοποίηση ανεξάρτητη από κλίμακα

Το διάνυσμα SIR, εφοδιασμένο με την μεθοδολογία ομαδοποίησης μέσω των SOM, χρησιμοποιείται για την διαδικασία ταυτοποίησης ανεξάρτητα από την κλίμακα. Για να καταλήξουμε στην αποθηκευμένη εκδοχή του προσώπου εισόδου, εφόσον αυτή υπάρχει, η SI αναπαράσταση του προσομοιώνεται μέσω του δικτύου SOM. Μέσω των πινάκων του δικτύου επιλέγεται η πλησιέστερη στο διάνυσμα εισόδου ομάδα. Τα μέλη της ομάδας αυτής αποτελούν τα πλησιέστερα στο πρόσωπο εισόδου αποθηκευμένα πρόσωπα σύμφωνα με την SIR (βλέπε Σχήμα Γ5.3).



Σχήμα Γ5.3: Ταυτοποίηση μέσω του Δικτύου SOM

Για να καταλήξουμε στην επικρατέστερη επιλογή από τα μέλη της ομάδας πρέπει να επιλεγεί το πλησιέστερο προς το διάνυσμα εισόδου με βάση κάποια μετρική. Για την επιλογή αυτή χρησιμοποιείται ένα δίκτυο LVQ (Learning Vector Quantization) το οποίο αρχικοποιείται, εκπαιδεύεται και εκτελεί την απαιτούμενη προσομοίωση με βάση την επιλεγείσα από το SOM ομάδα. Η επιλογή του συνδυασμού SOM και LVQ για την διαδικασία ταυτοποίησης δικαιώνεται από τα πειραματικά αποτελέσματα όπως φαίνεται στον Πίνακα Γ5.1 όπου χρησιμοποιήθηκαν ακόμη ο συντελεστής συσχέτισης και η μετρική I_1 .

Δεδομένα	Συντελεστής Συσχέτισης (%)	I_1 -νόρμα (%)	LVQ (%)
Μη ομαδοποιημένα	67	79	85
Ομαδοποίηση SOM	70	83	96

Πίνακα Γ5.1: Ποσοστά επιτυχούς ταυτοποίησης

3.2 Ταυτοποίηση ανεξάρτητη από φωτισμό

Όπως αναφέρθηκε στην προηγούμενη παράγραφο, τα ιδιάζοντα διανύσματα προσδιορίζουν τους βασικούς άξονες της εικόνας και επομένως περιέχουν πληροφορία υψίσυχνου περιεχομένου η οποία είναι ανεπηρέαστη από την κατανομή το φωτισμού. Η ταυτοποίηση που περιγράφεται στην τρέχουσα παράγραφο επικεντρώνεται στην εύρεση του αποθηκευμένου στη βάση προσώπου το οποίου το ουσιαστικό περιεχόμενο πλησιάζει περισσότερο από κάθε άλλο στο αντίστοιχο περιεχόμενο του προς ταυτοποίηση προσώπου. Η τεχνική βασίζεται στην σύγκριση των κατευθύνσεων των βασικών αξόνων των δύο εικόνων όπως αυτές περιγράφονται από τα αντίστοιχα ιδιάζοντα διανύσματα –βλέπε Κεφάλαιο Γ4.

Από τον ορισμό τους, τα ιδιάζοντα διανύσματα εκφράζουν κατευθύνσεις αξόνων. Κατά συνέπεια η ταυτοποίηση με βάση τα ιδιάζοντα διανύσματα δεν είναι σε καμία περίπτωση ανεξάρτητη από την περιστροφή.

3.3 Ταυτοποίηση βασισμένη στην πληροφορία υψής

Όπως έχει παρουσιαστεί στο Κεφάλαιο Γ4 η πληροφορία ουσιαστικού περιεχομένου μιας εικόνας διατηρείται μέσω των ιδιζόντων διανυσμάτων ενώ η πληροφορία υψής και φωτισμού περιέχεται στις ιδιάζουσες τιμές. Αυτό το τμήμα του αλγορίθμου επικεντρώνεται στην εύρεση του πλησιέστερου αποθηκευμένου προσώπου σύμφωνα με την πληροφορία υψής και ανεξάρτητα από διαφορές στο ουσιαστικό περιεχόμενο των εικόνων. Όπως αναφέρθηκε νωρίτερα οι ιδιάζουσες τιμές είναι ανεξάρτητες από μεταβολές στην μεταφορά, περιστροφή και ουσιαστικό περιεχόμενο.

Έστω $\mathbf{I} = \mathbf{U}_1 \mathbf{\Sigma}_1 \mathbf{V}_1^T$ το προς ταυτοποίηση πρόσωπο εισόδου. Για να επικεντρωθούμε αποκλειστικά στην πληροφορία υψής ανακατασκευάζουμε όλες τις αποθηκευμένες στη βάση εικόνες \mathbf{Y}_i χρησιμοποιώντας τα ιδιάζοντα διανύσματα της εικόνας \mathbf{I} και τις ιδιάζουσες τιμές της εικόνας \mathbf{Y}_i σύμφωνα με τη παρακάτω σχέση: $\hat{\mathbf{Y}}_i = \mathbf{U}_1 \mathbf{\Sigma}_{y_i} \mathbf{V}_1^T$.

Η βέλτιστη ταύτιση \mathbf{Y}_m θεωρείται εκείνη που ελαχιστοποιεί την Frobenius νόρμα της διαφοράς των πινάκων $\hat{\mathbf{Y}}_i$ και \mathbf{I} :

$$\mathbf{Y}_m = \arg \min_i \|\hat{\mathbf{Y}}_i - \mathbf{I}\| = \|\mathbf{U}_1 \mathbf{\Sigma}_{y_i} \mathbf{V}_1^T - \mathbf{U}_1 \mathbf{\Sigma}_1 \mathbf{V}_1^T\| = \|\mathbf{U}_1 (\mathbf{\Sigma}_{y_i} - \mathbf{\Sigma}_1) \mathbf{V}_1^T\| \quad (\Gamma 5.4)$$

Δεδομένου ότι $\|\hat{\mathbf{Y}}_i\| = \|\mathbf{Y}_i\| = \|\mathbf{\Sigma}_{y_i}\|$, χρησιμοποιώντας την εξίσωση (Γ5.4) καταλήγουμε στη σχέση:

$$\mathbf{Y}_m = \arg \min_i \|\mathbf{\Sigma}_{y_i} - \mathbf{\Sigma}_1\| = \arg \min_i \left\{ \sqrt{\sum_j (\lambda_{y_j} - \lambda_{1j})^2} \right\} \quad (\Gamma 5.5)$$

όπου λ_{yj} είναι η j -στη ιδιάζουσα τιμή της εικόνας \mathbf{Y}_i .

Με τη χρήση της εξίσωσης (Γ5.5), αποφεύγεται η ανακατασκευή των εικόνων της βάσης δεδομένου ότι η Frobenius νόρμα της διαφοράς μεταξύ των πινάκων $\hat{\mathbf{Y}}_i$ και \mathbf{I} υπολογίζεται άμεσα από τις ιδιάζουσες τιμές τους.

4. ΜΗΧΑΝΗ ΛΗΨΗΣ ΑΠΟΦΑΣΗΣ

Δεδομένου ότι ο αλγόριθμος περιλαμβάνει τρία διαφορετικά συστήματα ταυτοποίησης είναι λογικό να ενσωματώσουμε ένα τελευταίο στάδιο βασισμένο σε κανόνες το οποίο να τα συνδυάζει. Οι κανόνες δημιουργήθηκαν με βάση τις παραδοχές που έχουν γίνει κατά τη σχεδίαση του όλου συστήματος. Η επιτυχία των προσομοιώσεων, όσον αφορά στις επιτυχείς ταυτοποιήσεις, επιβεβαιώνει την καταλληλότητα της συνολικής προσέγγισης στην αναγνώριση προσώπων. Η μηχανή λήψης απόφασης παίρνει ως είσοδο τα αποτελέσματα των μερικών αναγνωρίσεων και επιλέγει την πιο αξιόπιστη, εφόσον υπάρχει. Η τελική απόφαση, είτε επιβεβαιώνει την αναγνώριση, είτε επαναλαμβάνει την διαδικασία δημιουργίας της βάσης δεδομένων, λαμβάνοντας υπόψη και το νεοεισαχθέν πρόσωπο (προσθέτει τα διανύσματα αναπαράστασής

του στη βάση). Το σύστημα λήψης απόφασης αποτελείται από ένα σύνολο κανόνων της δομής IF-THEN-ELSE πάνω σε χώρο συνόλων που ορίζονται βάση των πειραματικών αποτελεσμάτων.

Ορισμοί συνόλων

Έστω ότι τα στοιχεία I_S , I_L και I_C αντιπροσωπεύουν το αποτέλεσμα της αναγνώρισης από την ελάχιστη αναπαράσταση και συγκεκριμένα τις SIR, LIR και CIR αντίστοιχα. Ορίζουμε επίσης τα επόμενα σύνολα:

S: Αποθηκευμένα στη βάση πρόσωπα τα οποία ανήκουν στην ίδια ομάδα με το πρόσωπο εισόδου σύμφωνα με την κατηγοριοποίηση του SOM και με βάση την SI αναπαράσταση.

L: Σύνολο από τα m πιθανότερα πρόσωπα σύμφωνα με την LI αναπαράσταση ($m = 5\%$ του μεγέθους της βάσης).

C: Σύνολο από τα m πιθανότερα πρόσωπα σύμφωνα με την CI αναπαράσταση.

C_S: Υποσύνολο του **C** το οποίο περιέχει τα k ($k \ll m$) πιθανότερα πρόσωπα σύμφωνα με την SIR.

L_S: Υποσύνολο του **L** το οποίο περιέχει τα k πιθανότερα πρόσωπα σύμφωνα με την SIR.

S_L: Υποσύνολο του **S** το οποίο περιέχει τα k πιθανότερα πρόσωπα σύμφωνα με την LIR.

C_L: Υποσύνολο του **C** το οποίο περιέχει τα k πιθανότερα πρόσωπα σύμφωνα με την LIR.

S_C: Υποσύνολο του **S** το οποίο περιέχει τα k πιθανότερα πρόσωπα σύμφωνα με την CIR.

L_C: Υποσύνολο του **L** το οποίο περιέχει τα k πιθανότερα πρόσωπα σύμφωνα με την CIR.

Η δομή των κανόνων είναι ιεραρχική και η όλη διαδικασία σταματά όταν αποδοθεί τιμή στην έξοδο X .

Κανόνας 1: IF ($I_S = I_L$ OR $I_S = I_C$) THEN $X = I_S$ ELSEIF $I_L = I_C$ THEN $X = I_L$

Κανόνας 2: IF ($I_S \in L_S$ OR $I_S \in C_S$) THEN $X = I_S$

Κανόνας 3: IF ($I_L \in S_L$ OR $I_L \in C_L$) THEN $X = I_L$

Κανόνας 4: IF ($I_C \in S_C$ OR $I_C \in L_C$) THEN $X = I_C$

Κανόνας 5: IF $X = \emptyset$ THEN αναδιοργάνωσε τη βάση προσθέτοντας και το πρόσωπο εισόδου

Στους παραπάνω κανόνες συνδυάζεται το σιεπτικό της προτεινόμενης μεθόδου μαζί με τις υλοποιημένες τεχνικές ταυτοποίησης και τα αποτελέσματα τους. Ο πρώτος κανόνας εξασφαλίζει την επιβεβαίωση των αναγνωρίσεων στις περιπτώσεις που δεν υπάρχει σημαντική διαφορά ανάμεσα στο προς αναγνώριση πρόσωπο και στην αντίστοιχη αποθηκευμένη εκδοχή του. Ο κανόνας 1 στηρίζεται στο γεγονός ότι σε τέτοιες περιπτώσεις θα υπάρξει σωστή αναγνώριση από δύο τουλάχιστον τεχνικές. Ο παραπάνω κανόνας καλύπτει ακόμη και αναγνωρίσεις με μοναδικό πρόβλημα την τοποθέτηση αφού αυτό λύνεται κατά το

στάδιο της προεπεξεργασίας. Σύμφωνα και με την υλοποίηση των τεχνικών αναγνώρισης, αφού δίνουν σωστό αποτέλεσμα υπό διαφορετικές συνθήκες αποικλείεται να δώσουν το ίδιο λανθασμένο αποτέλεσμα στις ίδιες συνθήκες.

Οι κανόνες 2, 3 ελέγχουν το αποτέλεσμα της αναγνώρισης μέσω του SIR. Αυτό δίνει σωστή περιγραφή όταν υπάρχουν προβλήματα τοποθέτησης, κλίμακας και κλίσης. Σε μια τέτοια περίπτωση το σφάλμα ανακατασκευής της εικόνας θα είναι μικρό και το πρόσωπο θα βρίσκεται σε ανάλογη θέση στην πιθανή ομάδα και σε ακόμη ψηλότερη στην ταξινομημένη. Με αυτούς τους κανόνες ελέγχεται και ενδεχόμενη σύγκρουση ανάμεσα στην δεύτερη και τρίτη μέθοδο αναγνώρισης.

Ο κανόνας 4 ελέγχει τις αναγνωρίσεις της μεθόδου βασισμένης στο LIR. Αφού δεν ισχύουν οι κανόνες 1-3 σημαίνει ότι έχει αποτύχει η αναγνώριση μέσω του SIR. Ένα πρόβλημα που δεν λύνει η μέθοδος αυτή είναι αυτό του φωτισμού. Αν όντως συμβαίνει αυτό η μέθοδος της αναπαράστασης LIR θα δίνει σωστό αποτέλεσμα. Μια σχετικά ψηλή θέση στην πρώτη ομάδα το επιβεβαιώνει και έτσι γίνεται δεκτή η αναγνώριση που βασίστηκε στη απεικόνιση LIR.

Αν δεν ισχύουν τα προηγούμενα, σημαίνει ότι υπάρχει ένα πρόβλημα υψής ή μεταμπίεσης. Όπως έχει αποδειχθεί το CIR διάγραμμα δεν επηρεάζεται από αλλαγές της υψής. Ελέγχοντας και την θέση του στην δεύτερη ομάδα αυξάνεται η ευστάθεια του κανόνα. Σε διαφορετική περίπτωση, η συσκευή απόφασης θεωρεί το πρόσωπο εισόδου σαν ένα μη αποθηκευμένο πρόσωπο και καλεί μια διαδικασία που αναπρογραμματίζει την βάση δεδομένων αποθηκεύοντας και το νεοεισαχθέν πρόσωπο.

5. ΠΕΙΡΑΜΑΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ

Για τις προσομοιώσεις χρησιμοποιήσαμε τη βάση από πρόσωπα του Πανεπιστημίου του Bern. Τα προς ταυτοποίηση πρόσωπα ελέγχονται από τους τρεις αλγορίθμους ταυτοποίησης ανεξάρτητα, καθώς και από το συνολικό σχήμα. Χρησιμοποιήσαμε 100 πρόσωπα ομαδοποιημένα σύμφωνα με τις διαφορετικές επιδράσεις που έχουν υποστεί. Τα πειραματικά αποτελέσματα που παρουσιάζονται στον Πίνακα Γ5.2, επιδεικνύουν την αποδοτικότητα της χρήσης πολλαπλών αναπαραστάσεων κάθε μια από τις οποίες επικεντρώνεται στην αντιμετώπιση ενός επιμέρους προβλήματος. Μέσω της μηχανής λήψης απόφασης επιτεύχθηκε ένα συνολικό ποσοστό αναγνώρισης της τάξης του 94%.

Διακύμανση	Κλίμακα (%)	Βασικοί άξονες (%)	Φωτισμός (%)	Συνολική Επίδοση (%)
SIR	96	60	50	69
CIR	63	95	55	71
LIR	66	52	99	72
Συνολικό (Μηχανή λήψης απόφασης)				94

Πίνακας Γ5.2: Πειραματικά Αποτελέσματα

6. ΣΥΜΠΕΡΑΣΜΑΤΑ

Ένα σχήμα αναγνώρισης το οποίο δημιουργεί μια δυναμική βάση αποθήκευσης προσώπων και περιλαμβάνει ένα αλγόριθμο ταυτοποίησης κατάλληλο για την αναγνώριση προσώπων σε πραγματικό χρόνο παρουσιάστηκε στο κεφάλαιο αυτό. Οι πολλαπλές και συμπληρωματικές αναπαραστάσεις οι οποίες υιοθετούνται για τον κατακερματισμό του χώρου σφαλμάτων (ο οποίος θεωρείται η ένωση των μεταβολών που προέρχονται από τη διαφορετική κλίμακα των προσώπων, την οριοθέτηση, το γενικό φωτισμό και την διακύμανση της υψής) οδηγεί σε επιτυχημένες παράλληλες λύσεις σε επιμέρους προβλήματα της αναγνώρισης προσώπων. Για το σκοπό αυτό κάθε αναπαράσταση συνοδεύεται από ένα διαφορετικό αλγόριθμο ταυτοποίησης. Η παράλληλη λειτουργία επιτρέπει την υλοποίηση του συνολικού σχήματος σε εφαρμογές πραγματικού χρόνου.

BIBΛΙΟΓΡΑΦΙΑ

- [1]. P. Chellapa, C. Wilson and S. Sirohey, "Human and Machine Recognition of Faces: A Survey," *Proc. IEEE*, vol. 83, no. 5, pp. 705-740, 1995.
- [2]. J. Daugman, "Face and Gesture Recognition: Overview," *IEEE Trans. on PAMI*, vol. 19, no 7, pp 675-676, 1997.
- [3]. A. Doulamis, N. Tsapatsoulis, N. Doulamis and S. Kollias "Innovative Techniques for the Recognition of Faces Based on Multiresolution Analysis and Morphological Filtering," *Proc. of IWISP*, Manchester, November 1996.
- [4]. B. Moghaddam and A. Pentland, "Probabilistic Visual Learning for Object Representation," *IEEE Trans.on PAMI*, vol. 19, pp. 696-710, 1997.
- [5]. N. Tsapatsoulis, K. Karpouzis, G. Votsis and S. Kollias, "Analysis by Synthesis of Facial Images Based on Frontal and Profile Views," *Proc. of IWSNHC3DI Conference*, Rhodes, September 1997.
- [6]. M. Turk and A. Pentland, "Eigenfaces for Recognition," *Journ. Cognitive Neuroscience*, vol. 3, no. 1, 1991.
- [7]. N. Tsapatsoulis, N. Doulamis, A. Doulamis, and S. Kollias, "Face Extraction from Non-uniform Background and Recognition in Compressed Domain," *Proc. of ICASSP'98*, Seattle WA, May 1998.
- [8]. Z. Hong, "Algebraic feature extraction of image for recognition," *Pattern Recognition*, vol.24, pp. 211-219, 1991.
- [9]. L. Chiariglione, "MPEG and Multimedia Communications," *IEEE Trans. on Circuits and Systems for Video Techn.* vol. 7, no. 1, pp. 5-18, Feb. 1997.
- [10]. A. Doulamis, N. Doulamis, and S. Kollias, "Retrainable Neural Networks for Image Analysis and Classification" *Proc. of IEEE Int Conf. on Syst. Man & Cybern.*, Orlando, October 1997.
- [11]. S. Kollias, N. Doulamis, and A. Doulamis, "Improving the Performance of MPEG Compatible Encoders Using on Line Retractable Neural Networks," *Proc. of ICIP'97*, Santa Barbara, October 1997.
- [12]. P. Maragos, "Differential Morphology and Image Processing," *IEEE Trans. Image Processing*, vol. 5, pp. 922-937, June 1996.
- [13]. E. Reusens, T. Ebrahimi, C. Le Buhan, R. Castagno, V. Vaerman, C. de Sola Fabregas, S. Bhattacharjee, F. Bossen, and M. Kunt, "Dynamic Approach to Visual Data Compression". *IEEE Trans. on Cir. & Syst. Fo Video Techn.*, vol. 7, pp. 197-211, February. 1997.
- [14]. N. Tsapatsoulis, N. Doulamis, A. Doulamis, and S. Kollias, "Face Extraction from Non-uniform Background and Recognition in Compressed Domain," *Proc. of ICASSP'98*, Seattle WA, May 1998.
- [15]. G. Cottrell and M. Flemming, "Face recognition using unsupervised feature extraction," *Proc. Int. Conf. Neural Network*, pp 322-325, Paris 1990.
- [16]. S. Lawrence, et. al. "Face Recognition: A Convolutional Neural Network Approach," *IEEE Trans. on Neural Networks*, Vol. 8, No. 1, pp. 98-113, 1997.
- [17]. P. Maragos, "Morphological Signal and Image Processing," *Digital Signal Processing Handbook*, V. Madisetti and D. Williams, eds., IEEE Press.
- [18]. David Vernon, *Machine Vision*, Prentice Hall, 1991.
- [19]. T. Kohonen, "Self-Organisation and Associative Memory," Berlin: Springer, 1988.
- [20]. M. Leonidou, N. Tsapatsoulis and S. Kollias, "An Illumination Invariant Face Recognition Algorithm," in *Proc of the JCIS '98*, North Carolina, USA, October 1998.
- [21]. K. Fukunaga, "An Illumination Invariant Face Recognition Algorithm," 2nd ed, New York: Academic Press, 1990.

Η ΑΝΑΛΥΣΗ ΕΚΦΡΑΣΕΩΝ ΚΑΙ ΣΥΝΑΙΣΘΗΜΑΤΩΝ ΩΣ ΜΕΡΟΣ ΣΥΣΤΗΜΑΤΩΝ ΕΠΙΚΟΙΝΩΝΙΑΣ ΑΝΘΡΩΠΟΥ ΜΗΧΑΝΗΣ

Στο κεφάλαιο αυτό εισάγεται το πρόβλημα της ανάλυσης εκφράσεων και συναισθημάτων, παρατίθενται κάποια ζωτικά θέματα και γίνεται μια χονδρική αναφορά στις πιο δημοφιλείς προσεγγίσεις που έχουν εμφανιστεί στη βιβλιογραφία.

Ένα από τα πιο σημαντικά χαρακτηριστικά το συστημάτων πολυμέσων είναι αλληλεπίδραση που προσφέρουν στο χρήστη. Όλο και περισσότερο απαιτούμε από τα ηλεκτρονικά συστήματα να προσαρμόζονται στις απαιτήσεις του χρήστη και να διαμορφώνουν τη συμπεριφορά τους ανάλογα. Οι υπολογιστές θα γίνουν υπερβολικά ευφυείς όταν αρχίσουν να αντιλαμβάνονται τις προσδοκίες και τους στόχους του χρήστη. Ένας, ανώδυνος για το χρήστη, τρόπος με τον οποίο οι υπολογιστές μπορούν να επιτύχουν την κατανόηση των αντιδράσεων του χρήστη είναι η αποκωδικοποίηση των εκφράσεών του, μια και αυτές φέρουν μεγάλο ποσοστό της πληροφορίας σχετικά με την ικανοποίηση του. Η πληροφορία αυτή μπορεί να χρησιμοποιηθεί για να ερμηνευθεί η ικανοποίηση του χρήστη μετά την τελευταία ενέργεια του ή για να προκαταλάβει την επόμενη.

Ένα από τα πολλά προβλήματα της αναγνώρισης εκφράσεων είναι η διφορούμενη υφή τους η οποία απαιτεί περισσότερες από μια πηγές πληροφόρησης για την ερμηνεία τους. Για παράδειγμα η παρουσία φωνής η οποία συνοδεύει την οπτική διαμόρφωση μιας έκφρασης είναι σίγουρα υποβοηθητικός παράγοντας. Η αναγνώριση εκφράσεων και συναισθημάτων είναι η τρέχουσα πρόκληση στην έρευνα που σχετίζεται με την επικοινωνία ανθρώπου μηχανής. Σχετικά πρόσφατες εργασίες στον τομέα της όρασης υπολογιστών έχουν αφιερώσει μεγάλο ενδιαφέρον στη αναγνώριση και ταξινόμηση των εκφράσεων του προσώπου. Τεχνικές με τις οποίες ένας υπολογιστής μπορεί να αναγνωρίσει τις δράσεις και εκφράσεις σε ένα πρόσωπο μπορούν να χρησιμοποιηθούν σε μια σειρά από εφαρμογές σημαντικότερη από τις οποίες είναι η επικοινωνία ανθρώπου μηχανής.

Οι προσεγγίσεις για την ταξινόμηση εκφράσεων μπορούν να διαιρεθούν σε τρεις βασικές κατηγορίες: στατικές, ημιστατικές και δυναμικές. Οι στατικές προσεγγίσεις χρησιμοποιούν δεδομένα που προέρχονται από στατικές εικόνες στις οποίες οι εκφράσεις απεικονίζονται στην κορύφωση τους -“apex”. Παρόλο που αποτυγχάνουν να παρακολουθήσουν την χρονική εξέλιξη των εκφράσεων, παράγοντας ο οποίος είναι ιδιαίτερα σημαντικός για τον χαρακτηρισμό και την περιγραφή τους, είναι λιγότερο επιρρεπείς σε σφάλματα που προκύπτουν από τις διαδικασίες επεξεργασίας και ανάλυσης εικόνων που εμπλέκονται στο στάδιο προεπεξεργασίας. Οι ημιστατικές τεχνικές χρησιμοποιούν τουλάχιστον δύο εικόνες ή καρέ η μία από τις οποίες απεικονίζει το πρόσωπο σε ουδέτερη έκφραση και η άλλη την έκφραση στην κορύφωση της. Βασίζονται σε τεχνικές εκτίμησης κίνησης [25] και μπορούν να εντοπίσουν τα μέρη του προσώπου που

συμμετέχουν στη διαμόρφωση των εκφράσεων. Παρόλα αυτά δεν μπορούν να απεικονίσουν παράγοντες, όπως η ταχύτητα με την οποία εξελίσσονται οι εκφράσεις, οι οποίοι βοηθούν την διάκριση εκφράσεων οι οποίες εμπλέκουν τα ίδια χαρακτηριστικά προσώπου. Τέλος οι δυναμικές προσεγγίσεις χρησιμοποιούν ολόκληρη την ακολουθία στην οποία απεικονίζεται μια έκφραση και προσπαθούν να συλλάβουν τον τρόπο με τον οποίο εξελίσσεται χρονικά. Το βασικό πλεονέκτημα των δυναμικών μεθόδων είναι η ικανότητα τους να παρακολουθούν τις εκφράσεις σε όλο το χρονικό διάστημα της εξέλιξης τους. Συγκεκριμένα η γνώση της θέσης ενός κινητή σε σχέση με την συνολική ακολουθία είναι απολύτως απαραίτητη. Χωρίς την πληροφορία αυτή δεν μπορούμε να διακρίνουμε τη διαφορά ανάμεσα στην κορύφωση ελαφριάς μορφής εκφράσεων και το αρχικό στάδιο μιας έκφρασης με έντονα χαρακτηριστικά: μπορεί να μπερδέψουμε ένα ελαφρύ χαμόγελο με την αρχή ενός τρανταχτού γέλιου. Η παρατήρηση αυτή θέτει σοβαρούς περιορισμούς για το ποιες εκφράσεις μπορούν να αναγνωριστούν από συγκεκριμένα στιγμιότυπα. Η εξέταση της δυναμικής συμπεριφοράς των εκφράσεων, δηλαδή του τρόπου με τον οποίο αποκαλύπτονται με την πάροδο του χρόνου, μοιάζει απαραίτητη. Η ταχύτητα των κινήσεων κάποιων χαρακτηριστικών του προσώπου σε σχέση με την χρονική τους θέση στη συνολική ακολουθία μπορεί να βελτιώσει την αναγνώριση εκφράσεων ακόμα και σε συστήματα τα οποία δεν απαιτείται να εκτελούνται σε πραγματικό χρόνο. Πειραματικά αποτελέσματα δείχνουν ότι μερικές εκφράσεις μπορεί να αναγνωριστούν αμέσως μετά την έναρξη τους. Αριστέες εκφράσεις ομοιάζουν ως προς την αρχή τους αλλά διαφοροποιούνται στη συνέχεια. Ακόμα και σε αυτή την περίπτωση μπορεί με την παρατήρηση των πρώτων στιγμιότυπων να αποκλειστεί η πιθανότητα ορισμένων εκφράσεων και να τεθεί κάποια υπόθεση όσον αφορά κάποιες άλλες. Με τον τρόπο αυτό το σύστημα μπορεί να επικεντρώσει την προσοχή του σε ένα υποσύνολο των πιθανών αποκρίσεων που μπορεί να δώσει: λιγότερες υποθέσεις μπορούν να εξεταστούν περισσότερο λεπτομερώς και να οδηγήσουν σε πιο ακριβείς απαντήσεις – αντιδράσεις. Η δημιουργία κάποιας υπόθεσης στα πρώτα στάδια της έκφρασης μπορεί να είναι πολύ σημαντική σε συστήματα που απαιτείται να εκτελούνται σε πραγματικό χρόνο δεδομένου ότι άλλες ενέργειες περιμένουν τον χαρακτηρισμό της έκφρασης για να δράσουν ανάλογα: πρόβλεψη της επόμενης ενέργειας του χρήστη, επιλογή της αντίδρασης του υπολογιστή κ.ο.κ. Η χρονική εξέλιξη μπορεί επίσης να είναι εξαιρετικής σημασίας για την αναγνώριση ενδιάμεσων ή ελαφριάς μορφής εκφράσεων. Από τη άλλη πλευρά ένα σημαντικό μειονέκτημα που αντιμετωπίζουν η δυναμικές προσεγγίσεις είναι η έλλειψη διαθέσιμων βάσεων από εκφράσεις προσώπων και τα προβλήματα παρακολούθησης της κίνησης των σημείων στην επιφάνεια του προσώπου.

ΤΑΞΙΝΟΜΗΣΗ ΣΤΑΤΙΚΩΝ ΕΙΚΟΝΩΝ ΑΝΘΡΩΠΙΝΩΝ ΕΚΦΡΑΣΕΩΝ

Στο κεφάλαιο αυτό συγκρίνεται η απόδοση και η ικανότητα γενίκευσης διαφόρων σχημάτων για την ταξινόμηση συναισθημάτων με βάση στατικές εικόνες που δείχνουν τις εκφράσεις “χαρά”, “οργή” και “λύπη” καθώς και την ουδέτερη κατάσταση. Συγκρίνονται τρεις γενικές στρατηγικές: Η πρώτη προσέγγιση χρησιμοποιεί πρότυπα εκφράσεων για την εύρεση της βέλτιστης ταύτισης. Η δεύτερη χρησιμοποιεί συνδυασμό των αναλύσεων PCA και FLD για τον χαρακτηρισμό των εκφράσεων με χαμηλότερου επιπέδου αναπαράστασεις. Τέλος στην τρίτη χρησιμοποιείται ένα νευρωνικό δίκτυο multilayer perceptron εκπαιδευμένο με τον αλγόριθμο backpropagation.

1. ΕΙΣΑΓΩΓΗ

Τα τελευταία χρόνια πληθώρα αρχιτεκτονικών και αλγορίθμων για την αναγνώριση προσώπων και εκφράσεων έχουν καταγραφεί στη βιβλιογραφία της επιστήμης των υπολογιστών και της όρασης υπολογιστών. Ανασκόπηση της σχετικής βιβλιογραφίας έχει παρουσιαστεί στο πρώτο μέρος της διατριβής. Μια κατηγοριοποίηση, σε προσεγγίσεις βασισμένες σε (α) χαρακτηριστικά και (β) πρότυπα, περιγράφεται στην εργασία των Brunelli και Poggio [11], αλλά πειράματα από τον τομέα της Ψυχολογίας συγκλίνουν στο συμπέρασμα ότι το οπτικό σύστημα των ανθρώπων επεξεργάζεται τα πρόσωπα –τουλάχιστον μέχρι ενός σημείου- ολιστικά [12], γεγονός που καθιστά της τεχνικές βασισμένες σε πρότυπα περισσότερο τεκμηριωμένες τουλάχιστον όσον αφορά την βιολογική τους ορθότητα. Με βάση την τελευταία παρατήρηση στο τρέχον κεφάλαιο πραγματοποιούμε μια συγκριτική μελέτη των τεχνικών που χρησιμοποιούν πρότυπα και οι οποίες διαχωρίζονται περαιτέρω σε επιβλεπόμενες (MLP) και μη επιβλεπόμενες (ταύτιση με πρότυπο, PCA-FLD).

2. ΤΑ ΠΡΟΤΥΠΑ ΜΕΣΗΣ ΤΙΜΗΣ

Για τη διερεύνηση της ορθής ταξινόμησης και της ικανότητας γενίκευσης των διαφόρων μεθόδων χρησιμοποιήσαμε την βάση εικόνων CMU για την πραγματοποίηση μιας σειράς από πειράματα. Οι εικόνες αντιστοιχούν σε 20 άτομα (άνδρες και γυναίκες), έχουν χωρική ανάλυση 120x128 και απεικονίζουν τις εκφράσεις “χαρά”, “λύπη”, “ουδέτερη” και “οργή”. Για τον εντοπισμό των προσώπων στις εικόνες εφαρμόστηκε η διαδικασία η περιγράφηκε στα Κεφάλαια B1 και B2. Μετά τον εντοπισμό των προσώπων πραγματοποιήθηκε κανονικοποίηση ως προς την κλίμακα και τη θέση των ματιών. Η χωρική ανάλυση των κανονικοποιημένων προσώπων είναι 35x37 και χρησιμοποιήθηκαν στα πειράματα συνολικά 77 εικόνες. Σε τέσσερις περιπτώσεις υπήρξε αποτυχία εντοπισμού του προσώπου.

Το πρώτο πείραμα αφορούσε τον έλεγχο της τεχνικής ταύτισης με πρότυπο για την ταξινόμηση στατικών εικόνων με εκφράσεις. Για τη δημιουργία των προτύπων κάθε έκφρασης χρησιμοποιήθηκαν συνολικά 64

φωτογραφίες (16x4=16 εικόνες για κάθε μια από τις 4 εκφράσεις). Το πρότυπο δημιουργήθηκε αθροίζοντας τις εικόνες pixel προς pixel και λαμβάνοντας την μέση τιμή. Η Εικόνα Δ2.1 επιδεικνύει τα τέσσερα πρότυπα που αντιστοιχούν στις εκφράσεις “ουδέτερη”, “οργή”, “χαρά” και “λύπη” αντίστοιχα.



Εικόνα Δ2.1: Πρότυπα των εκφράσεων “ουδέτερη”, “οργή”, “χαρά” και “λύπη” τα οποία λήφθηκαν με χρήση του μέσου όρου σε ένα σύνολο από 16 εικόνες για κάθε έκφραση.

Παρόλο που τα πρότυπα είναι αναγνωρίσιμα από τον άνθρωπο όσον αφορά στην έκφραση που αναπαριστούν, ορθή ταξινόμηση επιτεύχθηκε μόνο στο 75% των εικόνων του συνόλου μάθησης και στο 50% των εικόνων του συνόλου ελέγχου –εικόνες που δεν χρησιμοποιήθηκαν για τη δημιουργία των προτύπων. Οι βασικότερες αιτίες για το χαμηλό βαθμό γενίκευσης ήταν οι μικρές διακυμάνσεις όσον αφορά τη κλίση, και θέση του προσώπου, οι οποίες δεν κατέστη δυνατόν να απαλειφθούν στη διαδικασία της κανονικοποίησης, καθώς και η ιδιομορφία στον τρόπο έκφρασης των συναισθημάτων από διαφορετικά άτομα.

Η μετρική που χρησιμοποιήθηκε στη διαδικασία ταύτισης είναι η $M3 = \frac{1}{k \cdot l} \sum \left| \frac{R1}{\text{mean}(R1)} - \frac{R2}{\text{mean}(R2)} \right|$, η

οποία όπως αναφέρθηκε στο πρώτο μέρος της διατριβής μοιάζει να είναι η καταλληλότερη για συγκρίσεις μονόχρωμων εικόνων.

3. Η ΜΕΘΟΔΟΣ FLD

Η χρήση της ανάλυσης – προβολής στις βασικές συντεταγμένες έχει χρησιμοποιηθεί κατά κόρο στην αναγνώριση προσώπων [12]-[18] και συνοψίζεται στο Κεφάλαιο Γ1. Η κεντρική ιδέα πίσω από την τεχνική αυτή είναι η εξαγωγή της βασικής πληροφορίας από το σύνολο μάθησης όπως αυτή παρουσιάζεται στις εικόνες. Οι εικόνες εισόδου προβάλλονται στους βασικούς άξονες οι οποίοι εκφράζουν την μέγιστη εντροπία του συνόλου μάθησης. Όπως αναφέρθηκε στο Κεφάλαιο Γ1 η ανάλυση PCA δεν μπορεί να θεωρηθεί βέλτιστη ως προς τη διαχωριστικότητα αλλά μπορεί πάντοτε να χρησιμοποιηθεί για δημιουργία συμπαγών αναπαραστάσεων. Στην μελέτη [15] επιδείχθηκε ότι η επιλογή των k βασικών αξόνων είναι βέλτιστη ως προς την κατηγοριοποίηση φυσικών κατηγοριών προσώπων, όπως το φύλο, αλλά όχι για την αναγνώριση προσώπων ή εκφράσεων οι οποίες στηρίζονται σε περισσότερο λεπτομερή στοιχεία του προσώπου.

Στο Κεφάλαιο Γ1 έγινε αναφορά στη μέθοδο FLD η οποία επιδιώκει να βρει τους άξονες εκείνους που μεγιστοποιούν ταυτόχρονα την ολική διασπορά του συνόλου μάθησης και τη διασπορά ανάμεσα στις κλάσεις όπως αυτές εκφράζονται από τα κέντρα τους. Στην περίπτωση της ταξινόμησης εκφράσεων οι κλάσεις γενικά είναι λίγες και για κάθε μια από αυτές υπάρχει ικανός αριθμός εικόνων. Επομένως η

εκτίμηση των πινάκων συμμεταβλητότητας C_i για κάθε κλάση δεν είναι επιρρεπής σε σφάλματα. Παραμένει όμως πάντα το πρόβλημα της μεγάλης διάστασης του πίνακα $\Sigma_w^{-1} \cdot \Sigma_b$. Για να ξεπεράσουμε και το πρόβλημα αυτό εφαρμόσαμε πρώτα την τεχνική PCA για να δημιουργήσουμε μια συμπαγή αναπαράσταση των εικόνων μέσω των συντελεστών του μετασχηματισμού KL και στη συνέχεια εφαρμόσαμε τη μέθοδο FLD για την επιλογή των χαρακτηριστικών –αξόνων- που συμβάλλουν περισσότερο στην διαχωριστικότητα των εκφράσεων.

3.1 Ανάλυση PCA για τη δημιουργία συμπαγούς αναπαράστασης

Η αποδοτική χρήση της τεχνικής PCA απαιτεί πολύ καλή κανονικοποίηση των προσώπων και ευθυγράμμιση των βασικών τους στοιχείων [13], διαδικασία που είχε πραγματοποιηθεί ήδη στην τεχνική ταύτισης με πρότυπο. Η διαδικασία δημιουργίας συμπαγούς αναπαράστασης των προσώπων με την τεχνική PCA συνοψίζεται παρακάτω:

Έστω το σύνολο μάθησης αποτελούμενο από N εικόνες $\{I_1, I_2, \dots, I_N\}$ όπου $I_i \in R^{n \times m}$, με εκφράσεις “ουδέτερη”, “οργή”, “χαρά” και “λύπη”, και $\{v_1, v_2, \dots, v_N\}$ με $v_i \in R^L$ και $L = n \cdot m$, τα αντίστοιχα διανύσματα. Οι συναρτήσεις βάσης του KLT υπολογίζονται από τη λύση του προβλήματος ιδιοτιμών:

$$\Lambda = \Phi^T \cdot \Sigma \cdot \Phi \quad (\Delta 2.1)$$

όπου Σ είναι ο πίνακας συμμεταβλητότητας, Φ είναι ο πίνακας των ιδιοδιανυσμάτων του Σ , και Λ είναι ο αντίστοιχος διαγώνιος πίνακας των ιδιοτιμών.

Για τη δημιουργία ενός διανύσματος χαρακτηριστικών $\hat{x} = [\hat{x}_1, \hat{x}_2, \dots, \hat{x}_m]^T$ για κάθε διάνυσμα v_i πραγματοποιείται προβολή στους m βασικούς άξονες –ιδιοδιανύσματα που αντιστοιχούν στις m μεγαλύτερες ιδιοτιμές- του μετασχηματισμού KLT:

$$\hat{x}_{(i)} = \Phi_k^T \cdot \hat{v}_i \quad (\Delta 2.2)$$

όπου $\hat{v}_i = v_i - \mu$, $\mu = \sum_{j=1}^N v_j$ και Φ_k είναι το τμήμα εκείνο του πίνακα των ιδιοδιανυσμάτων που αντιστοιχούν στις k μεγαλύτερες ιδιοτιμές.

Στην Εικόνα Δ2.2 παρουσιάζονται τα 36 βασικά ιδιοδιανύσματα όπως προέκυψαν από την παραπάνω διαδικασία. Όπως μπορεί κάποιος να παρατηρήσει μεγαλύτερη διακύμανση, ως προς τις εκφράσεις, παρουσιάζουν τα διανύσματα 12 έως 18 και όχι κάποια από τα πρώτα δέκα, γεγονός που ενισχύει την άποψη ότι η βέλτιστη ταξινόμηση δεν επιτυγχάνεται με προβολή στους βασικούς άξονες.



Εικόνα Δ2.2: Τα 36 ιδιοδιανύσματα που αντιστοιχούν στις μεγαλύτερες ιδιοτιμές υπολογισμένα σε ένα σύνολο από 64 φωτογραφίες προσώπων που δείχνουν τις εκφράσεις “ουδέτερη”, “οργή”, “χαρά” και “λύπη”.

3.2 Επιλογή των χαρακτηριστικών με τη μεγαλύτερη ικανότητα διαχωρισμού

Μετά την εφαρμογή της ανάλυσης PCA κάθε εικόνα αναπαρίσταται από το διάνυσμα $\hat{x}_{(i)} = [\hat{x}_{i1}, \hat{x}_{i2}, \dots, \hat{x}_{im}]^T$. Έστω οι κλάσεις $\omega_1, \omega_2, \dots, \omega_L$, με πληθικούς αριθμούς N_1, N_2, \dots, N_L αντίστοιχα, των παραπάνω διανυσμάτων. Έστω επίσης m_1, m_2, \dots, m_L τα μέσα διανύσματα των αναπαραστάσεων των κλάσεων αυτών και M το ολικό μέσο διάνυσμα, δηλαδή:

$$m_i = \frac{1}{N_i} \sum_{k=1}^{N_i} \hat{x}_{(k)}^{(i)}, \quad M = \frac{1}{L} \sum_{i=1}^L m_i \quad (\Delta 2.3)$$

όπου $\hat{x}_{(k)}^{(i)}$ το k -στο διάνυσμα της κλάσης ω_i

Στην περίπτωση που εξετάζουμε είναι $L=4$ –κλάσεις που αντιστοιχούν στις εκφράσεις “ουδέτερη”, “οργή”, “χαρά” και “λύπη”- και $N_1 = N_2 = \dots = N_L = 16$. Οι εντός κλάσης και ανάμεσα στις κλάσεις πίνακες διασποράς Σ_ω και Σ_b υπολογίζονται από τις σχέσεις:

$$\Sigma_\omega = \sum_{i=1}^L p(\omega_i) \cdot C_i = \sum_{i=1}^L p(\omega_i) \cdot E\{[\hat{x}^{(i)} - m_i] \cdot [\hat{x}^{(i)} - m_i]^T | \omega_i\} \quad (\Delta 2.4)$$

$$\Sigma_b = \sum_{i=1}^L p(\omega_i) \cdot (m_i - M) \cdot (m_i - M)^T \quad (\Delta 2.5)$$

Επειδή δεν υπάρχει η κάποια πληροφορία για την κατανομή της a priori πιθανότητας $p(\omega_i)$ τη θεωρούμε ομοιόμορφη, που ισοδυναμεί με ισοπίθανες κλάσεις. Οι πίνακες C_i εκτιμώνται από τη σχέση:

$$C_i = \frac{1}{N_i} \sum_{k=1}^{N_i} (\hat{x}_{(k)}^{(i)} - m_i) \cdot (\hat{x}_{(k)}^{(i)} - m_i)^T \quad (\Delta 2.6)$$

Ο πίνακας Ψ , τα διανύσματα του οποίου αποτελούν μια βάση –όχι ορθοκανονική– του χώρου αποτελείται από τα ιδιοδιανύσματα του πίνακα $\Sigma_{\omega}^{-1} \cdot \Sigma_b$ μετά από την επίλυση του προβλήματος ιδιοτιμών:

$$\Sigma_{\omega}^{-1} \cdot \Sigma_b \cdot \Psi = \Psi \cdot \Delta \quad (\Delta 2.7)$$

όπου οι πίνακες $\Psi, \Delta \in \mathfrak{R}^{m \times m}$.

Η τελική αναπαράσταση που χρησιμοποιείται για την ταξινόμηση δίνεται από τη σχέση:

$$y = \Psi^T \cdot \hat{x} \quad \text{ή} \quad y = \Psi_k^T \cdot \hat{x} \quad (\Delta 2.8)$$

όπου $\Psi_k = [\psi_1, \psi_2, \dots, \psi_k]$ ο πίνακας με τα ιδιοδιανύσματα που αντιστοιχούν στις k ($k < m$) μεγαλύτερες ιδιοτιμές του πίνακα $\Sigma_{\omega}^{-1} \cdot \Sigma_b$. Το κριτήριο ταύτισης μπορεί να είναι απλά η απόσταση:

$$k_0 = \arg \min_i \|y - \Psi_k^T \cdot m_i\| \quad (\Delta 2.9)$$

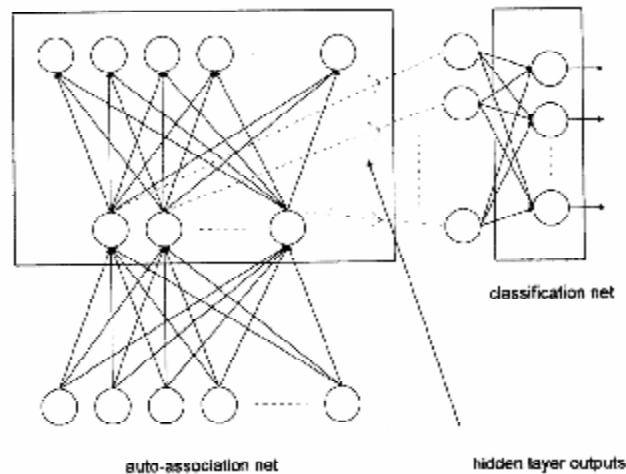
ή κάποια περισσότερο σύνθετη τεχνική.

Το τελικό αποτέλεσμα της χρήσης του συνδυασμού FLD και PCA ήταν 100% επιτυχία στην ταξινόμηση των εικόνων του συνόλου μάθησης αλλά σχετικά φτωχή ικανότητα γενίκευσης –ορθή ταξινόμηση εικόνων εκτός του συνόλου μάθησης– η οποία κυμάνθηκε περίπου στο 60%. Αυτό μπορεί να αποδοθεί σε όχι καλή εκτίμηση των πινάκων συμμεταβλητότητας των κλάσεων και σε προβλήματα μη ορθής ευθυγράμμισης των βασικών χαρακτηριστικών του προσώπου τα οποία έχουν τεράστια σημασία στην ταξινόμηση εκφράσεων.

4. ΤΕΧΝΙΚΕΣ ΕΠΙΒΛΕΠΟΜΕΝΗΣ ΜΑΘΗΣΗΣ

Η χρήση τεχνικών επιβλεπόμενης μάθησης μέσω νευρωνικών δικτύων multilayer perceptron (MLP) για αναγνώριση προσώπων και ταξινόμηση φυσικών κατηγοριών έχει υιοθετηθεί σε αρκετές εργασίες [19][20]. Η κεντρική ιδέα είναι η χρήση ενός ανατροφοδοτούμενου –feedforward– νευρωνικού δικτύου με ένα ή περισσότερα εσωτερικά επίπεδα και με αριθμό εξόδων ίσο με τον αριθμό των κλάσεων. Στη περίπτωση των φυσικών κατηγοριών ο αριθμός των κλάσεων είναι μικρός και εκ των προτέρων γνωστός. Αντίθετα σε συστήματα αναγνώρισης προσώπων ο αριθμός των κλάσεων είναι μεγάλος και μεταβαλλόμενος –αφού υπάρχει πάντα η περίπτωση αποθήκευσης νέων προσώπων στη βάση. Στην περίπτωση της ταξινόμησης εκφράσεων ο αριθμός των κλάσεων είναι περιορισμένος και η χρήση νευρωνικών δικτύων δόκιμη. Ένα πρόβλημα που υπάρχει πάντα όσον αφορά την ταξινόμηση εικόνων με χρήση νευρωνικών δικτύων είναι η μεγάλη διάσταση του χώρου εισόδου, γεγονός που επιφέρει δυσκολίες στη διαδικασία μάθησης και φτωχά αποτελέσματα όσον αφορά τη γενίκευση. Μια λύση που προτείνεται για το σκοπό αυτό είναι η χρήση δύο νευρωνικών δικτύων, όπως φαίνεται στο Σχήμα Δ2.1, που λειτουργούν το πρώτο ως αυτοσυσχετιστική μνήμη και το δεύτερο ως ταξινομητής. Το δίκτυο της αυτοσυσχετιστικής μνήμης αποτελείται από n εισόδους και n εξόδους και περιέχει $p \ll n$ εσωτερικούς νευρώνες. Κατά τη διάρκεια της εκπαίδευσης το

δίκτυο μαθαίνει να δημιουργεί στην έξοδο την καλύτερη δυνατή προσέγγιση του διανύσματος εισόδου. Με τον τρόπο αυτό η έξοδος των νευρώνων του εσωτερικού επιπέδου δημιουργεί μια συμπαγή αναπαράσταση του διανύσματος εισόδου η οποία μπορεί να χρησιμοποιηθεί ως είσοδος του δικτύου ταξινόμησης. Το θέμα είναι τι αναπαριστά η αναπαράσταση αυτή και αν συνεισφέρει στη διαχωριστικότητα των εκφράσεων. Σύμφωνα με την μελέτη [30] στην καλύτερη περίπτωση –όταν το δίκτυο τη αυτοσυσχετιστικής μνήμης είναι γραμμικό- η αναπαράσταση στην έξοδο των νευρώνων του εσωτερικού δικτύου είναι όμοια με την προβολή στους p βασικούς άξονες της τεχνικής PCA.



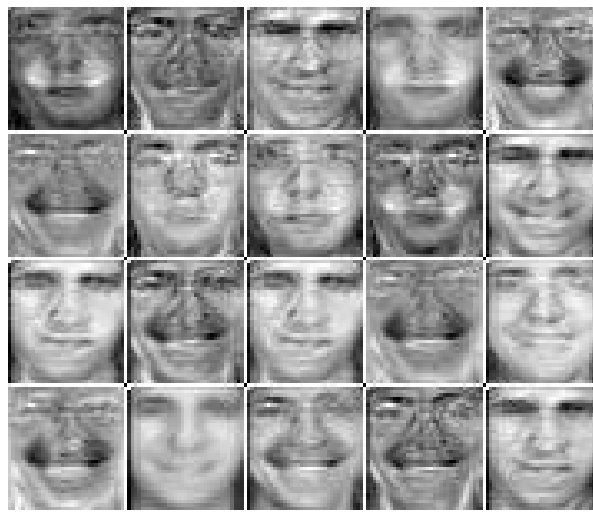
Σχήμα Δ2.1: Εξαγωγή χαρακτηριστικών για περιγραφή εκφράσεων με χρήση δύο νευρωνικών δικτύων

Δεδομένου ότι οι εικόνες που απεικονίζουν εκφράσεις στην περίπτωση που εξετάσαμε ήταν μικρής χωρικής ανάλυσης -35x37- μπορούσαμε να εφαρμόσουμε ταξινόμηση είτε με βάση τις τιμές των pixels των εικόνων είτε χρησιμοποιώντας τις αναπαραστάσεις $\hat{x}_{(i)} = [\hat{x}_{i1}, \hat{x}_{i2}, \dots, \hat{x}_{im}]^T$ που προήλθαν μετά από την ανάλυση PCA. Το επίπεδο εξόδου των δικτύων αποτελείτο και στις δύο περιπτώσεις από τέσσερις νευρώνες και αριθμός των νευρώνων στο εσωτερικό επίπεδο ποικίλει. Για την εκπαίδευση χρησιμοποιήθηκε ο αλγόριθμος μάθησης backpropagation [23].

4.1 Ταξινόμηση μέσω του MLP δικτύου εφαρμοσμένη άμεσα στα δεδομένα εισόδου

Το δίκτυο με τα καλύτερα αποτελέσματα προήλθε με χρήση 20 νευρώνων στο εσωτερικό επίπεδο. Η συνολική δομή του δικτύου ήταν 35x37 είσοδοι, 20 νευρώνες στο εσωτερικό επίπεδο και 4 έξοδοι. Η ικανότητα ταξινόμησης των εικόνων του συνόλου μάθησης προσέγγισε το 100% ενώ η ικανότητα γενίκευσης έφτασε στην καλύτερη περίπτωση το 75%. Δεδομένου ότι το επίπεδο τυχαιότητας αντιστοιχεί στο 25%, η ταξινόμηση μπορεί να θεωρηθεί ικανοποιητική αν ληφθούν υπόψη οι παραλλαγές στον τρόπο έκφρασης των συναισθημάτων από διαφορετικούς ανθρώπους και τα προβλήματα ανεπιτυχούς κανονικοποίησης. Πάντως με τυχαία επιλογή του συνόλου επαλήθευσης η γενίκευση περιορίζεται στο 40-60%. Στην Εικόνα Δ2.3 επιδεικνύονται τα βάρη των 20 νευρώνων του εσωτερικού επιπέδου. Μερικές από

τις αναπαραστάσεις είναι αντεστραμμένες, όσον αφορά τη φωτεινότητα, για καλύτερη απεικόνιση των εξαγομένων χαρακτηριστικών. Συγκρίνοντας τις Εικόνες Δ2.2 και Δ2.3 παρατηρούμε ότι το εσωτερικό επίπεδο του δικτύου MLP επικεντρώνεται περισσότερο στα χαρακτηριστικά που διαμορφώνουν τις εκφράσεις και λιγότερο στην γενικότερη πληροφορία του προσώπου. Για παράδειγμα το γεγονός αυτό γίνεται αντιληπτό από τις φωτεινές και σκούρες σιές γύρω από την περιοχή των ρυτίδων του στόματος οι οποίες καταδεικνύουν τη σημαντικότητα του χαρακτηριστικού αυτού τη διαμόρφωση κάποιων εκφράσεων (βλέπε τις εικόνες της πρώτης γραμμής και την τρίτη και τέταρτη εικόνα της δεύτερης γραμμής). Ένα άλλο προφανές χαρακτηριστικό είναι η θέση των φρυδιών η οποία χαρακτηρίζει σε μεγάλο βαθμό την έκφραση “οργή” (βλέπε τρίτη εικόνα της πρώτης γραμμής και τελευταία εικόνα της πέμπτης γραμμής).



Εικόνα Δ2.3: Τα βάρη για τους 20 νευρώνες του εσωτερικού –hidden- επιπέδου ενός MLP-δικτύου μεγέθους 1295x20x4 για την αναγνώριση των εκφράσεων “ουδέτερη”, “οργή”, “χαρά” και “λύπη”.

4.2 Συμπύεση και αναπαράσταση

Σε ένα δεύτερο πείραμα μειώσαμε τον αριθμό των νευρώνων στο εσωτερικό επίπεδο σε 5. Δεδομένου ότι το δίκτυο είχε 100% ορθή ταξινόμηση των εικόνων, επιτυγχάνεται βαθμός συμπίεσης $64/5 = 12.8$ στο εσωτερικό επίπεδο του δικτύου. Η ικανότητα γενίκευσης ήταν αντίστοιχη με αυτό του δικτύου με 20 νευρώνες αλλά στην περίπτωση αυτή το δίκτυο αναγκάζεται να επικεντρωθεί στα πιο εκφραστικά χαρακτηριστικά όπως εμφανίζεται στην Εικόνα Δ2.4. Ο τρίτος και ο τέταρτος νευρώνας ομοιάζουν με ανιχνευτή της θέσης των φρυδιών, περιοχή η οποία θεωρείται από τις πιο σημαντικές στην διαμόρφωση των εκφράσεων. Εξετάζοντας λεπτομερέστερα τους νευρώνες τρία και τέσσερα παρατηρούμε στον τρίτο νευρώνα μια μετατόπιση αμφοτέρων των φρυδιών προς τα κάτω και προς τα μέσα σε σχέση με το μέσο πρόσωπο, κίνηση που χαρακτηρίζει την έκφραση “οργή”, και στον τέταρτο νευρώνα μετατόπιση των φρυδιών προς τα άνω, κίνηση που χαρακτηρίζει την έκφραση “χαρά”. Ο πρώτος και ο πέμπτος νευρώνας

επικεντρώνονται στην περιοχή του στόματος και των ρυτίδων που δημιουργούνται στην περιοχή αυτή κατά την διαμόρφωση εκφράσεων όπως είναι οι “λύπη” και “χαρά”. Οι ίδιοι νευρώνες μοιάζει να μετρούν επίσης και την κύρτωση των χειλιών. Τέλος η στροφή που εμφανίζεται στον δεύτερο νευρώνα απεικονίζει την τάση της πλειονότητα των εικόνων του συνόλου μάθησης να παρουσιάζουν μια κλίση ως προς τον άξονα z λόγω αποτυχίας στη διαδικασία ευθυγράμμισης των βασικών χαρακτηριστικών του προσώπου. Η παρουσία του νευρώνα αποδεικνύει πόσο σημαντική είναι η σωστή κανονικοποίηση και ευθυγράμμιση των εικόνων και πως αποτυχία της διαδικασίας αυτή επηρεάζει άμεσα την ικανότητα γενίκευσης του δικτύου.



Εικόνα Δ2.4: Τα βάρη για τους 5 νευρώνες του εσωτερικού –hidden- επιπέδου ενός MLP-δικτύου μεγέθους $1295 \times 5 \times 4$ για την αναγνώριση των εκφράσεων “ουδέτερη”, “οργή”, “χαρά” και “λύπη”.

4.3 Ταξινόμηση των αναπαραστάσεων της PCA ανάλυσης

Η διαφορά στην περίπτωση αυτή ήταν η μεγαλύτερη ταχύτητα στην εκπαίδευση του δικτύου λόγω του περιορισμένου αριθμού εισόδων –δίκτυο $36 \times 12 \times 4$. Το δίκτυο με τα καλύτερα αποτελέσματα προήλθε με χρήση 12 νευρώνων στο εσωτερικό επίπεδο. Η βασικός στόχος του δικτύου ήταν η επιλογή των αξόνων εκείνων της PCA αναπαράστασης οι οποίοι παρουσιάζουν την καλύτερη ικανότητα διαχωρισμού των κλάσεων και υπό αυτή την έννοια μοιάζει με τη μέθοδο FLD. Ορθή ταξινόμηση του συνόλου μάθησης πραγματοποιείται και με τη χρήση λιγότερων από 36 στοιχείων για τα διανύσματα εισόδου –οι συντελεστές που αντιστοιχούν στις 10 βασικότερους άξονες είναι αρκετοί- αλλά για καλύτερη γενίκευση χρειάζεται ένα μεγαλύτερο διάνυσμα αναπαράστασης. Η βέλτιστη ικανότητα γενίκευσης πλησίασε το 65%, κάπως καλύτερη από την αντίστοιχη της μεθόδου FLD. Η διαφορά μπορεί να αποδοθεί στον επιβλεπόμενο τρόπο μάθησης.

5. ΣΥΜΠΕΡΑΣΜΑΤΑ

Στις προηγούμενες παραγράφους διερευνήθηκε και συγκρίθηκε η ικανότητα ορθής ταξινόμησης και γενίκευσης στατικών εικόνων που απεικονίζουν εκφράσεις για διάφορες επιβλεπόμενες και μη τεχνικές. Η τεχνική ταύτισης με πρότυπο παρουσιάζει χαμηλή επίδοση εξαιτίας (α) της απαλοιφής της λεπτομέρειας η οποία είναι απαραίτητη για το διαχωρισμό εκφράσεων και (β) ανεπιτυχούς ευθυγράμμισης των βασικών χαρακτηριστικών του προσώπου στη διαδικασία κανονικοποίησης. Επιπλέον σε πολλές από τις εικόνες η έκφραση δεν ήταν άμεσα αναγνωρίσιμη ούτε και από τους ανθρώπους. Η PCA-αναπαράσταση είχε εξαιρετική επιτυχία στην ταξινόμηση και στην ανακατασκευή των εικόνων του συνόλου μάθησης αλλά μικρή ικανότητα γενίκευσης εξαιτίας κυρίως της ιδιομορφίας στον τρόπο έκφραση των συναισθημάτων από τους ανθρώπους. Το δίκτυο MLP εκπαιδευμένο με τον αλγόριθμο backpropagation παρουσίασε τέλεια

ικανότητα ταξινόμησης των εικόνων του συνόλου μάθησης και αποδεικτή ικανότητα γενίκευσης, στη περίπτωση που οι εικόνες του συνόλου επαλήθευσης δεν πολύ διαφορετικές από αυτές του συνόλου μάθησης. Με βάση τα ανωτέρω προκύπτει ότι η μη γραμμική εξαγωγή χαρακτηριστικών, η οποία επιτυγχάνεται με ένα δίκτυο multilayer perceptron, μπορεί να μεγιστοποιήσει την ορθή ταξινόμηση στο σύνολο μάθησης αλλά η ικανότητα γενίκευσης εξαρτάται σε μεγάλο βαθμό από την άψογη κανονικοποίηση των εικόνων, την εκφραστικότητα των συναισθημάτων μέσω παρατηρήσιμων μεταβολών στο πρόσωπο και την σχετικά υψηλή χωρική ανάλυση η οποία απαιτείται για να συλλάβει τις μικρές λεπτομέρειες που διαφοροποιούν τις εκφράσεις.

Συνοπτικά οι παράγοντες που επηρεάζουν την αποδοτικότητα των αλγορίθμων ταξινόμησης εκφράσεων που στηρίζονται σε ολικές αναπαραστάσεις και που παρουσιάστηκαν στο κεφάλαιο αυτό είναι:

1) Περιορισμοί στο σύνολο δεδομένων:

(a). Χαμηλή ανάλυση απεικονιζόμενων προσώπων - 35x37 pixel

(b). Μικρός
αριθμός απεικονιζόμενων εκφράσεων (“λύπη”, “οργή” και “χαρά”)

(c). Οι
εκφράσεις στις φωτογραφίες είναι δύσκολα αναγνωρίσιμες ακόμα και από τον άνθρωπο

2) Περιορισμοί στη διαδικασία κανονικοποίησης των προσώπων

3) Χρήση στατικών εικόνων μόνο

Για τη βελτίωση της ικανότητας γενίκευσης μπορούν να γίνει αρκετές τροποποιήσεις: (α) Εισαγωγή ενός δεύτερου σταδίου κανονικοποίησης που να περιλαμβάνει και την ευθυγράμμιση του στόματος, (β) Η χρήση δυναμικής πληροφορίας όπως η εξέλιξη των εκφράσεων με την πάροδο του χρόνου μέσω της παρακολούθησης των διανυσμάτων κίνησης [24][25] ή της διαφοράς των καρέ [26] οδηγούν σε μεγαλύτερη ικανότητα γενίκευσης με κόστος την αύξηση της υπολογιστικής πολυπλοκότητας. Τέλος ο συνδυασμός πολλών πηγών εξαγωγής πληροφορίας όπως η δράση των μυών [27, 28] και η χρήση πληροφορίας φωνής [29] είναι ίσως ο μοναδικός τρόπος για επίτευξη ικανοποιητικών αποτελεσμάτων όσον αφορά την ορθή ταξινόμηση εκφράσεων.

ΤΑΞΙΝΟΜΗΣΗ ΣΥΝΑΙΣΘΗΜΑΤΩΝ ΜΕ ΒΑΣΗ ΤΑ ΣΥΝΟΛΑ FDP ΚΑΙ FAP ΤΟΥ MPEG-4 ΜΕΣΩ ΑΣΑΦΟΥΣ ΛΟΓΙΚΗΣ

Στο κεφάλαιο αυτό προτείνεται μια περιγραφή των έξι πρότυπων εκφράσεων με τη χρήση των συνόλων Facial Definition Parameter Set (FDP) και Facial Animation Parameter Set (FAP) [1] του MPEG-4. Σύμφωνα με το πρότυπο MPEG-4, το σύνολο FAP είναι ένα σύνολο από σύμβολα – τεκμήρια τα οποία περιγράφουν τις ελάχιστες δυνατές αντιληπτές μεταβολές που λαμβάνουν χώρα στη περιοχή του προσώπου. Ομάδες τέτοιων διαταραχών – δράσεων δημιουργούν την αντίληψη κάποιας έκφρασης [24]. Μια συστηματική προσέγγιση προς την κατεύθυνση της αναγνώρισης και ταξινόμησης των εκφράσεων μπορεί να βασιστεί στην αυτόματη ανίχνευση και παρακολούθηση μερικών χαρακτηριστικών σημείων της περιοχής του προσώπου τα οποία αποτελούν υποσύνολο του συνόλου FDP. Μετρήσεις βασισμένες στα σημεία αυτά, και οι οποίες εκφράζουν την μετατόπιση των σημείων αυτών ανάμεσα σε δύο καρέ, τροφοδοτούν ένα σύστημα ασφούς λογικής του οποίου η έξοδος αποτελείται από ένα σύνολο παραμέτρων οι οποίοι εκφράζουν το βαθμό με τον οποίο η παρατηρηθείσα έκφραση ομοιάζει σε κάθε μια από τις πρότυπες εκφράσεις. Εκτός από την ανάλυση των πρότυπων εκφράσεων προχωράμε ένα βήμα παρακάτω επιχειρώντας να διευρύνουμε τον αριθμό των εκφράσεων οι οποίες μπορούν να αναγνωριστούν. Τροποποιώντας της συναρτήσεις συμμετοχής των εμπλεκόμενων στην αναγνώριση χαρακτηριστικών σύμφωνα με την παράμετρο «ενεργοποίηση» [3] δημιουργούμε ένα αποδοτικό τρόπο για την αναγνώριση παραλλαγών των έξι βασικών εκφράσεων.

1. ΕΙΣΑΓΩΓΗ

Η έρευνα στην ανάλυση των συναισθημάτων του ανθρώπου, μέσω οπτικών στοιχείων, έχει επικεντρωθεί στα πρωτεύοντα ή πρότυπα συναισθήματα τα οποία είναι άμεσα συσχετισμένα με διακεκριμένες εκφράσεις [4]. Πολύ λίγες μελέτες [5] οι οποίες εξετάζουν μη πρότυπα συναισθήματα έχουν εμφανιστεί στη βιβλιογραφία της επιστήμης των υπολογιστών. Αντίθετα ερευνητές από το πεδίο της Ψυχολογίας έχουν διερευνήσει εκτεταμένα [3][6] ένα ευρύτερο φάσμα συναισθημάτων. Παρόλο που η εκμετάλλευση των αποτελεσμάτων που προήλθαν από τους Ψυχολόγους είναι πολύ δύσκολο να γίνει οι ερευνητές της επιστήμης των υπολογιστών μπορούν να χρησιμοποιήσουν κάποια στοιχεία στις μελέτες τους. Ο Whissel [3] ισχυρίζεται ότι τα συναισθήματα είναι σημεία σε ένα χώρο μικρής διάστασης ο οποίος σε μια πρώτη προσέγγιση μπορεί να θεωρηθεί διδιάστατος με άξονες την *ενεργοποίηση* - *activation* και *επαλήθευση* - *evaluation*. Από την οπτική της πρακτικής εφαρμογής, η *επαλήθευση* εκφράζει μάλλον κάποια εσωτερικά αισθήματα και σιέψεις του υποκειμένου οι οποίες προφανώς δεν μπορούν να ανιχνευθούν από τη μορφοποίηση του προσώπου. Από τη άλλη πλευρά η *ενεργοποίηση* σχετίζεται άμεσα με την ολική κίνηση του κεφαλιού καθώς και με τη δραστηριοποίηση των μυών του προσώπου. Η Εικόνα Δ2.1 και ο Πίνακας Δ2.1 παρέχουν μια ένδειξη για τη σχέση ανάμεσα στις διαταραχές του προσώπου, όπως αυτές μπορούν να προσεγγιστούν από την μετακίνηση κάποιων από τα σημεία στην επιφάνεια του, και του άξονα της *ενεργοποίησης* σύμφωνα με τον Whissel (η *ενεργοποίηση* που σχετίζεται με τον όρο “delighted” είναι 4.2 ενώ για τον όρο “eager” είναι 5).

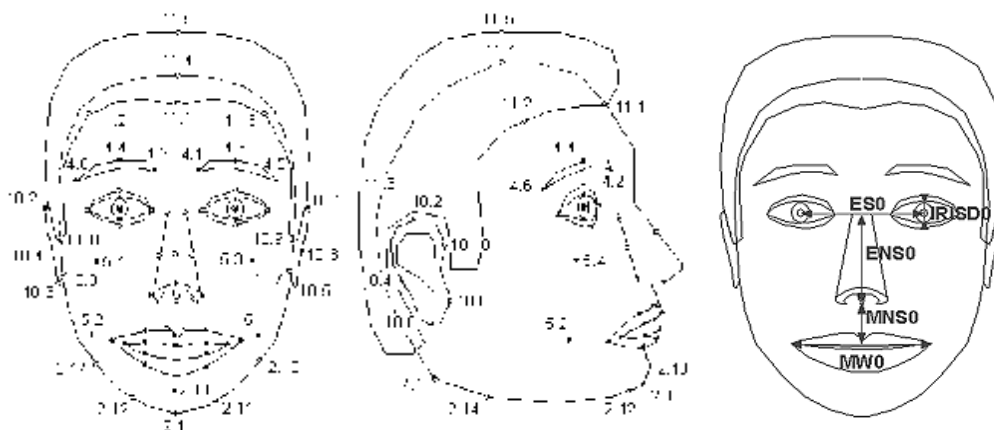
Η καθιέρωση των προτύπων MPEG και ειδικά του MPEG-4 υποδεικνύει κάποιους εναλλακτικούς τρόπου για την μοντελοποίηση των εκφράσεων προσώπου και των συναισθημάτων που σχετίζονται με αυτές. Οι παράμετροι απόδοσης κίνησης -Facial Animation Parameters (FAPs)- και προσδιορισμού του προσώπου - Facial Definition Parameter Set (FDP)- χρησιμοποιούνται στο πλαίσιο του MPEG-4 για τη δημιουργία μοντέλων προσώπου και απόδοσης κίνησης και μορφοποιήσεων σε αυτό. Στο κεφάλαιο αυτό προτείνουμε ένα τρόπο με τον οποίο τα σύνολα FAP και FDP μπορούν να χρησιμοποιηθούν και για την αντίστροφη διαδικασία: αυτή της αναγνώρισης εκφράσεων. Η αυτόματη ανίχνευση και παρακολούθηση των σημείων που ορίζονται στο σύνολο FDP, σε μια ακολουθία βίντεο είναι μια περιοχή ενεργής ερευνητικής δραστηριότητας. Περισσότερα για το συγκεκριμένο θέμα αναφέρονται στο Κεφάλαιο Β1. Η χρήση της πληροφορίας κίνησης αυτών των σημείων μπορεί να ενσωματωθεί στο πρότυπο MPEG-4 για την ανάλυση και κωδικοποίηση των εκφράσεων του προσώπου.



(α)

(β)

Εικόνα Δ3.1: Οπτική αναπαράσταση των συναισθημάτων “απόλαυση-delighted” και “ενθουσιασμός-eager”



Εικόνα Δ3.2: (α) Τα σημεία του συνόλου FDP στο τρισδιάστατο χώρο (β) Οι μονάδες των FAPs (The Facial Animation Parameter Units -ES = ESo/1000, NS = ENSo/1000, MNS = MNSo/1000, MW = MWo/1000) [1]

Η συνέχεια του χώρου των συναισθημάτων, η αβεβαιότητα που προκύπτει από τη διαδικασία της αυτόματης ανίχνευσης των σημείων που ορίζονται στο σύνολο FDP και η οποία επηρεάζει τις τιμές των χαρακτηριστικών που χρησιμοποιούνται για την ταξινόμηση τους καθώς και η δυσκολία ερμηνείας των συναισθημάτων με χρήση μόνο οπτικών στοιχείων καθιστούν την ασαφή λογική κατάλληλη για την απεικόνιση των χαρακτηριστικών στο χώρο των συναισθημάτων. Τέλος η ερευνητική εμπειρία των επιστημόνων στον τομέα της Ψυχολογίας, όπως αυτή εκφράζεται μέσα από την παράμετρο της *ενεργοποίησης*, μπορεί να ενσωματωθεί στο προτεινόμενο σχήμα για την ανάλυση παραλλαγών των πρότυπων συναισθημάτων [3].

	Activation	Evaluation		Activation	Evaluation
Afraid	4.9	3.4	Angry	4.2	2.7
Bashful	2	2.7	Delighted	4.2	6.4
Disgusted	5	3.2	Eager	5	5.1
Guilty	4	1.1	Joyful	5.4	6.1
Patient	3.3	3.8	Sad	3.8	2.4
Surprised	6.5	5.2			

Πίνακας Δ3.1: Επιλεγμένοι όροι – συναισθήματα από τις μελέτες [3] και [6]

2. ΣΥΝΟΛΑ ΠΑΡΑΜΕΤΡΩΝ ΓΙΑ ΤΟΝ ΟΡΙΣΜΟ ΠΡΟΣΩΠΩΝ ΚΑΙ ΤΗΣ ΚΙΝΗΣΗΣ ΣΕ ΑΥΤΑ

Το σύνολο παραμέτρων για τον ορισμό ενός προσώπου (Facial Definition Parameter set FPD) και το σύνολο παραμέτρων αναπαράστασης προσώπου (Facial Animation Parameter set FAP) έχουν σχεδιαστεί για να επιτρέπουν τον ορισμό του σχήματος και της υψής ενός προσώπου, την αναπαράσταση εκφράσεων, συναισθημάτων καθώς και της προφοράς ομιλίας. Το σύνολο FAP εάν ερμηνευτεί σωστά παράγει παρόμοια αποτελέσματα με το πραγματικό πρότυπο σε ότι αφορά εκφράσεις, προφορά ομιλίας σε διαφορετικά μοντέλα προσώπων, χωρίς να χρειάζεται να αρχικοποιήσουμε η να ρυθμίσουμε το μοντέλο. Το σύνολο παραμέτρων για τον ορισμό προσώπου (FDPs) χρησιμοποιείται ώστε να αντιστοιχίζει ένα δοθέν μοντέλο προσώπου σε ένα συγκεκριμένο πρόσωπο. Οι παράμετροι ορισμού προσώπου συνήθως αποστέλλονται μια φορά και ακολουθούνται από ένα «κύμα» παραμέτρων αναπαράστασης προσώπου. Εντούτοις ακόμα και αν ο αποκωδικοποιητής δεν λάβει τις παραμέτρους ορισμού προσώπου (FDPs) η χρήση των FAPUs εγγυάται ότι θα ερμηνευτούν σωστά οι παράμετροι αναπαράστασης προσώπου (FAP).

Το σύνολο παραμέτρων απόδοσης κίνησης προσώπου (FAPs) είναι βασισμένο στην μελέτη των ελάχιστων δράσεων του προσώπου και είναι στενά συνδεδεμένο με τις δράσεις των μυών. Αποτελούν ένα πλήρες σύνολο από βασικές δράσεις του προσώπου και ως εκ τούτου επιτρέπουν την αναπαράσταση των πιο φυσικών μορφοποιήσεων του. Πολύ υψηλές τιμές βοηθούν στον ορισμό δράσεων που δεν είναι φυσιολογικές για το ανθρώπινο πρόσωπο αλλά μπορούν να βρουν χρησιμότητα στον σχεδιασμό μοντέλων που χρησιμοποιούνται σε περιβάλλοντα εικονικής πραγματικότητας. Όλες οι παράμετροι που αφορούν την απόδοση κίνησης εκφράζονται σύμφωνα με τις μονάδες παραμέτρων απόδοσης κίνησης (Facial Animation Parameters Units). Αυτές οι μονάδες έχουν σχεδιαστεί έτσι ώστε να επιτρέπουν την απόδοση των παραμέτρων αναπαράστασης σε κάθε μοντέλο προσώπου με έναν συνεχή τρόπο παράγοντας λογικά

αποτελέσματα σε ότι αφορά τις εκφράσεις και την ανθρώπινη προφορά. Ανταποκρίνονται στον τεμαχισμό των αποστάσεων ανάμεσα σε κάποια χαρακτηριστικά σημεία κλειδιά του ανθρώπινου προσώπου.

Το σύνολο FDP επιτρέπει τον ορισμό ενός ακριβούς σχήματος προσώπου καθώς και της υψής του στην φάση της αρχικής μορφοποίησης και αναπαράστασης. Εάν αυτά τα στοιχεία χρησιμοποιηθούν στην φάση της αρχικοποίησης της διαδικασίας επικοινωνίας δύο μερών αναπαράγουν, στη πλευρά της εφαρμογής πελάτη, με ακόμα μεγαλύτερη ακρίβεια τις κινήσεις του συγκεκριμένου μοντέλου προσώπου.

Οι εκφράσεις προσώπων και τα συναισθήματα που αντιστοιχούν σε αυτές μπορούν να περιγραφούν ως ένα σύνολο μετρήσεων, που αφορούν τις παραμέτρους FDP και παραγόμενα από αυτές χαρακτηριστικά, καθώς και δράσεων όπως αυτή εκφράζεται από τον ορισμό των FAPs. Μερικές από τις παραμέτρους FAP, οριζόμενες πάνω σε ένα υποσύνολο των FDP μπορούν να χρησιμοποιηθούν για την απεικόνιση μετρήσεων που προκύπτουν από τη διαμόρφωση της επιφάνειας του προσώπου όπως αυτή καταγράφεται σε μια ακολουθία βίντεο και με τη βοήθειά τους να αναπαραχθεί η κίνηση της έκφρασης και ενδεχομένως να προσεγγιστεί το αντίστοιχο συναισθημα.

3. ΣΥΣΧΕΤΙΣΗ ΤΩΝ FDPs ΚΑΙ ΤΩΝ ΠΡΟΤΥΠΩΝ ΕΚΦΡΑΣΕΩΝ

Παρόλο που η κίνηση των μυών [4] είναι ιδιαίτερα σημαντική, όσον αφορά την απόδοση κίνησης και δημιουργία μορφοποιήσεων στο πρόσωπο, είναι πρακτικά αδύνατο να την παρακολουθήσει κάποιος χωρίς τη βοήθεια ηλεκτρομαγνητικών αισθητήρων. Εντούτοις ένα ποσοστό της κίνησης τους μπορεί να εξαχθεί με την παρατήρηση των διαμορφώσεων των ιστών του δέρματος του προσώπου όπως αυτή καταγράφεται από την κίνηση μερικών σημείων στην επιφάνεια του. Η λογική αυτή προσομοιάζει τον τρόπο τον οποίο οι άνθρωποι αντιλαμβάνονται τα συναισθήματα με χρήση της οπτικής πληροφορίας και συγκεκριμένα παρατηρώντας τις πιο εκφραστικές περιοχές του προσώπου όπως η περιοχή γύρω από τα μάτια και το στόμα. Τα έξι πρότυπα συναισθήματα, όπως και παραλλαγές τους, εμπλέκουν διαμορφώσεις του προσώπου που περιλαμβάνουν την κίνηση χαρακτηριστικών σημείων σε αυτό. Τα σημεία αυτά σχετίζονται άμεσα με ένα μικρό υποσύνολο του συνόλου FDP και υπό κάποιες προϋποθέσεις μπορούν να εντοπιστούν και να παρακολουθηθούν αυτόματα.

Στον Πίνακα Δ3.2 επιδεικνύεται ένας τρόπος περιγραφής των πρωτεύουσών εκφράσεων σύμφωνα με την ορολογία των FAPs [1]. Η περιγραφή στηρίχθηκε σε δύο μελέτες από τον τομέα της Νευροφυσιολογίας [2][3], οι οποίες αναφέρονται στις μορφοποιήσεις του προσώπου κατά τη διάρκεια μιας έκφρασης, καθώς και από πειραματικά δεδομένα από τις βάσεις προσώπων του Ekman –η οποία περιέχει μόνο στατικές εικόνες- και του MediaLab –ακολουθίες βίντεο που αναπαριστούν εκφράσεις.

Ο Πίνακας Δ3.3 δείχνει ποιοι παράμετροι FDPs εμπλέκονται στην προσέγγιση των FAPs καθώς επίσης και ποια είναι τα πραγματικά χαρακτηριστικά τα οποία χρησιμοποιούνται για την περιγραφή τους. Η συσχέτιση των χρησιμοποιούμενων υποσυνόλων FAP και FDP επιτυγχάνεται κυρίως μέσω αποστάσεων ανάμεσα σε παραμέτρους του συνόλου FDP. Οι χρονικές παράγωγοι των αποστάσεων εξυπηρετούν δύο σκοπούς: ορίζουν τη θετική τιμή των FAPs και περιγράφουν την εξέλιξη των εκφράσεων και ανιχνεύουν την

κορύφωση -“αρεχ”- τους. Η μεταβλητή $f_{i,NEUTRAL}$ αναφέρεται στη συγκεκριμένη απόσταση όταν το πρόσωπο είναι σε ουδέτερη κατάσταση –δεν υπάρχει έκφραση.

«οργή»	squeeze_l_eyebrow (+) lower_t_midlip(-) raise_l_i_eyebrow (+) close_t_r_eyelid (-) close_b_r_eyelid (-)	squeeze_r_eyebrow (+) raise_b_midlip (+) raise_r_i_eyebrow (+) close_t_l_eyelid (-) close_b_l_eyelid (-)
«λύπη»	raise_l_i_eyebrow (+) close_t_l_eyelid (+) raise_l_m_eyebrow (-) raise_l_o_eyebrow (-) close_b_l_eyelid (+)	raise_r_i_eyebrow (+) close_t_r_eyelid (+) raise_r_m_eyebrow (-) raise_r_o_eyebrow (-) close_b_r_eyelid (+)
«έκπληξη»	raise_l_o_eyebrow (+) raise_l_i_eyebrow (+) raise_l_m_eyebrow (+) squeeze_l_eyebrow (-) open_jaw (+)	raise_r_o_eyebrow (+) raise_r_i_eyebrow (+) raise_r_m_eyebrow (+) squeeze_r_eyebrow (-)
«χαρά»	close_t_l_eyelid (+) close_b_l_eyelid (+) stretch_l_cornerlip (+) raise_l_m_eyebrow (+) lift_r_cheek (+) lower_t_midlip (-) OR open_jaw (+)	close_t_r_eyelid (+) close_b_r_eyelid (+) stretch_r_cornerlip (+) raise_r_m_eyebrow (+) lift_l_cheek (+) raise_b_midlip (-)
«απέχθεια»	close_t_l_eyelid (+) close_t_r_eyelid (+) lower_t_midlip (-) squeeze_l_cornerlip (+)	close_b_l_eyelid (+) close_b_r_eyelid (+) open_jaw (+) squeeze_r_cornerlip (+)
«φόβος»	raise_l_o_eyebrow (+) raise_l_m_eyebrow (+) raise_l_i_eyebrow (+) squeeze_l_eyebrow (+) open_jaw (+) OR close_t_l_eyelid (-) lower_t_midlip (-) OR lower_t_midlip (+)	raise_r_o_eyebrow (+) raise_r_m_eyebrow (+) raise_r_i_eyebrow (+) squeeze_r_eyebrow (+) close_t_r_eyelid (-)

Πίνακας Δ3.2: Οι παράμετροι FAPs που εμπλέκονται στις έξι πρώτες εκφράσεις

Όνομα FAP	Χαρακτηριστικά περιγραφής	Θετικό πρόσημο
squeeze_l_eyebrow	$f_1 = s(1,3) / F_1 = f_1 - f_{1-NEUTRAL}$	$F_1 < 0$
squeeze_r_eyebrow	$f_2 = s(4,6) / F_2 = f_2 - f_{2-NEUTRAL}$	$F_2 < 0$
lower_t_midlip	$f_3 = s(16,30) / F_3 = f_3 - f_{3-NEUTRAL}$	$F_3 < 0$
raise_b_midlip	$f_4 = s(16,33) / F_4 = f_4 - f_{4-NEUTRAL}$	$F_4 < 0$
raise_l_l_eyebrow	$f_5 = s(3,8) / F_5 = f_5 - f_{5-NEUTRAL}$	$F_5 > 0$
raise_r_l_eyebrow	$f_6 = s(6,12) / F_6 = f_6 - f_{6-NEUTRAL}$	$F_6 > 0$
raise_l_o_eyebrow	$f_7 = s(1,7) / F_7 = f_7 - f_{7-NEUTRAL}$	$F_7 > 0$
raise_r_o_eyebrow	$f_8 = s(4,11) / F_8 = f_8 - f_{8-NEUTRAL}$	$F_8 > 0$
raise_l_m_eyebrow	$f_9 = s(2,7) / F_9 = f_9 - f_{9-NEUTRAL}$	$F_9 > 0$
raise_r_m_eyebrow	$f_{10} = s(5,11) / F_{10} = f_{10} - f_{10-NEUTRAL}$	$F_{10} > 0$
open_jaw	$f_{12} = s(30,33) / F_{11} = f_{11} - f_{11-NEUTRAL}$	$F_{11} > 0$
close_t_l_eyelid –close_b_l_eyelid	$f_{12} = s(9,10) / F_{12} = f_{12} - f_{12-NEUTRAL}$	$F_{12} < 0$
close_t_r_eyelid –close_b_r_eyelid	$f_{13} = s(13,14) / F_{13} = f_{13} - f_{13-NEUTRAL}$	$F_{13} < 0$
stretch_l_cornerlip –stretch_r_cornerlip	$f_{14} = s(28,29) / F_{14} = f_{14} - f_{14-NEUTRAL}$	$F_{14} > 0$
Squeeze_l_eyebrow & squeeze_r_eyebrow	$f_{15} = s(3,6) / F_{15} = f_{15} - f_{15-NEUTRAL}$	$F_{15} < 0$

Πίνακας Δ3.3(α): Περιγραφή των FAPs με χρήση ενός υποσυνόλου των FDPs. Σημείωση: $s(i,j)$ είναι η Ευκλείδεια απόσταση ανάμεσα στα FDPs i και j

Όνομα	Χαρακτηριστικά περιγραφής	Θετικό πρόσημο
Mouth_corners_down	$f_{16} = s([16,17],15) / F_{16} = f_{16} - f_{16-NEUTRAL}$	$F_{16} > 0$
Mouth_vertical_asymmetry	$f_{17} = slope(16,17) - f_{21}$	$f_{17} \neq 0$
Mouth_occlusion (detection failure)	$f_{18} = boolean\ value$	
Head_up/down	$f_{19} = \frac{ENSo}{ESo} / F_{19} = f_{19} - f_{19-NEUTRAL}$	$F_{19} < 0$
Head_rotation	$f_{20} = \frac{ESo}{ENSo} / F_{20} = f_{20} - f_{20-NEUTRAL}$	$F_{20} < 0$
Head_tilt	$f_{21} = slope([7,8],[11,12])$	$f_{21} \neq 0$

Πίνακας Δ3.3(β): Πρόσθετα χαρακτηριστικά –μη ορισμένα από το MPEG-4- για την περιγραφή εκφράσεων

3.1 Ανίχνευση των χαρακτηριστικών σημείων προσώπου

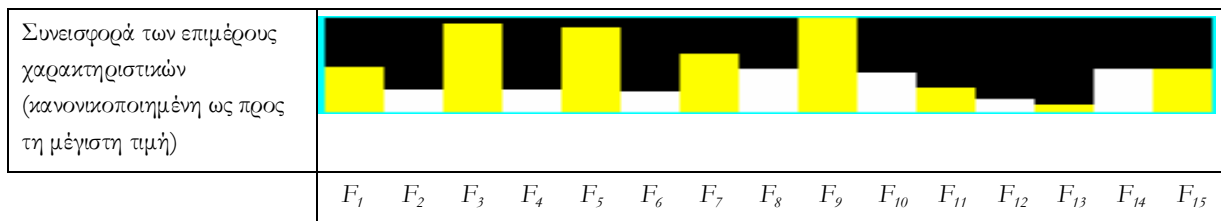
Η ανίχνευση των σημείων του προσώπου μέσω των οποίων ορίζονται οι παράμετροι FAPs που χρησιμοποιούνται βασίστηκε στη μεθοδολογία που περιγράφεται στο Κεφάλαιο Β1 και στην εργασία [8]. Παρόλα αυτά για επακριβή εντοπισμό πολλές φορές χρειάστηκε η παρέμβαση του χρήστη.

3.2 Αξιολόγηση του διανύσματος χαρακτηριστικών

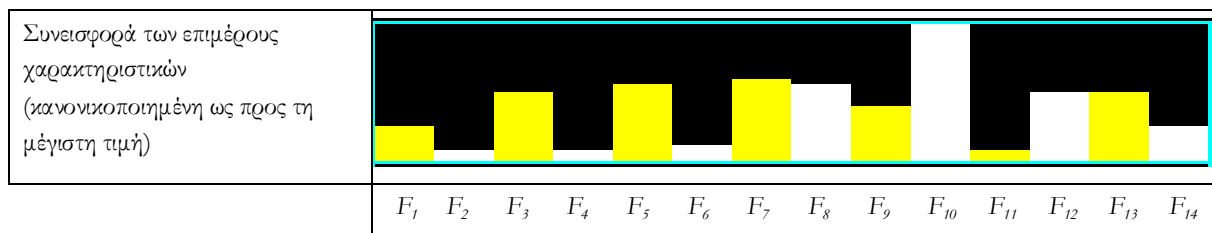
Ένα κρίσιμο θέμα σχετικά με τα χαρακτηριστικά του Πίνακα Δ3.3 είναι η ικανότητα τους να περιγράψουν τα αντίστοιχα FAPs, η επιμέρους συνεισφορά τους στην ταξινόμηση των πρωτεύουσών εκφράσεων καθώς η ικανότητα τους να διαχωρίζουν ομοειδείς εκφράσεις. Για την αξιολόγηση της ανωτέρω συμπεριφοράς των χαρακτηριστικών εκτελέσαμε δύο πειράματα: Χρησιμοποιήσαμε τις ακολουθίες εκφράσεων του MediaLab, οι οποίες αναφέρονται στις εκφράσεις «χαρά», «έκπληξη», «οργή» και «απέχθεια», και υπολογίσαμε το διάνυσμα χαρακτηριστικών του Πίνακα Δ3.3(α) για κάθε καρέ. Η ανίχνευση των σημείων FDP έγινε με χρήση των τεχνικών που περιγράφηκαν στο Κεφάλαιο Β1 αλλά στις περισσότερες περιπτώσεις η παρέμβαση του χρήστη για ακριβή εντοπισμό ήταν απαραίτητη. Το διάνυσμα των χαρακτηριστικών οδηγήθηκε σε μια δομή νευρωνικού δικτύου το οποίο εκπαιδεύτηκε για την ταξινόμηση των καρέ στις επιμέρους εκφράσεις. Το πρόγραμμα που χρησιμοποιήθηκε για την εκπαίδευση του νευρωνικού δικτύου (NeuralWare) μας έδωσε την δυνατότητα να αποκρυπτογραφήσουμε την συνεισφορά καθενός από τα 15 χαρακτηριστικά όσον αφορά την ταξινόμηση των καρέ. Τα αποτελέσματα επιδεικνύονται στο Σχήμα Δ3.1. Παρατηρούμε ότι για την ορθή ταξινόμηση των παραπάνω εκφράσεων ουσιαστική συνεισφορά είχαν 8 από τα 15 χαρακτηριστικά χωρίς απαραίτητα ο βαθμός συνεισφοράς τους να είναι ίδιος για κάθε έκφραση. Για παράδειγμα οι κάθετες ρυτίδες ανάμεσα στα φρύδια –χαρακτηριστικό F_{15} - ήταν εξαιρετικά σημαντικό στοιχείο για τον χαρακτηρισμό της έκφρασης «οργή» αλλά δεν συμμετείχε στην ουσιαστικά στην απόφαση για κάποιες άλλες εκφράσεις. Κάποια χαρακτηριστικά ιδιαίτερα επιρρεπή σε σφάλματα όσον αφορά τον ακριβή εντοπισμό των σημείων FDP, όπως τα *open_eyelid* και *close_eyelid*, ουσιαστικά παραβλέπονται από τον ταξινομητή. Από την πλευρά της ανάλυσης εκφράσεων τα χαρακτηριστικά αυτά μπορούν να αγνοηθούν αλλά για σκοπούς σύνθεσης και απόδοσης κίνησης –animation- η αξία τους είναι υψηλή. Τέλος από το Σχήμα Δ3.1 παρατηρείται ένας πλεονασμός όσον αφορά τα συμμετρικά χαρακτηριστικά –πχ *raise_l_i_eyebrow* και *raise_r_i_eyebrow*. Με δεδομένο ότι δεν είναι όλες οι πρωτεύουσες εκφράσεις συμμετρικές όσον αφορά την κίνηση των βασικών χαρακτηριστικών του προσώπου, το γεγονός αυτό είναι μάλλον περιεργό. Διατηρώντας και τα δύο συμμετρικά χαρακτηριστικά αυξάνουμε την ευρωστία του συστήματος και αποφεύγουμε σφάλματα σε εκφράσεις με μη συμμετρική μορφοποίηση –όπως η έκφραση «οργή». Δεδομένου ότι επικεντρωθήκαμε στην συνεισφορά των επιμέρους χαρακτηριστικών παρά στην συνολική επίδοση της διαδικασίας ταξινόμησης τα απόλυτα ποσοστά για κάθε κατηγορία έχουν ελάχιστη σημασία. Ούτως ή άλλως, όπως επιδεικνύεται στον Πίνακα Δ3.2, για ικανοποιητικά ποσοστά ταξινόμησης απαιτείται υψηλότερου επιπέδου συνδυασμός των χαρακτηριστικών καθώς και η χρήση λεκτικών κανόνων.

Στο δεύτερο πείραμα εξετάσαμε την ικανότητα του διανύσματος χαρακτηριστικών να διαχωρίζει μη πρωτεύουσες εκφράσεις οι οποίες έχουν γενικά κοινή βάση. Συγκεκριμένα χρησιμοποιήσαμε εικόνες με τις

εκφράσεις «ευθυμία», «χαρά» και «ενθουσιασμός» οι οποίες προέρχονται από την πιλοτική βάση του προγράμματος TMR PHYSTA και από τις ακολουθίες του MediaLab. Με τον ίδιο τρόπο όπως στο πρώτο πείραμα εκπαιδεύσαμε ένα νευρωνικό δίκτυο για την ταξινόμηση των εικόνων στις τρεις παραπάνω κατηγορίες. Η συνεισφορά των επιμέρους χαρακτηριστικών επιδεικνύεται στο Σχήμα Δ3.2. Όπως μπορεί εύκολα κάποιος να παρατηρήσει τα χαρακτηριστικά που σχετίζονται με την κίνηση των φρυδιών έχουν αυξημένη συμμετοχή στην διαδικασία ταξινόμησης. Το γεγονός αυτό υποδηλώνει ότι με κατάλληλη απόδοση βαρύτητας στα επιμέρους χαρακτηριστικά είναι η εφικτή η διάκριση ανάμεσα σε ομοιογενή χαρακτηριστικά.



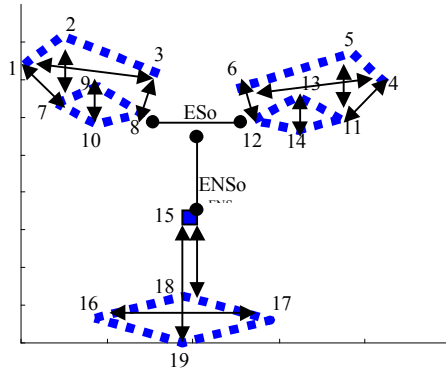
Σχήμα Δ3.1: Συνεισφορά των επιμέρους χαρακτηριστικών για την ταξινόμηση των πρωτευουσών εκφράσεων.



Σχήμα Δ3.2: Συνεισφορά των επιμέρους χαρακτηριστικών για την ταξινόμηση των εκφράσεων «ευθυμία», «χαρά» και «ενθουσιασμός».

4. ΤΟ ΣΥΣΤΗΜΑ ΑΣΑΦΟΥΣ ΛΟΓΙΚΗΣ

Η χονδρική δομή του συστήματος ασαφούς λογικής που χρησιμοποιείται επιδεικνύεται στο Σχήμα Δ3.4. Για κάθε εικόνα ή καρτέ που εμφανίζει κάποιο πρόσωπο σε μια δεδομένη συναισθηματική κατάσταση υπολογίζεται ένα διάνυσμα δεκαπέντε στοιχείων, τα οποία αντιστοιχούν στα FAPs του Πίνακα Δ3.3, και στη συνέχεια οδηγείται στην είσοδο του συστήματος ασαφούς λογικής. Το διάνυσμα εισόδου ασαφροποιείται με την τεχνική singleton. Για την μαθηματική θεμελίωση της ανωτέρω τεχνικής θεωρούμε $X_i, i=1, \dots, 15$ τα πεδία ορισμού των μεταβλητών $F_i, X = X_1 \times X_2 \times \dots \times X_{15}$ το καρτεσιανό γινόμενο τους και $\mathfrak{I}(X)$ το δυναμοσύνολο του συνόλου X . Στην ασαφροποίηση singleton ένα διάνυσμα $x_0 \in X$ αντιστοιχίζεται με ένα ασαφές σύνολο $A'(x) \in \mathfrak{I}(X)$ για το οποίο ισχύει $A'(x_0) = 1$ και $A'(x) = 0$ για $x \neq x_0$.



Σχήμα Δ3.3: Τα 19 FDP σημεία που χρησιμοποιούνται για την περιγραφή των FAPs του Πίνακα Δ3.3

Η έξοδος είναι ένα διάνυσμα n στοιχείων όπου n είναι ο αριθμός των διακεκριμένων εκφράσεων στις οποίες επιθυμούμε να κατατάξουμε το διάνυσμα εισόδου. Στη περίπτωση των πρότυπων συναισθημάτων κάθε τιμή του διανύσματος εξόδου εκφράζει το βαθμό συγγένειας της εισόδου με τα συναισθήματα θυμός, λύπη, χαρά, αγδία, φόβος και έκπληξη. Στο πεδίο ορισμού κάθε διακεκριμένου στοιχείου του διανύσματος ορίζεται μια ασαφής διαμέριση η οποία μπορεί να συσχετιστεί άμεσα με γλωσσικούς όρους όπως *λίγο*, *πολύ*, *μέτρια* κοκ. Οι γλωσσικοί όροι των ασαφών διαμερίσεων (πχ. *medium open jam*) συνδέονται μέσω κανόνων της μορφής IF-THEN οι οποίοι έχουν κατασκευαστεί ευριστικά με τη βοήθεια των Πινάκων Δ3.2 και Δ3.4. Οι Πίνακες Δ3.2 και Δ3.4 εκφράζουν την εκ των προτέρων γνώση του συστήματος.



Σχήμα Δ3.4: Η δομή του ασαφούς συστήματος

4.1 Ασαφής διαμέριση των διαστημάτων διακύμανσης των στοιχείων της εισόδου

Για να ορίσουμε τις διαμερίσεις χρειάζεται πρώτα να ορίσουμε τα διαστήματα διακύμανσης $X_i, i=1, \dots, 15$ των μεταβλητών F_i . Στην εργασία [1] προτείνεται ένα διάστημα διακύμανσης για όλα τα FAPs αλλά είναι πολύ γενικά για να μπορούν να χρησιμοποιηθούν σε συστήματα ανάλυσης εκφράσεων. Αντίθετα είναι αποδεκτά για σκοπούς σύνθεσης εκφράσεων και περιβάλλοντα εικονικής πραγματικότητας. Για την εκτίμηση των πεδίων ορισμού X_i χρησιμοποιήσαμε τα στατιστικά στοιχεία του Πίνακα Δ3.4 τα οποία προέκυψαν από μια στατιστική μελέτη σε δύο βάσεις από πρότυπες εκφράσεις: τις στατικές εικόνες του Ekman και τις ακολουθίες βίντεο του MediaLab. Αμφότερες περιέχουν ακριβείς αναπαραστάσεις των

πρότυπων εκφράσεων παρά τυπικές καθημερινές εκφράσεις αλλά το γεγονός αυτό τις καθιστά χρήσιμες στον ορισμό των διαστημάτων διακύμανσης. Η στρατηγική που υιοθετήθηκε είναι η παρακάτω:

Έστω m_{ij} και σ_{ij} η μέση τιμή και η τυπική απόκλιση αντίστοιχα, της μεταβλητής F_j στην έκφραση i (όπου $i = \{1 \Rightarrow \text{«οργή»}, 2 \Rightarrow \text{«λύπη»}, 3 \Rightarrow \text{«χαρά»}, 4 \Rightarrow \text{«απέχθεια»}, 5 \Rightarrow \text{«φόβος»}, 6 \Rightarrow \text{«έκπληξη»}\}$) όπως αυτή εκτιμήθηκε στο Πίνακα 4. Το διάστημα διακύμανσης X_j για κάθε δικατευθυντική παράμετρο FAP –βλέπε εργασία [1]- ορίζεται ως:

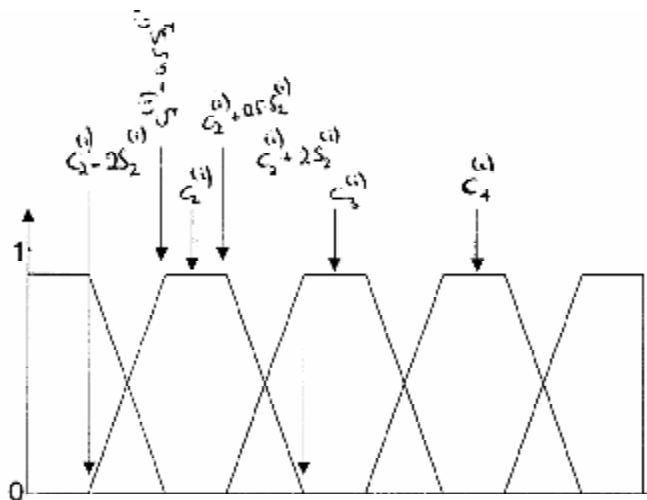
$$X_j = \left[\min_i \{m_{ij} - 3 \cdot \sigma_{ij}\} \quad \max_i \{m_{ij} + 3 \cdot \sigma_{ij}\} \right] \quad (\Delta 3.1)$$

Για τις μονοκατευθυντικές παραμέτρους FAP (όπως η F_{11}) ισχύει:

$$X_j = \left[\min_i \{m_{ij} - 3 \cdot \sigma_{ij}\} \quad 0 \right] \quad \text{ή} \quad X_j = \left[0 \quad \max_i \{m_{ij} + 3 \cdot \sigma_{ij}\} \right] \quad (\Delta 3.2)$$

ανάλογα με το αν αυτές είναι άνω ή κάτω φραγμένες αντίστοιχα.

Οι Πίνακες Δ3.2 και Δ3.4 χρησιμοποιήθηκαν και για τον ορισμό των διαμερίσεων. Ο Πίνακας Δ3.2 περιέχει ποιοτικά στοιχεία, προερχόμενα από μελέτες ερευνητών από τον τομέα της Ψυχολογίας - Νευροφυσιολογίας, για το ποια FAP συμμετέχουν σε κάθε συγκεκριμένη έκφραση και το πρόσημο της τιμής τους. Σημειώνεται ότι τα στοιχεία του Πίνακα Δ3.2 είναι προσαρμογή στην ορολογία του MPEG-4 προτάσεων όπως «τα φρύδια κινούνται προς τα άνω» ή «οι άκρες του στόματος κινούνται προς τα έξω και άνω». Η πληροφορία προσήμου στα FAP του Πίνακα Δ3.2 δεν αρκεί για τον ορισμό των διαμερίσεων γιατί είναι τελείως ποιοτική. Αντίθετα τα στατιστικά στοιχεία του Πίνακα Δ3.4 μπορούν να αποτελέσουν τη βάση για μια πιο ουσιαστική διαμέριση –πλήθος κλάσεων και κέντρα αυτών. Για παράδειγμα από τον Πίνακα Δ3.4 προκύπτει ότι το διάστημα X_{11} μπορεί να διαμεριστεί σε δύο κλάσεις που μπορούν να συσχετιστούν με τους γλωσσικούς όρους *λιγό* και *πολύ*.



Σχήμα Δ3.5: Συναρτήσεις συμμετοχής τραπεζοειδούς μορφής

Έστω $A^{(1)}, A^{(2)}, \dots, A^{(15)}$ οι ασαφείς διαμερίσεις των διαστημάτων $X_i, i=1, \dots, 15$ με τάξεις -αριθμούς κλάσεων- $|A^{(i)}|=k_i$. Κάθε διαμέριση ορίζει ένα αριθμό κλάσεων: $A^{(i)} = \{A_1^{(i)}, A_2^{(i)}, \dots, A_{k_i}^{(i)}\}$ οι οποίες εκφράζονται από αντίστοιχες συναρτήσεις συμμετοχής $\mu_j^{(i)}, j=1, \dots, k_i$. Η μορφή των συναρτήσεων συμμετοχής είναι τραπεζοειδής και καθορίζεται από τις παραμέτρους $c_j^{(i)}$ και $s_j^{(i)}$ όπως φαίνεται στο Σχήμα Δ3.5.

Η επιλογή των παραμέτρων $c_j^{(i)}$ και $s_j^{(i)}$ γίνεται σύμφωνα με τις τιμές των m_{ij} και σ_{ij} και συνοψίζεται στον Πίνακα Δ3.5.

		«οργή»	«λύπη»	«χαρά»	«απέχθεια»	«φόβος»	«έκπληξη»
F ₁ (ES)	Mean	-57	*	*	*	*	*
	StD	28	*	*	*	*	*
F ₂ (ES)	Mean	-58	*	*	*	*	*
	StD	31	*	*	*	*	*
F ₃ (MNS)	Mean	-73	*	-271	-234	*	*
	StD	51	*	110	109	*	*
F ₄ (MNS)	Mean	*	*	*	-177	218	543
	StD	*	*	*	108	135	203
F ₅ (ENS)	Mean	-83	85	*	*	104	224
	StD	48	55	*	*	69	103
F ₆ (ENS)	Mean	-85	80	*	*	111	211
	StD	51	54	*	*	72	97
F ₇ (ENS)	Mean	-66	*	*	*	*	54
	StD	35	*	*	*	*	31
F ₈ (ENS)	Mean	-70	*	*	*	*	55
	StD	38	*	*	*	*	31
F ₉ (ENS)	Mean	-149	*	24	-80	72	144
	StD	40	*	22	53	58	64
F ₁₀ (ENS)	Mean	-144	*	25	-82	75	142
	StD	39	*	22	54	60	62
F ₁₁ (MNS)	Mean	*	*	*	*	291	885
	StD	*	*	*	*	189	316
F ₁₂ (IrisD)	Mean	*	-153	-254	-203	244	254
	StD	*	112	133	148	126	83
F ₁₃ (IrisD)	Mean	*	-161	-242	-211	249	252
	StD	*	109	122	145	128	81
F ₁₄ (MW)	Mean	*	*	234	*	*	-82
	StD	*	*	98	*	*	39
F ₁₅ (ES)	Mean	-69	-56	*	-52	*	86
	StD	51	35	*	34	*	60

Πίνακας Δ3.4: Πειραματικά αποτελέσματα για το ποιοι FAPs εμπλέκονται στις πρότυπες εκφράσεις («οργή», «λύπη», «χαρά», «απέχθεια», «φόβος», «έκπληξη»)

4.2 Οι κανόνες απόφασης

Οι Πίνακες Δ3.2 και Δ3.4 σε συνδυασμό με τον ορισμό των διαμερίσεων για τις μεταβλητές εισόδου χρησιμοποιήθηκαν για την δημιουργία των κανόνων απόφασης οι οποίοι συνοψίζονται στον Πίνακα Δ3.6. Έστω $X_i, i=1, \dots, 15$ τα πεδία ορισμού των μεταβλητών F_i , όπως αυτά ορίστηκαν στη δεύτερη στήλη του πίνακα Πίνακα Δ3.5, $X = X_1 \times X_2 \times \dots \times X_{15}$ το καρτεσιανό γινόμενο τους και $\mathfrak{Z}(X)$ το δυναμοσύνολο του συνόλου X . Έστω επίσης ότι Y είναι η έξοδος του ασαφούς συστήματος και $B = \{B_1, B_2, \dots, B_6\}$ μια ασαφής διαμέριση της. Μετά τη διαδικασία της ασαφοποίησης ένα διάνυσμα εισόδου $x \in X$ αντιστοιχίζεται με το ασαφές σύνολο $A'(x) \in \mathfrak{Z}(X)$ το οποίο απαρτίζεται από τα επιμέρους στοιχεία: $A'(x) = \{a_1, a_2, \dots, a_{15}\}$. Με βάση τα παραπάνω οι κανόνες απόφασης διαμορφώνονται όπως εμφανίζεται στον Πίνακα Δ3.6

Μεταβλητή	Διάστημα διακύμανσης X_i	Διαμέριση		
		Κλάσεις	Κέντρο κλάσης	Τυπική απόκλιση
F_1	$[m_{11} - 3 \cdot \sigma_{11} \quad 0]$	$A_1^{(1)}$	$c_1^{(1)} = m_{11}$	$s_1^{(1)} = \sigma_{11}$
F_2	$[m_{12} - 3 \cdot \sigma_{12} \quad 0]$	$A_1^{(2)}$	$c_1^{(2)} = m_{12}$	$s_1^{(2)} = \sigma_{12}$
F_3	$[m_{33} - 3 \cdot \sigma_{33} \quad m_{13} + 3 \cdot \sigma_{13}]$	$A_1^{(3)}$	$c_1^{(3)} = \frac{(m_{33} + m_{43})}{2}$	$s_1^{(3)} = \max\{\sigma_{33}, \sigma_{43}\}$
		$A_2^{(3)}$	$c_2^{(3)} = m_{13}$	$s_2^{(3)} = \sigma_{13}$
F_4	$[m_{44} - 3 \cdot \sigma_{44} \quad m_{64} + 3 \cdot \sigma_{64}]$	$A_1^{(4)}$	$c_1^{(4)} = m_{43}$	$s_1^{(4)} = \sigma_{43}$
		$A_2^{(4)}$	$c_2^{(4)} = m_{54}$	$s_2^{(4)} = \sigma_{54}$
		$A_3^{(4)}$	$c_3^{(4)} = m_{64}$	$s_3^{(4)} = \sigma_{64}$
F_5	$[m_{15} - 3 \cdot \sigma_{15} \quad m_{65} + 3 \cdot \sigma_{65}]$	$A_1^{(5)}$	$c_1^{(5)} = m_{15}$	$s_1^{(5)} = \sigma_{15}$
		$A_2^{(5)}$	$c_2^{(5)} = \frac{(m_{25} + m_{55})}{2}$	$s_2^{(5)} = \max\{\sigma_{25}, \sigma_{55}\}$
		$A_3^{(5)}$	$c_3^{(5)} = m_{65}$	$s_3^{(5)} = \sigma_{65}$
F_6	$[m_{16} - 3 \cdot \sigma_{16} \quad m_{66} + 3 \cdot \sigma_{66}]$	$A_1^{(6)}$	$c_1^{(6)} = m_{16}$	$s_1^{(6)} = \sigma_{16}$
		$A_2^{(6)}$	$c_2^{(6)} = \frac{(m_{26} + m_{56})}{2}$	$s_2^{(6)} = \max\{\sigma_{26}, \sigma_{56}\}$
		$A_3^{(6)}$	$c_3^{(6)} = m_{66}$	$s_3^{(6)} = \sigma_{66}$
F_7	$[m_{17} - 3 \cdot \sigma_{17} \quad m_{67} + 3 \cdot \sigma_{67}]$	$A_1^{(7)}$	$c_1^{(7)} = m_{17}$	$s_1^{(7)} = \sigma_{17}$
		$A_2^{(7)}$	$c_2^{(7)} = m_{67}$	$s_2^{(7)} = \sigma_{67}$
F_8	$[m_{18} - 3 \cdot \sigma_{18} \quad m_{68} + 3 \cdot \sigma_{68}]$	$A_1^{(8)}$	$c_1^{(8)} = m_{18}$	$s_1^{(8)} = \sigma_{18}$
		$A_2^{(8)}$	$c_2^{(8)} = m_{68}$	$s_2^{(8)} = \sigma_{68}$
F_9	$[m_{19} - 3 \cdot \sigma_{19} \quad m_{69} + 3 \cdot \sigma_{69}]$	$A_1^{(9)}$	$c_1^{(9)} = m_{19}$	$s_1^{(9)} = \sigma_{19}$
		$A_2^{(9)}$	$c_2^{(9)} = m_{49}$	$s_2^{(9)} = \sigma_{49}$
		$A_3^{(9)}$	$c_3^{(9)} = \frac{(m_{39} + m_{59})}{2}$	$s_3^{(9)} = \max\{\sigma_{39}, \sigma_{59}\}$

		$A_4^{(9)}$	$c_4^{(9)} = m_{69}$	$s_4^{(9)} = \sigma_{69}$
F ₁₀	$[m_{1,10} - 3 \cdot \sigma_{1,10} \quad m_{6,10} + 3 \cdot \sigma_{6,10}]$	$A_1^{(10)}$	$c_1^{(10)} = m_{1,10}$	$s_1^{(10)} = \sigma_{1,10}$
		$A_2^{(10)}$	$c_2^{(10)} = m_{4,10}$	$s_2^{(10)} = \sigma_{4,10}$
		$A_3^{(10)}$	$c_3^{(10)} = \frac{(m_{3,10} + m_{5,10})}{2}$	$s_3^{(10)} = \max\{\sigma_{3,10}, \sigma_{5,10}\}$
		$A_4^{(10)}$	$c_4^{(10)} = m_{6,10}$	$s_4^{(10)} = \sigma_{6,10}$
F ₁₁	$[0 \quad m_{6,11} + 3 \cdot \sigma_{6,11}]$	$A_1^{(11)}$	$c_1^{(11)} = m_{5,11}$	$s_1^{(11)} = \sigma_{5,11}$
		$A_2^{(11)}$	$c_2^{(11)} = m_{6,11}$	$s_2^{(11)} = \sigma_{6,11}$
F ₁₂	$[m_{3,12} - 3 \cdot \sigma_{3,12} \quad m_{5,12} + 3 \cdot \sigma_{5,12}]$	$A_1^{(12)}$	$c_1^{(12)} = \frac{(m_{3,12} + m_{4,12})}{2}$	$s_1^{(12)} = \max\{\sigma_{3,12}, \sigma_{4,12}\}$
		$A_2^{(12)}$	$c_2^{(12)} = m_{2,12}$	$s_2^{(12)} = \sigma_{2,12}$
		$A_3^{(12)}$	$c_3^{(12)} = \frac{(m_{5,12} + m_{6,12})}{2}$	$s_3^{(12)} = \max\{\sigma_{5,12}, \sigma_{6,12}\}$
F ₁₃	$[m_{4,13} - 3 \cdot \sigma_{4,13} \quad m_{5,13} + 3 \cdot \sigma_{5,13}]$	$A_1^{(13)}$	$c_1^{(13)} = \frac{(m_{3,13} + m_{4,13})}{2}$	$s_1^{(13)} = \max\{\sigma_{3,13}, \sigma_{4,13}\}$
		$A_2^{(13)}$	$c_2^{(13)} = m_{2,13}$	$s_2^{(13)} = \sigma_{2,13}$
		$A_3^{(13)}$	$c_3^{(13)} = \frac{(m_{5,13} + m_{6,13})}{2}$	$s_3^{(13)} = \max\{\sigma_{5,13}, \sigma_{6,13}\}$
F ₁₄	$[m_{6,14} - 3 \cdot \sigma_{6,14} \quad m_{3,14} + 3 \cdot \sigma_{3,14}]$	$A_1^{(14)}$	$c_1^{(14)} = m_{6,14}$	$s_1^{(14)} = \sigma_{6,14}$
		$A_2^{(14)}$	$c_2^{(14)} = m_{3,14}$	$s_2^{(14)} = \sigma_{3,14}$
F ₁₅	$[m_{1,15} - 3 \cdot \sigma_{1,15} \quad m_{6,15} + 3 \cdot \sigma_{6,15}]$	$A_1^{(15)}$	$c_1^{(15)} = \frac{(m_{1,15} + m_{4,15})}{2}$	$s_1^{(15)} = \max\{\sigma_{1,15}, \sigma_{4,15}\}$
		$A_2^{(15)}$	$c_2^{(15)} = m_{6,15}$	$s_2^{(15)} = \sigma_{6,15}$

Πίνακας Δ3.5: Αναλυτική περιγραφή των διαμερίσεων για όλες τις μεταβλητές

4.3 Υπολογισμός της εξόδου του συστήματος ασαφούς λογικής

Για τον προσδιορισμό της εξόδου του ασαφούς συστήματος ορίζουμε τον βαθμό συνάφειας r_{ij} του i -στού στοιχείου του διανύσματος εισόδου με την κλάση $A_j^{(i)}$ όπως εκφράζεται από τη σχέση:

$$r_{ij} = \max\{a_i \cap A_j^{(i)}\} \quad (\Delta 3.3)$$

Ορίζουμε επίσης ως $h(k)$ των αριθμό των υποθέσεων που πρέπει να πληρούνται ταυτόχρονα στον κανόνα k . Για παράδειγμα για τον Κανόνα 4 είναι $h(4) = 7$ ενώ για τον κανόνα 5 είναι $h(5) = 5$.

Στους κανόνες του Πίνακα E2.6 ο γλωσσικός όρος «και» αντιστοιχεί στην ασαφή τομή δύο συνόλων και υλοποιείται από μια τ-νόρμα ενώ ο γλωσσικός όρος «ή» αντιστοιχεί στην ασαφή ένωση δύο συνόλων και

υλοποιείται από μια σ -νόρμα. Στην υλοποίηση χρησιμοποιήσαμε την τ -νόρμα $t(a,b)=a \cdot b$ και την σ -νόρμα $u(a,b)=\max(a,b)$.

Έστω $b(l)$ ο βαθμός πεποίθησης ότι η παρατηρούμενη, δια του διανύσματος εισόδου x , έκφραση είναι η l (όπου $l \in \{1 \Rightarrow \text{«οργή»}, 2 \Rightarrow \text{«λύπη»}, 3 \Rightarrow \text{«χαρά»}, 4 \Rightarrow \text{«απέχθεια»}, 5 \Rightarrow \text{«φόβος»}, 6 \Rightarrow \text{«έμπληξη»}\}$) και $T(l) \in \mathfrak{Z}(X)$, $U(l) \in \mathfrak{Z}(X)$, τα σύνολα των εμπλεκομένων, αποκλειστικά, με συνθήκη «και» και «ή» αντίστοιχα, κλάσεων στον κανόνα l –για παράδειγμα $T(4) = \{A_1^{(3)}, A_1^{(4)}, A_2^{(9)}, A_2^{(10)}, A_1^{(12)}, A_1^{(13)}, A_1^{(15)}\}$ και $U(4) = \{A_1^{(11)}, A_1^{(14)}\}$ - τότε μπορούμε να υπολογίσουμε τα $b(l)$ από την επόμενη σχέση:

$$b(l) = \frac{\prod_{A_j^{(i)} \in T(l)} r_{ij} \cdot \max_{A_j^{(i)} \in U(l)} \{r_{ij}\}}{h(l)} \quad (\Delta 3.4)$$

Αν χρειάζεται να βγει κάποιο οριστικό συμπέρασμα σε σχέση με την παρατηρούμενη έκφραση αυτό λαμβάνεται από τη σχέση:

$$q = \arg \max_l b(l) \quad (\Delta 3.5)$$

Κανόνας 1: Αν a_1 είναι $A_1^{(1)}$ και a_2 είναι $A_1^{(2)}$ και a_3 είναι $A_2^{(3)}$ και a_5 είναι $A_1^{(5)}$ και a_6 είναι $A_1^{(6)}$ και a_7 είναι $A_1^{(7)}$ και a_8 είναι $A_1^{(8)}$ και a_9 είναι $A_1^{(9)}$ και a_{10} είναι $A_1^{(10)}$ και a_{15} είναι $A_1^{(15)}$ τότε b είναι B_1

Κανόνας 2: Αν a_5 είναι $A_2^{(5)}$ και a_6 είναι $A_2^{(6)}$ και a_{12} είναι $A_2^{(12)}$ και a_{13} είναι $A_2^{(13)}$ και a_{15} είναι $A_1^{(15)}$ τότε b είναι B_2

Κανόνας 3: Αν a_9 είναι $A_3^{(9)}$ και a_{10} είναι $A_3^{(10)}$ και a_{12} είναι $A_1^{(12)}$ και a_{13} είναι $A_1^{(13)}$ και a_{14} είναι $A_2^{(14)}$ και $\{a_3$ είναι $A_1^{(3)}$ ή a_{11} είναι $A_1^{(11)}$ ή a_4 είναι $A_2^{(4)}\}$ τότε b είναι B_3

Κανόνας 4: Αν a_3 είναι $A_1^{(3)}$ και a_4 είναι $A_1^{(4)}$ και a_9 είναι $A_2^{(9)}$ και a_{10} είναι $A_2^{(10)}$ και a_{12} είναι $A_1^{(12)}$ και a_{13} είναι $A_1^{(13)}$ και a_{15} είναι $A_1^{(15)}$ και $\{a_{11}$ είναι $A_1^{(11)}$ ή a_{14} είναι $A_1^{(14)}\}$ τότε b είναι B_4

Κανόνας 5: Αν a_4 είναι $A_2^{(4)}$ και a_{11} είναι $A_1^{(11)}$ και a_{12} είναι $A_3^{(12)}$ και a_{13} είναι $A_3^{(13)}$ και $\{a_1$ είναι $A_1^{(1)}$ ή a_2 είναι $A_1^{(2)}$ ή a_5 είναι $A_2^{(5)}$ ή a_6 είναι $A_2^{(6)}$ ή a_9 είναι $A_3^{(9)}$ ή a_{10} είναι $A_3^{(10)}\}$ τότε b είναι B_5

Κανόνας 6: Αν a_4 είναι $A_3^{(4)}$ και a_5 είναι $A_3^{(5)}$ και a_6 είναι $A_3^{(6)}$ και a_7 είναι $A_2^{(7)}$ και a_8 είναι $A_2^{(8)}$ και a_9 είναι $A_4^{(9)}$ και a_{10} είναι $A_4^{(10)}$ και a_{11} είναι $A_2^{(11)}$ και a_{12} είναι $A_3^{(12)}$ και a_{13} είναι $A_3^{(13)}$ και a_{14} είναι $A_1^{(14)}$ και a_{15} είναι $A_2^{(15)}$ τότε b είναι B_6

Πίνακας Δ3.6: Οι κανόνες του συστήματος απόφασης

5. ΑΝΑΓΝΩΡΙΣΗ ΠΑΡΑΛΛΑΓΩΝ ΤΩΝ ΠΡΟΤΥΠΩΝ ΣΥΝΑΙΣΘΗΜΑΤΩΝ

Το σύστημα το οποίο περιγράφηκε στην παράγραφο 4 μπορεί να τροποποιηθεί ώστε να χρησιμοποιηθεί για την ανάλυση παραλλαγών των προτύπων συναισθημάτων. Για να επιτευχθεί αυτό χρειάζεται (α) να εκτιμηθεί ποια από τα χαρακτηριστικά του Πίνακα Δ3.3 συμμετέχουν στη διαμόρφωση των συναισθημάτων αυτών και (β) να τροποποιήσουμε με κάποιο λογικό τρόπο τις συναρτήσεις συμμετοχής για τις διάφορες κλάσεις των διαμερίσεων των επιμέρους στοιχείων του διανύσματος εισόδου.

Σε μια γενική θεώρηση μπορούμε να διακρίνουμε έξι κατηγορίες συναισθημάτων που αντιστοιχούν σε κάθε ένα από τα πρότυπα συναισθήματα. Στο πλαίσιο κάθε μιας από τις κατηγορίες αυτές ενδιάμεσα συναισθήματα εκφράζονται από διαφορετική ένταση των εσωτερικών αισθημάτων και των παρατηρούμενων οπτικών στοιχείων καθώς και από μικρές διακυμάνσεις στη γενικότερη δομή τους. Για παράδειγμα η κατηγορία που αντιστοιχεί στο πρότυπο συναίσθημα «φόβος» περιλαμβάνει και τα συναισθήματα «ανησυχία» και «τρόμος». Τα συναισθήματα αυτά μπορούν να μοντελοποιηθούν, στο πλαίσιο του ασαφούς συστήματος, τροποποιώντας τις διαμερίσεις των μεταβλητών εισόδου ώστε να συμπεριλάβουμε και νέες κλάσεις οι οποίες αντιστοιχούν στα νέα συναισθήματα. Για το σκοπό αυτό –και με δεδομένη την απουσία πειραματικών στοιχείων– οι νέες κλάσεις θα είναι απλά μετακινημένες ως προς την κεντρική τιμή εκδοχές των κλάσεων που αντιστοιχούν στο πρότυπο συναίσθημα.

Έστω a_x και a_y οι τιμές ενεργοποίησης των συναισθημάτων X και Y τα οποία ανήκουν στην ίδια κατηγορία και $\mu_j^{(i)}$ η συνάρτηση συμμετοχής της κλάσης $A_j^{(i)}$ τότε για την μοντελοποίηση του συναισθήματος Y με δεδομένη την αντίστοιχη του X ακολουθούμε τους παρακάτω κανόνες:

Κανόνας 1: Συναισθήματα της ίδιας κατηγορίας περιγράφονται από τα ίδια χαρακτηριστικά F_i και με την ίδια δομή κανόνων απόφασης

Κανόνας 2: Ο βαθμός συνάφειας r_{ij} του i -στού στοιχείου του διανύσματος εισόδου με την κλάση $A_j^{(i)}$ εκφράζεται από τη σχέση: $r_{ij} = \max\{a_i \cap A_j^{(i)}\}$ όπου $A_j^{(i)}$ είναι η μετατοπισμένη εκδοχή της κλάσης $A_j^{(i)}$ όπως αυτή περιγράφεται από την συνάρτηση συμμετοχής $\mu_j^{(i)} = \frac{a_y}{a_x} \mu_j^{(i)}$

Κανόνας 3: Οι τιμές των a_x και a_y προέρχονται από τη μελέτη του Whissel [3]

6. ΠΕΙΡΑΜΑΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ

Για την αξιολόγηση της αποδοτικότητας του προτεινόμενου αλγορίθμου πραγματοποιήσαμε πειράματα με δύο διαφορετικά σύνολα δεδομένων. Το πρώτο αποτελείται από στατικές εικόνες που επιδεικνύουν πρότυπες εκφράσεις στην κορύφωση τους -“αρεχ”- ενώ το δεύτερο περιέχει ακολουθίες βίντεο με μια

μεγαλύτερη ποικιλία απεικονιζόμενων συναισθημάτων. Τα αποτελέσματα συνοψίζονται στους Πίνακες Δ3.7 και Δ3.8.

Υλικό: Το πρώτο σύνολο αποτελείται από (α) 80 εικόνες από τη βάση προσώπων CMU οι οποίες επιδεικνύουν πρόσωπα σε ουδέτερη έκφραση καθώς και εκφράσεις όπως «χαρά», «λύπη» και «οργή», (β) 60 εικόνες από τη βάση προσώπων Yale που αντιστοιχούν σε ουδέτερη έκφραση και τις εκφράσεις «χαρά», «λύπη» και «έκπληξη», (γ) 100 καρτέ από τις ακολουθίες του MediaLab που αντιστοιχούν σε ουδέτερη έκφραση και τις εκφράσεις «χαρά», «λύπη», «απέχθεια», «οργή» και «έκπληξη», (δ) 30 εικόνες από διάφορες πηγές με ουδέτερη έκφραση και έκφραση «φόβος». Οι εικόνες με ουδέτερη έκφραση χρησιμοποιήθηκαν σε όλες τις περιπτώσεις ως το πρώτο καρτέ για να εκτιμηθούν οι παράμετροι FAPs σε ηρεμία.

Το δεύτερο σύνολο αποτελείται από την πιλοτική βάση δεδομένων που δημιουργήθηκε στο πλαίσιο του προγράμματος PHYSTA και περιλαμβάνει ακολουθίες βίντεο και ήχου οι οποίες επιδεικνύουν ανθρώπους σε διάφορες συναισθηματικές καταστάσεις –όχι μόνο πρότυπες. Στις προσομοιώσεις χρησιμοποιήθηκε μόνο το τμήμα βίντεο των ακολουθιών αυτών.

	«οργή»	«λύπη»	«χαρά»	«απέχθεια»	«φόβος»	«έκπληξη»
Στατικό Σύνολο	85%	76%	92%	73%	67%	94%
PHYSTA	68%	61%	87%	64%	58%	85%

Πίνακας Δ3.7: Ποσοστά επιτυχούς ταξινόμησης πρότυπων εκφράσεων

Σχολιασμός: Από τον Πίνακα Δ3.7 προκύπτει ότι τα ποσοστά αναγνώρισης στο σύνολο PHYSTA είναι χαμηλότερα από τα αντίστοιχα στο σύνολο με τις στατικές εικόνες. Αυτό οφείλεται στο γεγονός ότι στο σύνολο PHYSTA υπάρχουν συναισθηματικές καταστάσεις από την καθημερινή ζωή και όχι ακραίες περιπτώσεις όπως αυτές του στατικού συνόλου. Φαίνεται επίσης από τον ίδιο πίνακα ότι εκφράσεις με μεγαλύτερο παράγοντα ενεργοποίησης που αντιστοιχούν σε ισχυρότερες δράσεις των μυών του προσώπου είναι ευκολότερα αναγνωρίσιμες. Ο Πίνακας E2.8 δείχνει κάποια αποτελέσματα ταξινόμησης παραλλαγών των πρότυπων εκφράσεων. Παρόλο που τα αποτελέσματα είναι σχετικά χαμηλά είναι πάντα υψηλότερα από το ποσοστό τυχαιότητας πράγμα που υποδηλώνει ότι η διαδικασία που περιγράφηκε στην παράγραφο 5 είναι κατάλληλη για την ταξινόμηση των παραλλαγών των πρότυπων εκφράσεων.

	«περιφρόνηση»	«απέχθεια»	«αντιπάθεια»	«απόλαυση»	«ενθουσιασμός»	«χαρά»
Ποσοστά Αναγνώρισης	48%	60%	52%	61%	65%	75%

Πίνακας Δ3.8: Ποσοστά αναγνώρισης παραλλαγών των πρότυπων εκφράσεων

7. ΣΥΜΠΕΡΑΣΜΑΤΑ

Στο κεφάλαιο αυτό παρουσιάστηκε μια μελέτη η οποία επιδεικνύει ότι οι παράμετροι FAP και FDP του MPEG-4 συνδυαζόμενοι με κάποιο σύστημα ασαφούς λογικής μπορεί να χρησιμοποιηθούν και για την

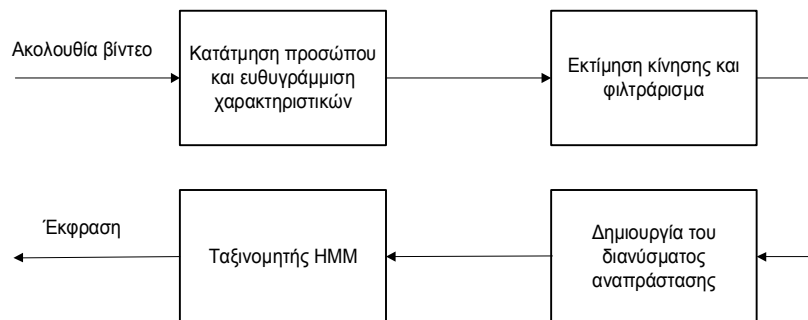
ανάλυση εκφράσεων και συναισθημάτων. Το σύστημα ασφούς λογικής ενσωματώνει την πρότερη γνώση για κάποια χαρακτηριστικά των εκφράσεων η οποία προέρχεται από μελέτες στον τομέα της Ψυχολογίας και στατιστική ανάλυση πειραματικών δεδομένων. Επιπλέον η αβεβαιότητα για τον επακριβή εντοπισμό των σημείων προσώπου, τα οποία παρουσιάζουν επίπεδη μετακίνηση κατά την δράση των μυών του, επιβάλλει το χρήση κάποιου σχήματος το οποίο να δέχεται το γεγονός αυτό. Τέλος με τη χρήση της παραμέτρου ενεργοποίηση μπορεί να υπάρξει μοντελοποίηση παραλλαγών προτύπων εκφράσεων για τις οποίες δεν υπάρχουν επαρκή πειραματικά δεδομένα για τη εξαγωγή στατιστικών συμπερασμάτων.

ΠΑΡΑΚΟΛΟΥΘΗΣΗ ΤΗΣ ΧΡΟΝΙΚΗΣ ΕΞΕΛΙΞΗΣ ΤΩΝ ΕΚΦΡΑΣΕΩΝ ΜΕ HIDDEN MARKOV MODELS

Στο κεφάλαιο αυτό παρουσιάζεται μια δυναμική προσέγγιση της αναγνώρισης ανθρωπίνων εκφράσεων, η οποία στηρίζεται στα HMM και χρησιμοποιεί τόσο την χωρική πληροφορία, όσον αφορά την κατανομή της κίνησης σε διάφορες περιοχές του προσώπου, καθώς και πως η πυκνότητα αυτής της κίνησης μεταβάλλεται χρονικά. Τα μοντέλα HMM έχουν την ικανότητα να χειρίζονται χρονικά μεταβαλλόμενα δεδομένα ανεξάρτητα από την κλίμακα του χρόνου. Η ιδιότητα τους αυτή είναι χρήσιμη στην ταξινόμηση εκφράσεων από ακολουθίες βίντεο. Για το σκοπό αυτό μπορούν να τροφοδοτηθούν με κάποιο διάνυσμα αναπαράστασης το οποίο εξάγεται από την κίνηση των διαφόρων σημείων του προσώπου.

1. Ο ΠΡΟΤΕΙΝΟΜΕΝΟΣ ΑΛΓΟΡΙΘΜΟΣ

Το χονδρικό μπλοκ διάγραμμα του προτεινόμενου αλγορίθμου επιδεικνύεται στην Εικόνα Ε4.1. Οι εκφράσεις δημιουργούνται κατά βάση από την κίνηση των μυών του προσώπου και λιγότερο από την ολική κίνηση του η οποία άλλωστε δεν μπορεί εύκολα να διαχωριστεί στο τμήμα που συμμετέχει σε κάποιες εκφράσεις. Για να αποφευχθεί η επίδραση της ολικής κίνησης στην εκτίμηση της κίνησης των χαρακτηριστικών του προσώπου, το πρόσωπο εντοπίζεται και απομονώνεται από το φόντο με χρήση κάποιων από τις τεχνικές που περιγράφονται στα Κεφάλαια Β1 και Β2 της διατριβής. Τα υπόλοιπα τμήματα του διαγράμματος περιγράφονται στις επόμενες παραγράφους.



Σχήμα Δ4.1: Η αρχιτεκτονική του προτεινόμενου σχήματος

1.1 Εκτίμηση κίνησης

Για την εκτίμηση της κίνησης των σημείων στην επιφάνεια του προσώπου ακολουθείται η παρακάτω διαδικασία:

- Έστω F_k και F_{k+1} δύο διαδοχικά καρέ στα οποία το πρόσωπο έχει καταταμηθεί και κανονικοποιηθεί ως προς τη κλίμακα και τη θέση. Κάθε pixel $p_k(x, y)$ του k -στου καρέ περιγράφεται μέσω του $2n \times 2n$ μπλοκ $b_k(x, y)$ που το περιλαμβάνει και αντιστοιχίζεται με το επόμενο σφάλμα (MAD-Mean Absolute Difference):

$$e_k(x, y) = |b_k(x, y) - b_{k+1}(x, y)| = \sum_{l=-n}^n \sum_{m=-n}^n |p_k(x+l, y+m) - p_{k+1}(x+l, y+m)| \quad (\Delta 4.1)$$

- Τα διανύσματα μετατόπισης υπολογίζονται μόνο για εκείνα τα μπλοκ του καρέ τα οποία αντιστοιχούν σε μεγάλο σφάλμα $e_k(x, y)$. Έστω ότι m_b και s_b η μέση τιμή και τυπική απόκλιση των e_k υπολογισμένα σε όλα τα μπλοκ του k -στου καρέ. Υπολογίζουμε την μετατόπιση μόνο για τα μπλοκ που ανήκουν στο σύνολο I το οποίο δίνεται από τη σχέση:

$$I = \{b_k \mid e_k > (m_b + s_b)\} \quad (\Delta 4.2)$$

- Το διάνυσμα μετατόπισης $\hat{v}_k(x, y)$ του μπλοκ $b_k(x, y)$ υπολογίζεται με τη διαδικασία ταύτισης μπλοκ σε μια γειτονιά του $b_{k+1}(x, y)$ σύμφωνα με την εξίσωση:

$$\hat{v}_k(x, y) = (\hat{v}_x, \hat{v}_y) = \arg \min_{(v_x, v_y) \in Q} \sum_{l=-n}^n \sum_{m=-n}^n |p_k(x+l, y+m) - p_{k+1}(x+l-v_x, y+m-v_y)| \quad (\Delta 4.3)$$

όπου $Q = \{-q, \dots, q\} \times \{-q, \dots, q\}$ ένα τετραγωνικό πλέγμα που αντιπροσωπεύει την περιοχή έρευνας. Για την μείωση του χρόνου εκτέλεσης πραγματοποιείται λογαριθμική αντί για εξαντλητική έρευνα, δηλαδή μόνο ένα υποσύνολο των συνδυασμών $(\hat{v}_x, \hat{v}_y) \in Q$ ελέγχεται στη διαδικασία ταύτισης

Τονίζεται ότι η διαδικασία υπολογισμού της μετατόπισης για τα σημεία του προσώπου είναι η απλούστερη δυνατή για σκοπούς περιορισμού της υπολογιστικής πολυπλοκότητας και ταχείας εκτέλεσης. Επίσης έγινε προσπάθεια συμβατότητας με τη διαδικασία εκτίμησης κίνησης των προτύπων MPEG-1 και MPEG-2.

1.2 Φιλτράρισμα των διανυσμάτων κίνησης

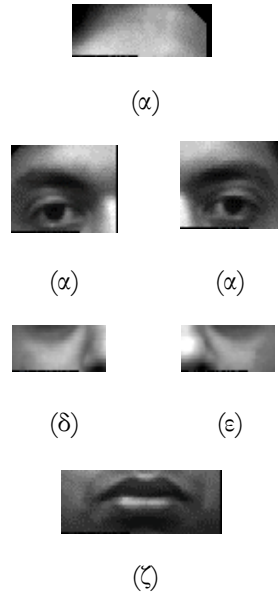
Η διαδικασία της εκτίμησης της μετατόπισης των σημείων στην επιφάνεια του προσώπου είναι φυσικό να παρουσιάζει και σφάλματα. Με βάση τη συνέχεια των ιστών του δέρματος του προσώπου το φιλτράρισμα με φίλτρα ενδιάμεσης τιμής -median filtering- όσον αφορά την κατεύθυνση των διανυσμάτων κίνησης είναι μια προφανής επιλογή. Μετά το κατευθυντικό φιλτράρισμα πραγματοποιείται και φιλτράρισμα όσον αφορά το μέτρο των διανυσμάτων.

2 ΔΗΜΙΟΥΡΓΙΑ ΤΟΥ ΔΙΑΝΥΣΜΑΤΟΣ ΑΝΑΠΑΡΑΣΤΑΣΗΣ

Έστω S η περιοχή του προσώπου η οποία έχει εντοπιστεί και απομονωθεί με τεχνικές όμοιες με αυτές που παρουσιάστηκαν στο Μέρος Β της διατριβής. Μπορούμε να διαμερίσουμε την περιοχή του προσώπου σε προκαθορισμένες περιοχές όπως επιδεικνύεται στην Εικόνα Δ4.1. Εμπνευσμένοι από τη εργασία του

Bassili [33] διαιρούμε την περιοχή S σε έξι υποπεριοχές που αντιστοιχούν στο μέτωπο, το αριστερό και δεξί μάτι, το αριστερό και δεξί μάγουλο και το στόμα:

$$S = \{s_1, s_2, s_3, s_4, s_5, s_6\} \quad (\Delta 4.4)$$



Εικόνα Δ4.1: Κατάτμηση προσώπου σε υποπεριοχές (α) s_1 (β) s_2 (γ) s_3 (δ) s_4 (ε) s_5 (ζ) s_6

Έστω $A_i(s_j)$ η i -στή ομάδα διανυσμάτων κίνησης που αντιστοιχεί στη περιοχή s_j , και $\mathbf{v}_k = a_k e^{j\varphi_k}$ η μιγαδική αναπαράσταση του διανύσματος μετατόπισης \mathbf{v}_k -βλέπε Εικόνα Δ4.3. Η ομαδοποίηση των διανυσμάτων κίνησης μπορεί να πραγματοποιηθεί σύμφωνα με την κατεύθυνση τους -οριζόντια, διαγώνια και κάθετα- με βάση την επόμενη σχέση:

$$\{A_i : \mathbf{v}_k \in s_j, (2i-1) \cdot \frac{\pi}{8} \leq \varphi_k < (2i+1) \cdot \frac{\pi}{8} \mid i=1...8, j=1...6\} \quad (\Delta 4.5)$$

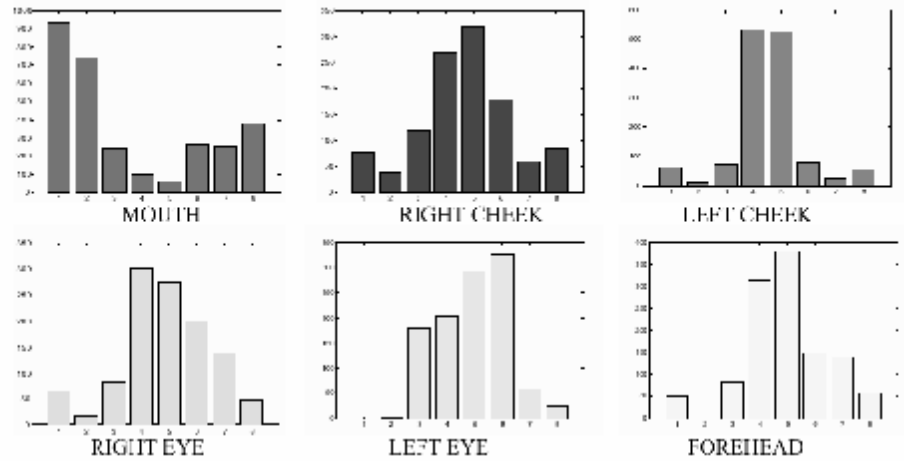
Η ενέργεια $E_i(s_j)$ σε κάθε κατεύθυνση στην υποπεριοχή s_j υπολογίζεται σύμφωνα με τη σχέση:

$$E_i(s_j) = \sum_{\mathbf{v}_k \in A_i(s_j)} \|\mathbf{v}_k\| \quad (\Delta 4.6)$$

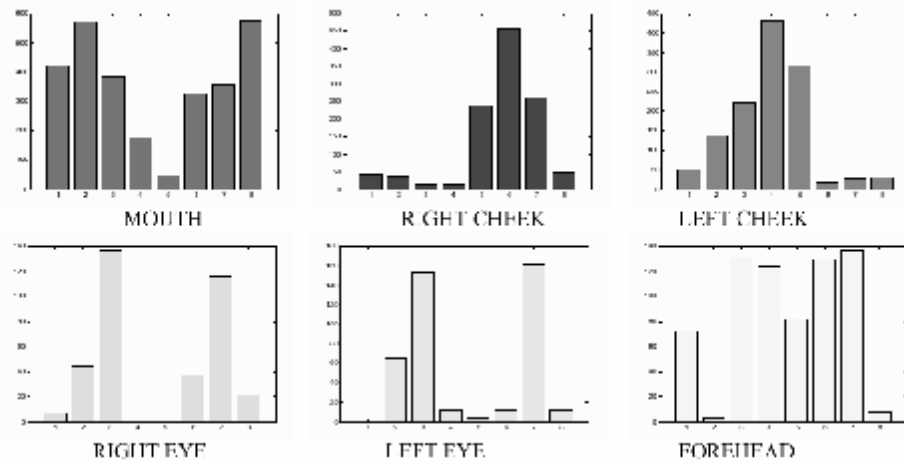
Το διάνυσμα χαρακτηριστικών που περιγράφει τη συνολική μετατόπιση αποτελείται από το σύνολο των E_i σε ολόκληρη την επιφάνεια S . Σύμφωνα με την περιγραφή που δόθηκε προηγουμένως το διάνυσμα χαρακτηριστικών αποτελείται από 48 στοιχεία που αντιστοιχούν στη μέση ενέργεια σε έξι διευθύνσεις για κάθε μια από τις οκτώ υποπεριοχές του προσώπου. Το διάνυσμα χαρακτηριστικών μπορεί να περιοριστεί σε 24 στοιχεία αν αφαιρεθούν τα στοιχεία που αντιστοιχούν κάποιες μη επιτρεπτές κινήσεις ή κινήσεις που δεν διαμορφώνουν εκφράσεις των χαρακτηριστικών του προσώπου -για παράδειγμα η οριζόντια κίνηση στη περιοχή των ματιών δεν είναι σημαντική για τον χαρακτηρισμό κάποιας από τις εκφράσεις. Περιοχές

του προσώπου με περισσότερους βαθμούς ελευθερίας, όπως το στόμα, περιγράφονται με χρήση περισσότερων στοιχείων- ενέργεια και για τις οκτώ διευθύνσεις.

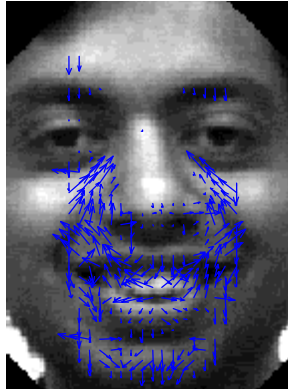
SURPRISE:



HAPPINESS:



Εικόνα Δ4.2: Ενέργεια στις οκτώ κύριες κατευθύνσεις για κάποιες από τις περιοχές του προσώπου. Κατευθύνσεις 1: Κάτω 2: Αριστερά κάτω 3: Αριστερά 4: Αριστερά πάνω 5: Πάνω 6: Δεξιά πάνω 7: Δεξιά 8: Δεξιά κάτω.



Εικόνα Δ4.3: Διανύσματα κίνησης σχετιζόμενα με την έκφραση “χαρά”

3. Ο ΤΑΞΙΝΟΜΗΤΗΣ HMM

Η ικανότητα των HMMs να χειρίζονται ακολουθιακά δεδομένα, η ανεξαρτησία τους από μεταβολές της κλίμακας του χρόνου –η διάρκεια των εκφράσεων δεν πρέπει να θεωρείται ως χαρακτηριστικό τους αντίθετα με τις καταστάσεις από τις οποίες διέρχονται– καθώς και η δυνατότητα μάθησης, τα καθιστούν κατάλληλα για την ταξινόμηση αγνώστων ακολουθιών από διανύσματα. Για αναλυτική περιγραφή των HMMs ο αναγνώστης παραπέμπεται στην εργασία [34]. Στη συνέχεια επικεντρωνόμαστε σε κάποια στοιχεία που αφορούν την υλοποίηση και την εκπαίδευση του ταξινομητή HMM.

Ένα γενικό μπλοκ διάγραμμα του ταξινομητή HMM επιδεικνύεται στο Σχήμα Δ4.2. Έξι διαφορετικά HMMs που αντιστοιχούν στις πρότυπες εκφράσεις –«χαρά», «λύπη», «οργή», «απέχθεια», «φόβος» και «έκπληξη»– χρησιμοποιούνται για την διάρθρωση του ταξινομητή. Τα μοντέλα είναι πρώτης τάξης, με μεταβάσεις από αριστερά στα δεξιά μόνο και αποτελούνται από τέσσερις καταστάσεις G_1 , G_2 , G_3 και G_4 . Οι καταστάσεις αυτές αντιστοιχούν χονδρικά σε ουδέτερη κατάσταση, έκφραση σε εξέλιξη, κορύφωση και χαλάρωση. Στη συνέχεια δίνονται κάποια χαρακτηριστικά της υλοποίησης των HMMs.

- Η πιθανότητα στην έξοδο των καταστάσεων λαμβάνεται με χρήση μιας συνεχούς κατανομής πυκνότητας πιθανότητας. Η επιλογή αυτή έγινε για να μειωθούν οι απαιτήσεις για δεδομένα εκπαίδευσης. Στη διακριτή περίπτωση η τάξη της διαμέρισης –codebook– πρέπει να είναι αρκετά μεγάλη ώστε να ελαχιστοποιείται το σφάλμα βαντισμού. Ένα πρόβλημα με της συνεχείς κατανομές πυκνότητας πιθανότητας είναι η αρχικοποίηση των παραμέτρων τους ώστε να αποφεύγεται η σύγκλιση σε τοπικά ελάχιστα.
- Η συνάρτηση πυκνότητα πιθανότητας στην έξοδο της κατάστασης G_j προσεγγίζεται από μια κατανομή Gauss πολλών μεταβλητών:

$$b_j(\mathbf{O}_i) = \frac{\exp\left\{-\frac{1}{2}(\mathbf{O}_i - \boldsymbol{\mu}_j)^T \mathbf{C}_j^{-1}(\mathbf{O}_i - \boldsymbol{\mu}_j)\right\}}{(2\pi)^{\frac{k}{2}} \cdot |\mathbf{C}_j|^{\frac{1}{2}}} \quad (\Delta 4.7)$$

όπου \mathbf{O}_i είναι η i -στη παρατήρηση (διάνυσμα χαρακτηριστικών εισόδου), $\boldsymbol{\mu}_j$ είναι το μέσο διάνυσμα της κατάστασης G_j , \mathbf{C}_j είναι ο αντίστοιχος πίνακας συμμεταβλητότητας και k είναι ο αριθμός των στοιχείων στο διάνυσμα \mathbf{O}_i (στη περίπτωση που εξετάζουμε $k=24$). Οι αρχικές τιμές για τα διανύσματα $\boldsymbol{\mu}_j$ και \mathbf{C}_j εκτιμώνται με στατιστικές τεχνικές και πριν την έναρξη της διαδικασίας εκπαίδευσης. Επανεκτίμηση των παραμέτρων γίνεται με τη βοήθεια μιας παραλλαγής του αλγόριθμου Baum-Welch η οποία εφαρμόζεται σε διανύσματα (όπως το $\boldsymbol{\mu}_j$) και πίνακες (όπως ο \mathbf{C}_j).

Οι πιθανότητες μετάβασης a_{mn} ανάμεσα στις καταστάσεις G_m και G_n εκτιμώνται κατά τη διάρκεια της εκπαίδευσης και λαμβανοντας υπόψη τους περιορισμούς: (α) $a_{mn}=0$ για κάθε $n < m$ -μεταβάσεις μόνο από αριστερά προς τα δεξιά (β) $a_{mm}=1-a_{mn}$ -δυνατή μόνο η μετάβαση στην επόμενη κατάσταση ή παραμονή στην ίδια.

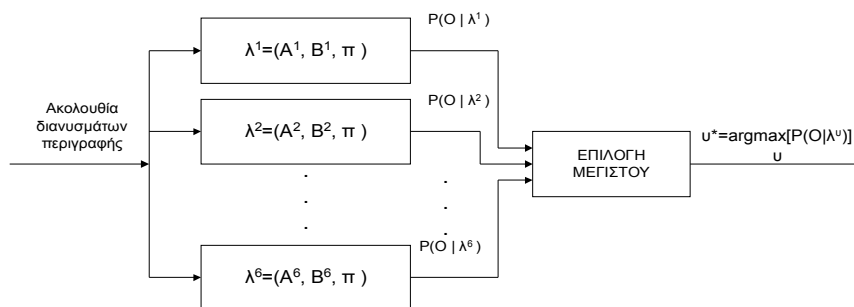
- Ο βαθμός ταύτισης της τροχιάς του διανύσματος αναπαράστασης, όπως αυτή περιγράφεται από την ακολουθία $\mathbf{O} = \mathbf{O}_1 \mathbf{O}_2 \dots \mathbf{O}_T$, δεδομένου του μοντέλου $\lambda_m(\mathbf{A}_m, \mathbf{B}_m, \boldsymbol{\pi}_m)$ επιδεικνύεται στο Σχήμα Δ4.2 και συνοψίζεται παρακάτω:

- 1) Υπολογίζουμε την βέλτιστη ακολουθία καταστάσεων \mathbf{Q}^* , με δεδομένη την παρατηρούμενη ακολουθία \mathbf{O} , χρησιμοποιώντας τον αλγόριθμο του Viterbi:

$$\mathbf{Q}^* = \arg \max_{\mathbf{Q}} \{P(\mathbf{Q} / \mathbf{O}, \lambda_m)\} \quad (\Delta 4.8)$$

- 2) Ο βαθμός ταύτισης της παρατηρούμενης ακολουθίας \mathbf{O} δεδομένης της ακολουθίας καταστάσεων \mathbf{Q}^* δίνεται από την ποσότητα:

$$P^* = P(\mathbf{O} / \mathbf{Q}^*, \lambda_m) \quad (\Delta 4.9)$$



Σχήμα Δ4.2 Διάγραμμα του ταξινομητή HMM

4. ΠΕΙΡΑΜΑΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ

Για την αξιολόγηση της απόδοσης του αλγορίθμου χρησιμοποιήσαμε τις ακολουθίες του MediaLab καθώς και επιλεγμένες ακολουθίες από δεδομένα του project PHYSTA ώστε να δημιουργήσουμε τουλάχιστον 30 ακολουθίες για κάθε μια από τις έξι πρότυπες εκφράσεις. Τρεις ακολουθίες χρησιμοποιήθηκαν για την αρχικοποίηση του HMM άλλες δώδεκα για την εκπαίδευση του και οι υπόλοιπες δεκαπέντε για τη διαδικασία επαλήθευσης. Οι ακολουθίες που χρησιμοποιήθηκαν για την εκπαίδευση του HMM αποτελούνται από 15 καρέ έκαστη και επελέγησαν έτσι ώστε να περιέχουν χαρακτηριστικά παραδείγματα των εκφράσεων. Οι ακολουθίες επαλήθευσης υπέστησαν υποδειγματοληψία με ρυθμό 5 καρέ ανά δευτερόλεπτο ώστε να υπάρχει εμφανής μετατόπιση στα χαρακτηριστικά του προσώπου ανάμεσα σε διαδοχικά καρέ. Τα πειραματικά αποτελέσματα επιδεικνύονται στον Πίνακα Δ4.1.

Έκφραση	«φόβος»	«απέχθεια»	«χαρά»	«λύπη»	«έκπληξη»	«οργή»
«φόβος»	10	1	0	2	0	0
«απέχθεια»	2	11	0	0	0	1
«χαρά»	0	0	14	0	1	0
«λύπη»	3	2	1	12	0	1
«έκπληξη»	0	0	0	0	14	0
«οργή»	0	1	0	1	0	13
Επιτυχής ταξινόμηση	67%	73%	93%	80%	93%	87%

Πίνακας Δ4.1: Ποσοστά επιτυχούς ταξινόμησης

5. ΣΥΜΠΕΡΑΣΜΑΤΑ

Ο αλγόριθμος που παρουσιάστηκε στο κεφάλαιο αυτό δίνει ικανοποιητικά αποτελέσματα όσον αφορά την ταξινόμηση εκφράσεων υπό την προϋπόθεση ότι δεν υπάρχει ολική κίνηση του κεφαλιού κατά τη διάρκεια της έκφρασης κάτι το οποίο δεν είναι πάντα προφανές. Πάντως χρησιμοποιώντας διαδικασίες παρακολούθησης της κίνησης του κεφαλιού η ολική κίνηση μπορεί να αντισταθμιστεί. Με δεδομένη την επακριβή εκτίμηση της κίνησης των σημείων της επιφάνειας του προσώπου η μέθοδος είναι αρκετά εύρωστη γιατί η ταξινόμηση στηρίζεται στην τροχιά που διαγράφει το διάνυσμα των χαρακτηριστικών και όχι σε μεμονωμένα στιγμιότυπα που λαμβάνονται σε κάθε καρέ. Σε κάθε περίπτωση πάντως πρέπει να τονιστεί ότι η αποτελεσματική εκτίμηση της μετατόπισης των σημείων μεταξύ διαδοχικών καρέ δεν είναι καθόλου απλή υπόθεση.

BIBΛΙΟΓΡΑΦΙΑ

- [1]. M. Tekalp, "Face and 2-D Mesh Animation in MPEG-4," *Tutorial Issue On The MPEG-4 Standard, Image Communication Journal*, Elsevier, 1999.
- [2]. G. Faigin, "The Artist's Complete Guide to Facial Expressions," *Watson-Guptill*, New York, 1990.
- [3]. C. M. Whissel, "The dictionary of affect in language," R. Plutchnik and H. Kellerman (Eds) *Emotion: Theory, research and experience: vol 4, The measurement of emotions*. Academic Press, New York, 1989.
- [4]. P. Ekman and W. Friesen, "The Facial Action Coding System," *Consulting Psychologist Press*, San Francisco, CA, 1978.
- [5]. *EC TMR Project PHYSTA Report*, "Development of Feature Representation from Facial Signals and Speech," January 1999.
- [6]. R. Plutchik, "Emotion: A Psychoevolutionary Synthesis," *Harper and Row*, New York, 1980.
- [7]. F. Parke and K. Waters, "Computer Facial Animation," *A K Peters*, 1996.
- [8]. Kin-Man Lam and Hong Yan, "An Analytic-to-Holistic Approach for Face Recognition Based on a Single Frontal View," *IEEE Trans. on PAMI*, vol. 20, no. 7, July 1998.
- [9]. K. Karpouzis, N. Tsapatsoulis and S. Kollias, "Moving to Continuous Facial Expression Space using the MPEG-4 Facial Definition Parameter (FDP) Set," in *Proc. of the Electronic Imaging 2000*, San Jose, CA, USA, Jan. 2000.
- [10]. P. Chellapa, C. Wilson and S. Sirohey, "Human and Machine Recognition of Faces: A Survey," *Proc. IEEE*, vol. 83, no. 5, pp. 705-740, 1995.
- [11]. R. Brunelli and T. Poggio, "Face recognition: Features versus templates," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 15, No. 10, 1993.
- [12]. M. Kirby and L. Sirovich, "Application of the Karhunen-Loeve procedure for the characterization of human faces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 12, No. 1, 1990, pp. 103-108.
- [13]. M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, Vol. 3, pp.71-86, 1991.
- [14]. G.W. Cottrell and J. Metcalfe, "EMPATH: Face, emotion, and gender recognition using holons," In R.P. Lippman, J. Moody and D.S. Touretzky (Eds.) *Advances in Neural Information Processing Systems 3*, San Mateo: Morgan & Kaufman, pp. 564-571, 1991.
- [15]. A.J. O'Toole, H. Abdi, K.A. Deffenbacher and D. Valentin, "A low-dimensional representation of faces in the higher dimensions of the space," *Journal of the Optical Society of America A*, Vol. 10, 405-411, 1993.
- [16]. P.N. Belhumeur, J.P. Hespanha and D.J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," *In Proceedings of the European Conference on Computer Vision*, 1996.
- [17]. P.J.B. Hancock, V. Bruce and A.M. Burton, "Testing principle component representations for faces," *Proc. 4th Neural Computation and Psychology Workshop*, London, 9-11 April 1997, Eds. J.A. Bullinaria, D.W. Glasspool, and G. Houghton, pp. 84-97, London, Springer, 1997.
- [18]. B. Moghaddam, W. Wahid and A. Pentlant, "Beyond Eigenfaces: Probabilistic Matching for Face Recognition," *Proc. 3rd IEEE Intl Conf. on Automatic Face and Gesture Recognition*, Nara, Japan, 1998.
- [19]. M.S. Gray, D.T. Lawrence, B.A. Golomb and T.J. Sejnowski, "A perceptron reveals the face of sex," *Neural Computation*, Vol. 7, No. 6, pp. 1160-1164, 1995.
- [20]. N. Intrator, D. Reisfeld, and Y. Yeshurun, "Face Recognition using a Hybrid Supervised/ Unsupervised Neural Network," *Pattern Recognition Letters*, Vol. 17, pp. 67-76, 1996.

- [21]. J. Daugman, "Complete discrete 2-D Gabor transforms by neural networks for image analysis and compression," *IEEE Transactions on Acoustic, Speech and Signal Processing*, Vol. 36, No. 7, pp. 1169-1179, 1988.
- [22]. J. Buhmann, J. Lange and C. von der Malsburg, "Distortion invariant object recognition by mat-ching hierarchically labeled graphs," *In IJCNN International Conference on Neural Networks*, Washington, DC, Vol. 1, pp. 155-159, 1989.
- [23]. Z. Zhang, M. Lyons, M. Schuster and S. Akamatsu, "Comparison between geometry-based and Gabor-wavelets-based facial expression recognition using multi-layer perceptrons," *In Proceedings of the 3rd IEEE International Conference on Automatic Face & Gesture Recognition*, Nara, Japan, 1998.
- [24]. I. Essa, T. Darrell and A. Pentland, "Tracking facial motion," *Proceedings of the Workshop on Motion of Nonrigid and Articulated Objects*, IEEE Computer Society, pp. 36-42, 1994.
- [25]. Y. Yacoob and L.S. Davis, "Recognizing human facial expressions from long image sequences using optical flow," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 18, No. 6, pp. 636-642, 1996.
- [26]. C. Padgett and G.W. Cottrell, and R. Adolphs, "Categorical perception in facial emotion classification," *manuscript*, 1998.
- [27]. M. Stewart Bartlett, P.A. Viola, T.J. Sejnowski, B.A. Golomb, J. Larsen, J.C. Hager, P. Ekman, "Classifying facial action," *Advances in Neural Information Processing Systems 8*, D. Touretzky, M. Mozer, and M. Hasselmo (Eds.), MIT Press, Cambridge, MA, 1996.
- [28]. J. Lien, T. Kanade, J.F. Cohn and C.C. Li, "A multi-method approach for discriminating between similar facial expressions, including expression intensity estimation," *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 853-859, 1998.
- [29]. R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz and J. Taylor, "Emotion recognition in human-computer inter-action," *to appear in the IEEE Signal Processing Magazine*, January 2001.
- [30]. H. Boulard and Y. Camp, "Autoassociation by Multilayer Perceptrons and Singular Value Decomposition," *Biological Cybern*, Vol. 59, pp. 291-294, 1988.
- [31]. I. Essa and A. Pentland, "Coding, Analysis, Interpretation, and Recognition of Facial Expressions," M.I.T. Media Laboratory, Perceptual Computing Group Report No. 325, 1995.
- [32]. N. Tsapatsoulis, N. Doulamis, A. Doulamis, and S. Kollias "Face Extraction from Non-uniform Background and Recognition in Compressed Domain," *Proc. of ICASSP'98*, Seattle WA, May 1998.
- [33]. J.N. Bassili, "Emotion Recognition: The role of facial movement and the relative importance of upper and lower areas of the face," *Journal of Personality and Social Psychology*, vol. 37, 2049-2059, 1979.
- [34]. L.R. Rabiner, "A tutorial on HMM and Selected Applications in Speech Recognition," *IEEE Proc.*, vol.77, no. 2, 1989.
- [35]. T. Otsuka and J. Ohya "Recognizing Multiple Persons' Facial Expressions Using HMM Based on Automatic Extraction of Significant frames from Image Sequences," *Journ. Cognitive Neuroscience*, vol. 3, no. 1, 1991.
- [36]. C. Kobayashi and J. Hara, "Dynamic Recognition of Basic Facial Expressions by Discrete time Recurrent Neural Networks," *Proc. of International Joint Conference on Neural Networks*, pp.155-158, 1993.
- [37]. "Review of existing techniques for human emotion understanding and their applications in human-computer interaction," *Technical Report*, Research contract FMRX-CT97-0098 (DG12-BDNC), October 1998.

ΕΠΙΛΟΓΟΣ - ΣΥΜΠΕΡΑΣΜΑΤΑ

Στην παρούσα διατριβή παρουσιάστηκε με συστηματικό τρόπο η προσέγγιση των θεμάτων της αναγνώρισης προσώπου και ανάλυσης εκφράσεων στο πλαίσιο εφαρμογών επικοινωνίας ανθρώπου μηχανής. Στο πρώτο μέρος της διατριβής πραγματοποιήθηκε μια σφαιρική καταγραφή των ανωτέρω προβλημάτων από την οποία ο αναγνώστης μπορεί να βρει επιπλέον πηγές για την ενημέρωση του. Με βάση τα στοιχεία που παρατίθενται γίνεται φανερό ότι αμφότερα τα προβλήματα της αναγνώρισης προσώπου και της ανάλυσης εκφράσεων δεν μπορούν να επιλυθούν –τουλάχιστον με τις υπάρχουσες συνθήκες- σε ένα συνολικό πλαίσιο. Η επικοινωνία ανθρώπου μηχανής όπως ορίστηκε στο Κεφάλαιο Α1 θέτει ικανούς περιορισμούς για την αντιμετώπιση τους τουλάχιστον μέχρι ενός σημείου. Συγκεκριμένα στο πλαίσιο της επικοινωνίας ανθρώπου υπολογιστή στο στενό εργασιακό και οικιακό χώρο πολλά επιμέρους ζητήματα απλουστεύονται:

Η αναγνώριση προσώπου είναι ευκολότερη με δεδομένο τον μικρό αριθμό προσώπων. Επιπλέον ο υπολογιστής έχει συνεχή επαφή με τα επιμέρους πρόσωπα και επομένως τη δυνατότητα να αναπροσαρμόζει τα δεδομένα – απεικονίσεις που διαθέτει για αυτά. Η διαδικασία ομοιάζει με την αντίστοιχη που επιτελείται από τα ανθρώπινα όντα Δύσκολα αντιλαμβανόμαστε τις σταδιακές αλλαγές σε πρόσωπα του στενού μας περιβάλλοντος γιατί η εικόνα τους στον εγκέφαλο μας διαρκώς αναπροσαρμόζεται. Σε περιβάλλον καθημερινής σχέσης και επικοινωνίας ο υπολογιστής διαθέτει ικανό αριθμό απεικονίσεων για το ίδιο άτομο –με γυαλιά, γενειάδα, διαφορετικό χρώμα και κτένισμα μαλλιών και. Όσον αφορά το στενό τεχνικό επίπεδο, ένα συγκεκριμένο περιβάλλον επικοινωνίας ανθρώπου υπολογιστή οδηγεί σε αποτελεσματική αντιμετώπιση προβλημάτων όπως οι συνθήκες φωτισμού και το καλιμπράρισμα των συσκευών οπτικής επικοινωνίας. Σύμφωνα με το Κεφάλαιο Β2 η υπαρχή τέτοιου είδους πληροφoρίας αυξάνει την ευρωστία του συστήματος εντοπισμού προσώπων με κατάλληλη προσαρμογή των παραμέτρων του μοντέλου για την χρωματική δομή του δέρματος.

Η ανάλυση εκφράσεων, ως πιο σύνθετο πρόβλημα, χρειάζεται περισσότερο τους περιορισμούς του στενού οικιακού και εργασιακού χώρου. Όπως γίνεται αντιληπτό από το πρώτο μέρος της διατριβής, πέρα από τα καθαρά τεχνικής φύσεως προβλήματα που διέπουν την ανάλυση εκφράσεων –παρακολούθηση της κίνησης του προσώπου και των σημείων του- η συναισθηματική έκφραση είναι εξαιρετικά ποικιλόμορφη στο πλάτος του ανθρώπινου πληθυσμού. Ο τρόπος με τον οποίο εκφράζουμε τα συναισθήματα μας είναι αρκετά προσωπικός. Υπό το πρίσμα αυτό η εκτίμηση των συναισθημάτων μέσα από τις εκφράσεις μπορεί να γίνει μόνο σε μια per person βάση και είναι εφικτή σε ένα περιορισμένο κύκλο προσώπων. Επιπλέον, όπως και οι άνθρωποι, ο υπολογιστής πρέπει να μπορεί να μαθαίνει σταδιακά τον τρόπο αντίδρασης των επιμέρους ατόμων. Το γεγονός αυτό είναι σύμφωνο με την χρήση κανόνων απόφασης στη διαδικασία ανάλυσης εκφράσεων που υιοθετήθηκε στο Κεφάλαιο Δ3.

Η δομή της διατριβής δίνει την ικανότητα στον αναγνώστη να κινηθεί μεθοδικά προς την πρακτική υλοποίηση ενός συστήματος αναγνώρισης προσώπων και ανάλυσης εκφράσεων έχοντας υπόψη τις προηγούμενες παρατηρήσεις. Σε κάθε επιμέρους πρόβλημα προτείνεται ένας σαφώς προσδιορισμένος τρόπος αντιμετώπισης. Οι προτεινόμενοι αλγόριθμοι είναι αυτοί που κρίθηκαν από τον συγγραφέα ως οι καταλληλότεροι και οι πλέον αξιόπιστοι με βάση το πρόβλημα. Η καταλληλότητα εξασφαλίζεται είτε με την προσαρμοσμένη στο πρόβλημα σχεδίαση, είτε -όταν πρόκειται για γνωστούς αλγορίθμους- με αποτελεσματική παρέμβαση ή και αναδιάρθρωση. Η αξιοπιστία βασίζεται σε σωρεία προσομοιώσεων που πραγματοποιήθηκαν από τον συγγραφέα και οποίες είτε είχαν συγκριτική υφή είτε προσπαθούσαν να εντοπίσουν τα όρια ανοχής των αλγορίθμων.

Στη διατριβή ενσωματώθηκαν και μερικά κεφάλαια στα οποία γίνεται νύξη σε άλλες εφαρμογές πλην της επικοινωνίας ανθρώπου μηχανής. Ο αναγνώστης μπορεί να τα χρησιμοποιήσει ως βάση για πιθανή μελλοντική εργασία σε παρεμφερείς τομείς. Αν έπρεπε υποχρεωτικά να επιλέξω κάποιο τμήμα της διατριβής στο οποίο αξίζει να δοθεί έμφαση σε μελλοντικές εργασίες θα επέλεγα τον εντοπισμό συγκεκριμένων σημείων -όχι χαρακτηριστικών- στο πρόσωπο. Ένα μεγάλο τμήμα της διατριβής αφιερώθηκε στη μοντελοποίηση εκφράσεων με βάση κάποια του προσώπου και την ορολογία του MPEG-4 στο σχετικό τομέα. Η ανάπτυξη προηγμένων τεχνικών για την ανίχνευση των σημείων αυτών είναι βέβαιο πως θα βρει πολύπλευρη εφαρμογή. Στο θέμα αυτό ίσως δεν δόθηκε η απαιτούμενη έμφαση δεδομένου ότι αποτελεί το θέμα της διατριβής άλλων συναδέλφων στο Εργαστήριο Ψηφιακής Επεξεργασίας Εικόνας, Βίντεο και Πολυμέσων.

Ολοκληρώνοντας το τμήμα αυτό της διατριβής θα ήθελα να διατυπώσω δύο συμπεράσματα τα οποία προήλθαν από την πεντάχρονη ενασχόληση μου με το συγκεκριμένο αντικείμενο: (α) Στο πλαίσιο της διατριβής δεν δόθηκε έμφαση μόνο στο επιστημονικό περιεχόμενο των λύσεων για τα επιμέρους θέματα αλλά προτάθηκαν λύσεις που αφορούν στους μηχανικούς, είναι δηλαδή άμεσα εφαρμόσιμες, (β) η επίλυση ενός σύνθετου προβλήματος προσεγγίζεται αποτελεσματικότερα από απλές επιμέρους λύσεις παρά από μια σύνθετη ολική λύση.

Ο ΜΕΤΑΣΧΗΜΑΤΙΣΜΟΣ KARHUNEN-LOEVE-ΙΔΙΟΔΙΑΝΥΣΜΑΤΑ

Ο μετασχηματισμός Karhunen-Loeve εφαρμόστηκε με ιδιαίτερη επιτυχία στην αναγνώριση προσώπων και είναι γνωστός ως τεχνική των eigenfaces. Για μεγάλες βάσεις δεδομένων που περιέχουν $O(10^3)$ πρόσωπα έχουν παρουσιάσει πολύ καλή απόδοση. Βασισμένη στην ίδια ιδέα η ανίχνευση χαρακτηριστικών με βάση τον μετασχηματισμό Karhunen-Loeve, μια τεχνική που είναι γνωστή ως eigenfeatures.

Έστω λοιπόν μια συλλογή από M πρόσωπα με $m \times n$ διαστάσεις το καθένα. Κάθε εικόνα E_i όπου $i=1,2,3,\dots,M$ αντιστοιχίζεται σε ένα διάνυσμα X_i διάστασης $1 \times MN$. Σχηματίζεται κατόπιν ο εκτιμώμενος πίνακας αυτοσυσχέτισης ως ακολούθως:

$$\hat{R} = \sum_{i=1}^M (X_i - X_\mu) \cdot (X_i - X_\mu)^T \quad [6]$$

$$\text{όπου } X_\mu = \frac{1}{M} \sum_i^M X_i \quad [7]$$

εκφράζει το μέσο διάνυσμα (πρόσωπο). Το γεγονός ότι η διάσταση του πίνακα είναι $MN \times MN$ δυσχεραίνει την επεξεργασία του και αποτελεί μειονέκτημα. Ακολουθεί διαγωνοποίηση της εκτιμήτριας

$$\hat{R} = U \cdot \Lambda \cdot U^T \quad [8]$$

όπου Λ είναι διαγώνιος πίνακας που περιέχει τις ιδιοτιμές της εκτιμήτριας λ_i . Προφανώς ο πίνακας U περιέχει τα ιδιοδιανύσματα της εκτιμήτριας:

$$U = [u_1 | u_2 | \dots | u_{mn}] \quad [9]$$

όπου υπάρχει μια αντιστοιχία ανάμεσα στα ιδιοδιανύσματα και τις ιδιοτιμές. Ως πιο σημαντικά ιδιοδιανύσματα θεωρούνται αυτά που αντιστοιχούν στις πιο μεγάλες ιδιοτιμές. Έτσι επιλέγοντας τα σημαντικότερα ιδιοδιανύσματα, μειώνουμε το πλήθος τους, ενώ παράλληλα μπορούμε μια σχετικά πιστή αναπαράσταση των εικόνων, αφού έχουμε κρατήσει μόνο τους συντελεστές των ιδιοδιανυσμάτων αυτών. Οι συντελεστές προκύπτουν από την προβολή της εικόνας στα αντίστοιχα ιδιοδιανύσματα. Η αναπαραγωγή των διανυσμάτων X_i γίνεται σύμφωνα με την σχέση:

$$X = a_1 u_1 + a_2 u_2 + \dots \quad [10]$$

και με βάση το γεγονός ότι τα ιδιοδιανύσματα αυτά είναι γνωστά, η αναπαράσταση των εικόνων γίνεται με απλή αντιστοίχιση τους με τα βάρη παραγωγής.

$$X \Rightarrow [a_1 \ a_2 \ \dots \ a_k] \quad [11]$$

Η παραπάνω διαδικασία όπως έχει αναφερθεί είναι δυνατόν να χρησιμοποιηθεί τόσο για την αναγνώριση ολόκληρου του προσώπου με την μέθοδο eigenfaces όσο και για αυτήν των επιμέρους χαρακτηριστικών με την μέθοδο eigenfeatures. Στην εργασία περιγράφεται μια επέκταση της τεχνικής των ιδιοπροσώπων σε κατηγοριοποίηση πολλαπλών ομάδων ιδιοδιανυσμάτων, που σχετίζονται με διαφοροποιήσεις κλίμακας και προσανατολισμού των προσώπων. Κατά αυτόν τον διαχωρισμό που βασίζεται σε διαφορετικές οπτικές γωνίες το πρώτο που πρέπει να κάνει κάποιος είναι η ανίχνευση της θέσης και ο καθορισμός του γενικού προσανατολισμού του προς αναγνώριση προσώπου, ώστε να γίνει η επιλογή της κατάλληλης αποθηκευμένης ομάδας ιδιοδιανυσμάτων. Ο κατάλληλος χώρος επιλέγεται μέσω κάποιος μετρικής απόστασης από αυτόν. Ακολουθεί κωδικοποίηση και αναγνώριση μέσω ταιριάσματος με τον κοντινότερο γείτονα. Μεγαλύτερο ενδιαφέρον παρουσιάζουν τεχνικές περιγραφής και κωδικοποίησης χαρακτηριστικών που είναι γνωστά ως ιδιοχαρακτηριστικά. Καλύτερη επίδοση στην διαδικασία της αναγνώρισης μπορεί να επιτευχθεί με εμπλουτισμό με ιδιοχαρακτηριστικά της παράστασης με ιδιοπρόσωπα. Κατά κάποιον τρόπο λοιπόν η τραχεία περιγραφή ενός προσώπου μπορεί να λειανθεί με την παρουσία των ιδιοχαρακτηριστικών. Μια εναλλακτική τεχνική του ταιριάσματος με πρότυπο είναι η ανίχνευση χαρακτηριστικών με χρήση ιδιοπροτύπων (eigen-templates). Με αυτή την τεχνική επιτυγχάνεται μεγαλύτερη ανοχή σε αλλοιώσεις από σήματα θορύβου. Ο συνδυασμός ιδιοπροσώπων και ιδιοχαρακτηριστικών παρουσιάζει μεγαλύτερη ανθεκτικότητα από την κάθε μέθοδο μόνη της και δίνει υψηλά ποσοστά επιτυχίας κατά την αναγνώριση.

Η μέθοδος eigenfeatures είναι ιδιαίτερα αποτελεσματική και παρόλο που είναι παρόμοια με την τεχνική ταιριάσματος με πρότυπο είναι σημαντικά ταχύτερη και πιο ευσταθής. Επιπλέον έχει ένα πολύ σοβαρό θεωρητικό υπόβαθρο. Εντούτοις παρουσιάζει και κάποια σημαντικά μειονεκτήματα.

Απαιτεί την στατιστική εκτίμηση του πίνακα συμμεταβλητότητας η οποία για να πραγματοποιηθεί αποτελεσματικά χρειάζεται ένα μεγάλο αριθμό από δείγματα (στην περίπτωση μας χαρακτηριστικά προσώπου).

Επίσης είναι ιδιαίτερα επιρρεπής σε μεταβολές της κλίσης του προσώπου ως προς τους τρεις άξονες. Τέλος παραμένει σαν τεχνική αρκετά υπολογιστικά πολύπλοκη και ως εκ τούτου ακατάλληλη για εφαρμογές πραγματικού χρόνου.

ΤΑ ΦΙΛΤΡΑ GABOR

Τα φίλτρα Gabor μπορούν να περιγράψουν την περιοχή που περιβάλλει ένα pixel μιας εικόνας. Αυτό γίνεται αν χρησιμοποιήσουμε μια ομάδα από φίλτρα διαφορετικών συχνοτήτων και προσανατολισμών τα οποία έχουν ως αναφορά το συγκεκριμένο pixel. Κατά αυτόν τον τρόπο δίνουν την δυνατότητα σχηματισμού ενός διανύσματος χαρακτηριστικών που δημιουργείται από τις αποκρίσεις τους.

Η ιδιότητα τους αυτή τα έχει καταστήσει τα έχει καταστήσει ως βασική επιλογή για την ανίχνευση μεμονωμένων σημείων στα χαρακτηριστικά προσώπου. Για παράδειγμα μετά την ανίχνευση των ματιών με χρήση της τεχνικής ταιριάσματος με πρότυπο, απαιτείται εντοπισμός των σημείων που αντιστοιχούν στην κόρη του ματιού και στις γωνίες τους. Σχεδόν σε όλους τους αλγόριθμους ανίχνευσης χαρακτηριστικών σημείων προσώπου τα φίλτρα Gabor αποτελούν το τελικό στάδιο του αλγόριθμου.

Τα φίλτρα έχουν την ακολουθεί μορφή:

$$\Psi_{\underline{k}}(\underline{x}) = \frac{k^2}{\sigma^2} \cdot \exp\left(-\frac{k^2 \cdot \underline{x}^2}{2\sigma^2}\right) \cdot [\exp(-ikx) - \exp(\sigma^{-2})] \quad [12]$$

όπου \underline{x} είναι η θέση της εικόνας, \underline{k} είναι η κεντρική συχνότητα, σ είναι το πλάτος του γκαουσιανού παραθύρου του πλάτους της συνάρτησης και $\underline{k}/\sigma = \alpha$. Το διάνυσμα χαρακτηριστικών έχει σαν στοιχεία του τις αποκρίσεις των φίλτρων στις διάφορες κεντρικές συχνότητες για μια συγκεκριμένη θέση. Η διαφορά φάσης σε ένα στοιχείο του διανύσματος κατά την σύγκριση των δυο εικόνων, μεταφράζεται σε τοπική διαφοροποίηση ως προς την αντίστοιχη κατεύθυνση. Οι διαφοροποιήσεις προς κάθε κατεύθυνση εκτιμούνται και συνδυάζονται για την δημιουργία μιας κοινής διαφοροποίησης θέσης με βάση το κριτήριο του ελαχίστου τετραγωνικού σφάλματος. Μόνο στα σημεία τα οποία ισχύει:

$$\xi = \sqrt{\sum_j a_j^2} > \xi_{TH} \quad [13]$$

λαμβάνεται υπόψη η διαφοροποίηση θέσης όπου j είναι ο δείκτης που κινείται στα στοιχεία ενός διανύσματος και a δείχνει το πλάτος των στοιχείων. Η απόκριση φάσης στην χαμηλότερη χωρική συχνότητα μας δίνει την αρχική εκτίμηση της διαφοράς θέσης ενός χαρακτηριστικού σημείου. Η θέση του χαρακτηριστικού σημείου επανατίθεται βάση αυτής της εκτίμησης και η διαδικασία επαναλαμβάνεται κατά τις υψηλότερες χωρικές συχνότητες προσδίδοντας συνεχώς μεγαλύτερη ακρίβεια. Η διαφορά θέσης που προκύπτει αντιστοιχίζεται με ένα βαθμό εμπιστοσύνης μεταξύ των δυο συγκρινόμενων διανυσμάτων. Υψηλής εμπιστοσύνης διαφορές θέσης χρησιμοποιούνται για τον εντοπισμό των σημείων. Προκειμένου να ανιχνεύσουμε χαρακτηριστικά σημεία της εικόνας μπορούμε να εφαρμόσουμε την τεχνική αυτή

συγκρίνοντας την εκάστοτε εικόνα με πρότυπο. Επίσης η τεχνική αυτή μπορεί να χρησιμοποιηθεί σε εκτίμηση κίνησης προσώπου σε video εικόνας προσώπου, καθώς και κατά την απεικόνιση των προσώπων με γράφους των οποίων οι κόμβοι φέρουν περιγραφή των αντίστοιχων αποκρίσεων των φίλτρων στις διάφορες συχνότητες και οι ακμές περιγράφουν τοπογραφικές σχέσεις. Η τελευταία χρήση σχετίζεται με ταίριασμα ελαστικών γράφων και βρίσκει εφαρμογή στην αναγνώριση και την απομόνωση προσώπων η άλλων αντικειμένων σε σύνθετες σκηνές. Μια άλλη μορφή της διδιάστατης συνάρτησης Gabor είναι:

$$g(x, y; u_0, v_0) = \exp \left\{ - \left[\frac{x^2}{2\sigma_x^2} + \frac{y^2}{2\sigma_y^2} \right] + 2\pi j [u_0 x + v_0 y] \right\} \quad [14]$$

όπου τα σ_x και σ_y συμβολίζουν τα χωρικά πλάτη της γαουσιανής περιβάλλουσας και στην ουσία ρυθμίζουν την επιλεκτικότητα των αξόνων. Επίσης (u_0, v_0) είναι η εκάστοτε κεντρική συχνότητα της μιγαδικής ημιτονοειδούς. Οι συναρτήσεις Gabor σχηματίζουν μια πλήρη, μη ορθογωνική βάση. Η αναπαράσταση αυτών σε μορφή wavelet είναι:

$$\Phi_\lambda(x, y, \theta) = \exp \left\{ - \lambda^2 (x'^2 + y'^2) + j\pi x' \right\} \quad [15]$$

όπου $x' = x \cdot \cos\theta + y \cdot \sin\theta$ και $y' = x \cdot \sin\theta + y \cdot \cos\theta$. Η παράμετρος θ δείχνει τον προσανατολισμό του φίλτρου ως προς τον χώρο. Ο προσανατολισμός αφορά την εξαγωγή χαρακτηριστικών ως προς κάποια επιθυμητή κατεύθυνση.

Η παράμετρος προσανατολισμού θ καθορίζει την κατεύθυνση των ανιχνεύσιμων ακμών. Υπάρχει ιδιαίτερη ευαισθησία σε ακμές μικρού μήκους, ενώ παράλληλα μειώνεται η απόκριση για αύξηση των μηκών των γραμμών. Τα χαρακτηριστικά εκτοπίζονται, δηλαδή επιλέγονται μέσω μιας διαδικασίας εύρεσης τοπικών μεγίστων. Υπάρχει και μια τρίτη μορφή του φίλτρου:

$$\Phi_\lambda(x, y, \theta) = \exp \left\{ - \lambda^2 \cdot (x^2 + y^2) + j\pi \cdot (x \cdot \cos\theta + y \cdot \sin\theta) \right\} \quad [16]$$

Συγκρίνοντας την προκύπτουσα σχέση με την δεύτερη των μορφών του φίλτρου εξάγουμε το συμπέρασμα:

$$\cos\theta = 2u_0 \quad \sin\theta = 2v_0 \quad [17]$$

δηλαδή η επιλογή της εκάστοτε συχνότητας καθορίζει και τον προσανατολισμό του φίλτρου.

Ένα άλλο σημαντικό συμπέρασμα για τις συναρτήσεις Gabor διατυπώνεται παρακάτω. Για την εκθετική συνάρτηση που εκφράζει την περιβάλλουσα ισχύει η ιδιότητα ότι το γινόμενο του πεδίου ορισμού της συνάρτησης με το εύρος ζώνης του μετασχηματισμού Fourier είναι ελάχιστο. Αυτό συνεπάγεται βέλτιστο εντοπισμό χαρακτηριστικών, δηλαδή η διακριτική ικανότητα γίνεται βέλτιστη τόσο ως προς τον φυσικό χώρο όσο και ως προς τον χώρο των συχνοτήτων ταυτόχρονα.

ΤΟ ΠΡΟΤΥΠΟ MPEG-4 ΚΑΙ Η ΜΟΝΤΕΛΟΠΟΙΗΣΗ ΤΟΥ ΠΡΟΣΩΠΟΥ

ΕΙΣΑΓΩΓΗ

Η εισαγωγή ενός προτύπου για την κωδικοποίηση ακολουθιών video καθώς και η ευφύστερη χρησιμοποίηση του ήδη υπάρχοντος υλικού καθίσταται πλέον αναγκαία στην εποχή μας καθώς συνηγορούν προς αυτή την κατεύθυνση τόσο οι εξελίξεις στις μεθόδους παραγωγής οπτικοακουστικού υλικού όσο και οι εξελίξεις στον εξοπλισμό μετάδοσης και κωδικοποίησης.

Το πρότυπο MPEG-4 είναι αυτό που επιχειρεί να καλύψει την συγκεκριμένη ανάγκη. Η σημαντικότερη αλλαγή που επιφέρει στην αντιμετώπιση του προβλήματος της κωδικοποίησης εντοπίζεται στο γεγονός ότι σαν δομικά στοιχεία της ακολουθίας video θεωρούνται πλέον τα οπτικοακουστικά αντικείμενα και όχι τα απλά πλαίσια εικόνων. Πλέον η ισχυρή αυτοσυσχέτιση στην πληροφορία περνάει σε δεύτερο ρόλο όχι τόσο εξαιτίας των διαδικασιών πρόβλεψης της κίνησης, αλλά κυρίως λόγω της περιγραφής χωρικών και χρονικών συσχετίσεων μέσω οπτικοακουστικών αντικειμένων, ή εναλλακτικά μέσω της χρονικής και χωρικής εξέλιξης των ιδιοτήτων ενός οπτικοακουστικού αντικειμένου.

Η προσέγγιση αυτή επιτρέπει την υποστηρίξη εξαιρετικά διαλογικών εφαρμογών πέρα από τα όρια της ταχείας αναζήτησης σε μια ακολουθία. Έτσι σε μια ακολουθία ο χρήστης μπορεί να επιλέξει το υπόβαθρο για τον τηλεπαραουσιαστή να τροποποιήσει μια σκηνή ή να χρησιμοποιήσει οποιοδήποτε αντικείμενο σαν υπερσύνδεσμο για αναζήτηση συναφούς πληροφορίας.

Το MPEG-4 έχει τις παρακάτω πολύ σημαντικές ιδιότητες:

- 1) Είναι κατά κύριο λόγο πρότυπο αποκωδικοποίησης δηλαδή περιγράφει ένα συντακτικό με το οποίο πρέπει να είναι συμβατή η προς αποκωδικοποίηση ακολουθία, αφήνοντας σημαντική ελευθερία στην προ-επεξεργασία, σύνθεση και κωδικοποίηση του οπτικοακουστικού υλικού.
- 2) Είναι πρότυπο ενδογενώς κλιμακούμενο επιτρέποντας έτσι την ιεραρχική κωδικοποίηση πληροφοριών οι οποίες μεταδίδονται από τα τηλεπικοινωνιακά μέσα μεγάλης αξιοπιστίας. Επιπρόσθετα στην χάραξη κατευθυντήριων γραμμών για το πρότυπο MPEG-4 έχει ληφθεί ιδιαίτερη μνεία για την δυνατότητα ανάληψης του αποκωδικοποιητή από διακοπή, καθυστέρηση ή παραμόρφωση τμήματος της ακολουθίας.

Οι σημαντικότερες προτεραιότητες που έχουν τεθεί στην διαμόρφωση του προτύπου MPEG-4 είναι:

- Πρόσβαση περιεχομένου μέσω εργαλείων πολυμέσων
- Διαχείριση και τροποποίηση ακολουθιών βάση περιεχομένου

- Κωδικοποίηση φυσικών, υβριδικών και συνθετικών σημάτων
- Βελτιστοποιημένη τυχαία πρόσβαση
- Αποτελεσματική κωδικοποίηση
- Συγχρονισμένη κωδικοποίηση πολλαπλών σημάτων
- Ανθεκτικότητα στην μετάδοση μέσω «θορυβωδών» δικτύων
- Κλιμακούμενη κωδικοποίηση βάση περιεχομένου

ΤΟ MPEG-4 ΚΑΙ Η ΑΝΑΓΝΩΡΙΣΗ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ

Το σύνολο παραμέτρων για τον ορισμό ενός προσώπου (Facial Definition Parameter set FDP) και το σύνολο παραμέτρων αναπαράστασης προσώπου (Facial Animation Parameter set FAP) έχουν σχεδιαστεί για να επιτρέπουν τον ορισμό του σχήματος και της υψής ενός προσώπου καθώς και την αναπαράσταση εκφράσεων, συναισθημάτων καθώς και προφοράς ομιλίας. Το FAP set εάν ερμηνευτεί σωστά παράγει παρόμοια αποτελέσματα με το πραγματικό πρότυπο σε ότι αφορά εκφράσεις προφορά ομιλίας σε διαφορετικά μοντέλα προσώπων, χωρίς να χρειάζεται να αρχικοποιήσουμε ή να ρυθμίσουμε το μοντέλο.

Το FDP set επιτρέπει τον ορισμό ενός ακριβούς σχήματος προσώπου καθώς και της υψής του στην φάση του setup. Εάν αυτά τα στοιχεία χρησιμοποιηθούν στην φάση της αρχικής ρύθμισης είναι επίσης πιθανόν να παράγουμε με ακόμα μεγαλύτερη ακρίβεια τις κινήσεις του συγκεκριμένου μοντέλου προσώπου.

Αναλυτικά για το FAPs

Το σύνολο παραμέτρων αναπαράστασης προσώπου (FAPs) είναι βασισμένο στην μελέτη των ελάχιστων αντιδράσεων του προσώπου και είναι στενά συνδεδεμένο με τις αντιδράσεις των μυών. Αποτελούν ένα πλήρες σύνολο από βασικές αντιδράσεις του προσώπου και ως εκ τούτου επιτρέπουν την αναπαράσταση των πιο φυσικών αντιδράσεων του προσώπου. Υπερβολικές τιμές βοηθούν στον ορισμό αντιδράσεων που δεν είναι φυσιολογικές για το ανθρώπινο πρόσωπο αλλά μπορούν να βρουν χρησιμότητα στον σχεδιασμό κινουμένων σχεδίων.

Όλες οι παράμετροι που αφορούν την μετάφραση κίνησης εκφράζονται σύμφωνα με τις Μονάδες Παραμέτρων Αναπαράστασης Προσώπου (Facial Animation Parameters Units). Αυτές οι μονάδες έχουν σχεδιαστεί έτσι ώστε να επιτρέπουν την απόδοση των παραμέτρων αναπαράστασης σε κάθε μοντέλο προσώπου με έναν συνεχή τρόπο παράγοντας λογικά αποτελέσματα σε ότι αφορά τις εκφράσεις και την ανθρώπινη προφορά. Ανταποκρίνονται στον τεμαχισμό των αποστάσεων ανάμεσα σε κάποια χαρακτηριστικά κλειδιά του ανθρώπινου προσώπου. Το μέγεθος του τεμαχισμού που εφαρμόζεται επιλέγεται ανάλογα με την επιθυμητή ακρίβεια. Ο παρακάτω πίνακας περιέχει το σύνολο των παραμέτρων. Για κάθε παράμετρο η λίστα περιέχει το όνομα της μια σύντομη περιγραφή, το κατά πόσο η παράμετρος αυτή παίρνει μόνο θετικές ή και αρνητικές τιμές, ορισμό της κατεύθυνσης κίνησης για θετικές τιμές, καθώς και ορισμό των μονάδων μέτρησης. Οι μονάδες μέτρησης μπορούν να είναι είτε μοίρες είτε κάποιο από το

FAPUs όπως φαίνεται στο σχήμα 2.1. Κάθε παράμετρος ορίζεται με όρους έντασης που παίρνουν τιμές από 1 έως 10.

Το σύνολο παραμέτρων περιέχει 3 υψηλού επιπέδου παραμέτρους:

Αναλυτικά για το FDPs

Το σύνολο παραμέτρων για τον ορισμό προσώπου (FDPs) χρησιμοποιείται ώστε να αντιστοιχίζει ένα δοθέν μοντέλο προσώπου σε ένα συγκεκριμένο πρόσωπο. Οι παράμετροι ορισμού προσώπου συνήθως αποστέλλονται μια φορά και ακολουθούνται από ένα «κύμα» παραμέτρων αναπαράστασης προσώπου. Εντούτοις ακόμα και αν ο αποκωδικοποιητής δεν λάβει τις παραμέτρους ορισμού προσώπου (FDPs) η χρήση των FAPUs κάνει σίγουρο το ότι θα ερμηνευτούν σωστά οι παράμετροι αναπαράστασης προσώπου (FAP). Αυτό εγγυάται ελάχιστη λειτουργία σε εφαρμογές αναμετάδοσης ή τηλεσυνδιάσκεψης.

Το σύνολο παραμέτρων ορισμού προσώπου περιέχει τα παρακάτω:

- Τρισδιάστατο δικτύωμα (με συντεταγμένες υψής αν χρησιμοποιείται υφή)
- Σημεία τρισδιάστατου πρότυπου
- Εικόνα υψής (προαιρετική)
- Λοιπά (μαλλιά, γυαλιά, ηλικία, γένος) (προαιρετικά)

Το τρισδιάστατο δικτύωμα χρησιμοποιείται για να ορίσει το σχήμα του προσώπου. Μπορεί να περιέχει και συντεταγμένες υψής εάν η υφή εφαρμόζεται. Το σύνολο των 3D χαρακτηριστικών σημείων χρησιμοποιείται για να εντοπίσει τα χαρακτηριστικά του προσώπου στο τρισδιάστατο σχήμα. Τα σημεία αυτά παρουσιάζονται στο παρακάτω σχήμα.

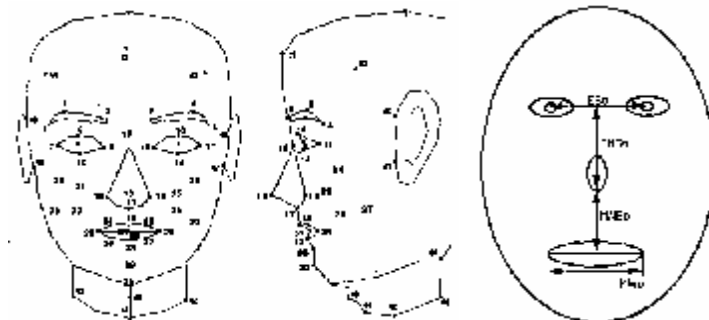
Πρέπει επίσης να τονιστεί ότι το ενιαίο πλαίσιο των χαρακτηριστικών σημείων όπως ορίζονται από το MPEG-4, μπορεί να χρησιμοποιηθεί εκτός από την μοντελοποίηση συνθετικών προσώπων και σε κλασσιές εφαρμογές όπως η αναγνώριση προσώπων, συναισθημάτων και εκφράσεων (βλ. [11]).

FAP name	FAP description	Bidirectional /Unidirectional	Direction of movements for positive intensities	Measurement units
“move_h_l_eyeball”	Horizontal movement of the left eyeball	B	Right	Degrees
“move_h_r_eyeball”	Horizontal movement of the right eyeball	B	Right	Degrees
“move_v_l_eyeball”	Vertical movement of the left eyeball	B	Downward	Degrees
“move_v_r_eyeball”	Vertical movement of the right eyeball	B	Downward	Degrees
?enlarge_l_pupil?	Enlarge th left pupil	B	Outward	ES
?enlarge_r_pupil?	Enlarge th right pupil	B	Outward	ES
“close_upper_l_eyelid”	Vertical movement of the upper left eyelid	B	Downward	ENS
“close_upper_r_eyelid”	Vertical movement of the upper right eyelid	B	Downward	ENS
“close_lower_l_eyelid”	Vertical movement of the lower left eyelid	B	Downward	ENS
“close_lower_r_eyelid”	Vertical movement of the lower right eyelid	B	Downward	ENS
“move_o_l_eyeball”	Outside/inside movement o the left eyeball	B	Forward	ES
“move_o_r_eyeball”	Outside/inside movement o the right eyeball	B	Forward	ES
“raise_l_o_eyebrow”	Vertical movement of the outer part of the left eyebrow	B	Upward	ENS
“raise_r_o_eyebrow”	Vertical movement of the outer part of the right eyebrow	B	Upward	ENS
“raise_l_m_eyebrow”	Vertical movement of the middle part of the left eyebrow	B	Upward	ENS
“raise_r_m_eyebrow”	Vertical movement of the middle part of the right eyebrow	B	Upward	ENS
“raise_l_i_eyebrow”	Vertical movement of the inner part of the left eyebrow	B	Upward	ENS
“raise_r_i_eyebrow”	Vertical movement of the inner part of the right eyebrow	B	Upward	ENS
“squeeze_l_eyebrow”	Horizontal movement of the left eyebrow	U	Inwards	ES

“squeeze_r_eyebrow”	Horizontal movement of the right eyebrow	U	Inwards	ES
“stretch_l_nose”	Stretch/squeeze movement of the left nostril	B	Toward the edge of the face	ENS
“stretch_r_nose”	Stretch/squeeze movement of the right nostril	B	Toward the edge of the face	ENS
“raise_nose”	Vertical movement of the nose	U	Upward	ENS
“open_jaw”	Rotational movement of the jaw	U	Downward	MNS
“move_hori_jaw”	Left/Right movement of the jaw	B	Right	MNS
“move_fwd_jaw”	Forward movement of the jaw	B	Forward	MNS
“depress_chin”	Upward and compressing movement of the chin (like in sadness)	U	Upward	Intensity 1-10
“puff_l_chek”	Puffing movement of the left cheek	B	Left	ES
“puff_r_check”	Puffing movement of the right cheek	B	Right	ES
“lift_l_cheek”	Lifting movement of the left cheek	U	Upward	ENS
“lift_r_cheek”	Lifting movement of the right cheek	U	Upward	ENS
“lower_lowerlip”	Vertical movement of the lowerlip	U	Downward	MNS
“lower_l_cornerlip”	Vertical movement of th left corner of the lips	U	Downward	MNS
“lower_r_cornerlip”	Vertical movement of the right corner of the lips	U	Downward	MNS
“raise_l_cornerlip”	Vertical (upward) movement of the left corner of the lips	U	Upward	MNS
“raise_r_cornerlip”	Vertical (upward) movement of the right corner of the lips	U	Upward	MNS
“raise_u_midlip”	Vertical movement of the middle part of the upper lip	U	Upward	MNS
“raise_l_midlip”	Vertical movement of the middle part of the lower lip	U	Upward	MNS
“push_pull_upperlip”	Protruding or sucking movement of the upper lip	U	Forward	MNS
“push_pull_lowerlip”	Protruding or sucking movement of the lower lip	U	Backward	MNS
“stretch_l_cornerlip”	Stretch the left corner of the lips	U	Left	MW
“stretch_r_cornerlip”	Stretch the right corner of the lips	U	Right	MW
“squeeze_l_cornerlip”	Squeeze the left corner of the lips	U	Left	MW

"squeeze_r_cornerlip"	Squeeze the right corner of the lips	U	Right	MW
?move_h_tongue_tip?	Left/Right movemnt of the tip of the tongue	B	Right	MW
?raie_tongue_tip?	Vertical movement of the tip of the tongue	B	Up	MW
?proturd_tongue?	Horizontal protruding movement of the tongue body	B	Forward	MW
?raise_tongue?	Vertical movement of the tongue body	B	Up	MW
?tongue_roll?	Rolling of the tongue in the U shape	U	Towards the center of the face	Degrees
?raise_left_ear?	Raise the left ear	B	Up	ENS
?raise_right_ear?	Raise the right ear	B	Up	ENS
?pull_left_ear?	Pull left ear forward with possible rotation	U	Forward	ENS
?pull_right_ear?	Pull right ear forward with possible rotation	U	Forward	ENS

Πίνακας F3.1: Ο ορισμός των FAPs στο MPEG-4



Εικόνα F3.1: FDP points και αποστάσεις κανονικοποίησης του προτύπου MPEG-

ΔΕΙΚΤΕΣ ΟΡΩΝ

Α
Aristotle,3

