

Πανεπιστήμιο Κρήτης
Σχολή Θετικών Επιστημών
Τμήμα Επιστήμης Υπολογιστών

Παρακολούθηση προσώπων και εκτίμηση της πόζας σε ψηφιακές ακολουθίες εικόνων

Γεώργιος Σημαντήρης

Μεταπτυχιακή Εργασία

Ηράκλειο, Φεβρουάριος 2003

Παρακολούθηση προσώπων και εκτίμηση της πόζας σε ψηφιακές ακολουθίες εικόνων

Εργασία που υποβλήθηκε από τον
Γεώργιο Σημαντήρη
ως μερική εκπλήρωση των απαιτήσεων για την απόκτηση
ΜΕΤΑΠΤΥΧΙΑΚΟΥ ΔΙΠΛΩΜΑΤΟΣ ΕΙΔΙΚΕΥΣΗΣ

Συγγραφέας:

Γεώργιος Σημαντήρης
Τμήμα Επιστήμης Υπολογιστών
Πανεπιστήμιο Κρήτης

Εισηγητική Επιτροπή:

Γεώργιος Τζιρίτας, Αναπληρωτής Καθηγητής, Επόπτης

Στέλιος Ορφανουδάκης, Καθηγητής, Μέλος

Κωνσταντίνος Στεφανίδης, Αναπληρωτής Καθηγητής, Μέλος

Δεκτή:

Πάνος Κωνσταντόπουλος
Πρόεδρος Επιτροπής Μεταπτυχιακών Σπουδών

Φεβρουάριος 2003

*Αφιερώνεται στη Στέλλα
και στην οικογένειά μου*

Παρακολούθηση προσώπων και εκτίμηση της πόζας σε ψηφιακές ακολουθίες εικόνων

Γεώργιος Σημαντήρης

Μεταπτυχιακή Εργασία

Τμήμα Επιστήμης Υπολογιστών

Πανεπιστήμιο Κρήτης

Περίληψη

Στην εργασία αυτή παρουσιάζουμε μία μέθοδο παρακολούθησης προσώπων και εκτίμησης των τρισδιάστατων παραμέτρων κίνησης και γεωμετρίας. Βασική υπόθεση κατά την ανάπτυξη της εργασίας είναι ότι το πρόσωπο βρίσκεται κοντά στην κάμερα και εκτελεί ομαλή και απλή κίνηση. Η μεθοδολογία μπορεί να βρει εφαρμογή σε συστήματα ασφαλείας, περιβάλλοντα διεπαφής ανθρώπου-υπολογιστή και περιβάλλοντα τηλε-συνδιασκέψεων. Επίσης αποτελεί θεμέλιο για την περαιτέρω επεξεργασία του προσώπου, όπως είναι η ανάλυση-αναγνώριση έκφρασης, η παρακολούθηση του βλέμματος, η αναγνώριση ομιλίας και η αναγνώριση-ταυτοποίηση προσώπων.

Η μέθοδος βασίζεται σε σημεία, στρατηγικά τοποθετημένα πάνω στο πρόσωπο, έτσι ώστε να ορίζουν τα κύρια χαρακτηριστικά του, που είναι τα μάτια, η μύτη και το στόμα. Αποτελείται από δύο στάδια : την εκτίμηση των παραμέτρων κίνησης και γεωμετρίας στον τρισδιάστατο χώρο και την αντιστοίχιση σημείων στο δισδιάστατο χώρο.

Στο πρώτο στάδιο θεωρούμε ότι τα επιλεγμένα σημεία αντιστοιχούν σε σημεία που ανήκουν σε μία επίπεδη επιφάνεια στον τρισδιάστατο χώρο. Η μοντελοποίηση αυτή είναι απλούστερη και πιο συμφέρουσα από άλλες στο σχετικό ερευνητικό τομέα. Οι 2-Δ συντεταγμένες των σημείων σε δύο συνεχόμενες εικόνες είναι ικανές να ανακτήσουν τις 3-Δ παραμέτρους κίνησης, μεταφορά και περιστροφή, καθώς και τον προσανατολισμό του προσώπου σχετικά με την κάμερα.

Στο δεύτερο στάδιο παρακολουθούνται τα σημεία στην ακολουθία. Επιστρατεύεται η εκτίμηση της κίνησης για να προκύψει μία πρόβλεψη των θέσεών τους στην επόμενη εικόνα. Η πρόβλεψη αυτή ορίζει μία περιοχή αναζήτησης, μέσα στην οποία

επιχειρείται η εύρεση των σημείων με βάση μία αντιστοίχιση πολλαπλών μπλοκ και γεωμετρικούς περιορισμούς.

Για την αξιολόγηση της προτεινόμενης μεθοδολογίας χρησιμοποιήθηκαν προσομοιώσεις καθώς και πραγματικές ακολουθίες εικόνων. Τα πειραματικά αποτελέσματα ικανοποιούν στην ακρίβεια σε σχέση με το κόστος. Ενθαρρύνουν τον προσανατολισμό της έρευνας στην κατεύθυνση αυτή στοχεύοντας στην ανάπτυξη ενός ολοκληρωμένου συστήματος παρακολούθησης προσώπων σε πραγματικό χρόνο με υψηλή απόδοση.

Επόπτης: Γεώργιος Τζιρίτας
Αναπληρωτής Καθηγητής
Επιστήμης Υπολογιστών
Πανεπιστήμιο Κρήτης

Face tracking and pose estimation in digital image sequences

George Simantiris

Master of Science Thesis

Computer Science Department

University of Crete

Abstract

In this work a face tracking and facial pose estimation methodology is presented. Facial features are tracked and the 3-D motion and geometrical parameters are estimated. Our work is based on the assumption that the face is close to the camera and face motion is smooth and simple. It can be applied in security systems, human-computer interfaces and telecommunication environments. Furthermore, it comprises the basis for more sophisticated image analysis regarding to faces, such as expression analysis-recognition, gaze tracking, lip reading followed by speech recognition and face recognition-authentication.

The method is based on points, strategically placed on the face, such that they determine the main facial features, which are eyes, nose and mouth. It consists of two stages : the estimation of the 3-D motion and geometrical parameters and the 2-D point matching.

In the first stage, the selected points are considered to correspond to points onto a planar surface in 3-D space. This modelling technique is simpler and more profitable, in terms of computational cost, comparative to others, presented in the respective research area. The 2-D image positions of the points in two consecutive frames are capable of retrieving the 3-D motion parameters, translation and rotation, as well as the face pose relative to the camera.

In the second stage, the points are tracked throughout the image sequence. The motion estimation is used in order to predict the point coordinates in the next frame. This prediction defines a search area, in which an attempt to find the corresponding points takes place, based on an multiple block matching technique in addition to geometrical constraints.

To evaluate the proposed methodology, simulations as well as real image sequences have been used. The experimental results satisfy in accuracy towards the cost.

Orientation of research towards this direction is encouraged, targeting the development of a face tracking system meeting high standards and real-time specifications.

Supervisor: George Tziritas
Associate Professor of Computer Science
University of Crete

Ευχαριστίες

Η απόκτηση του Μεταπτυχιακού Διπλώματος Ειδίκευσης αποτέλεσε μία διαδικασία δύσκολη και κουραστική. Πολλοί και αξιόλογοι άνθρωποι μου συμπαραστάθηκαν και με βοήθησαν κατά τη διάρκεια των μεταπτυχιακών μου σπουδών, τους οποίους θα ήθελα να ευχαριστήσω στο σημείο αυτό, ελπίζοντας να μην ξεχάσω κανέναν.

Πρώτο από όλους θα ήθελα να ευχαριστήσω τον κ. Γεώργιο Τζιρίτα, ο οποίος πίστεψε ότι είμαι ικανός να ξεκινήσω και να τελειώσω τις μεταπτυχιακές μου σπουδές και μου στάθηκε σε όλη τη διάρκεια ως Ακαδημαϊκός Σύμβουλος και Επόπτης. Οι πολύτιμες συμβουλές του καθώς και η ουσιαστική του βοήθεια αποτελούν σημαντική συμβολή για την περάτωση της μεταπτυχιακής μου εργασίας. Αφού τον ευχαριστήσω και πάλι, θα ήθελα να του ευχηθώ ό,τι καλύτερο για τον ίδιο και την οικογένειά του.

Στη συνέχεια θα ήθελα να εκφράσω την ευγνωμοσύνη μου σε έναν αξιόλογο νέο άνθρωπο, συνεργάτη και σπουδαίο φίλο, τον Κωνσταντίνο Παναγιωτάκη. Χωρίς τη βοήθειά του και χωρίς την ενθάρρυνσή του στις δύσκολες στιγμές θα μου ήταν αδύνατο να πετύχω τους στόχους μου. Από τα βάθη της καρδιάς μου, σε ευχαριστώ Κωστή, και σου εύχομαι πάντα επιτυχίες και ευτυχία στη ζωή σου.

Στο σημείο αυτό θα ήθελα να ευχαριστήσω και όλους τους συνεργάτες-φίλους μου, με τους οποίους πέρασα ατέλειωτες ώρες εντός αλλά και εκτός Πανεπιστημίου. Μιλώ για τους αξιόλογους ανθρώπους τους οποίους ο κ. Τζιρίτας επέλεξε για να στελεχώσουν την ερευνητική του ομάδα. Τους παλιότερους μεταπτυχιακούς φοιτητές Ηλία Γκρίνια, Σπύρο Λιάπη, Γεώργιο Ζήκο και Μαρία Τράκα. Το νυν διδακτορικό Νίκο Κομοντάκη και τους νεότερους Παναγιώτη Κουτσουράκη, Γιάννη Μαυρικάκη, Γεώργιο Παγώνη και Μανώλη Δελάκη. Επίσης ευχαριστώ και τους Χαράλαμπο Αποστολίδη, Ευτύχιο Σηφάκη και Θέμιδα Ζαμάνη με τους οποίους συνεργάστηκα για ένα μεγάλο διάστημα και τους οποίους εκτιμώ πολύ. Τέλος, οφείλω και ένα μεγάλο ευχαριστώ στον κ. Christophe Garcia, ο οποίος με βοήθησε κατά το ξεκίνημά μου στην έρευνα και μου συμπαραστεκόταν όποτε τον χρειαζόμουνα.

Δεν θα ήθελα να παραλείψω να ευχαριστήσω και τους άλλους δύο καθηγητές της επιτροπής εξέτασης της μεταπτυχιακής μου εργασίας, τους κ.κ. Στέλιο Ορφανουδάκη και Κωνσταντίνο Στεφανίδη, κατ' αρχήν για την υπομονή που είχαν να παρακολουθήσουν την παρουσίαση της εργασίας αλλά και για τις εύστοχες παρατηρήσεις τους σε αρκετά σημεία.

Πολλοί άνθρωποι μου στάθηκαν και με βοήθησαν, ο καθένας με τον τρόπο του, κατά τη διάρκεια των σπουδών μου, αλλά και στην προσωπική μου ζωή. Όλοι τους είναι φίλοι μου, με τους οποίους πέρασα πολλές στιγμές, καλές και κακές, και τους ευχαριστώ βαθύτατα. Ιδιαίτερες ευχαριστίες ανήκουν στους Απόστολο (Τόλη) Παπανικολάου, Ιάκωβο Μαυροειδή, Γεώργιο Σαπουντζή, Μανώλη Τζομπανάκη, Βασίλη Κοντογιάννη, Αντώνη Σταματάκη, Χρήστο Νικόλη και Γιάννη Χαριτάκη. Επίσης ευχαριστώ τους Γεώργιο Αρχων, Χαράλαμπο Παπαδάκη, Χάρη Μαρινάκη, Μιχάλη Κιαγιαδάκη, Νεκτάριο Χατζογιαννάκη, Μαρία Κόκκινου και Μάνια Μαρινάκη. Τέλος, θα ήθελα να ευχαριστήσω θερμά και μια ξεχωριστή παρέα για τη συμπαράσταση και τη φιλία που μου προσέφερε. Αναφέρομαι στους Κώστα Χατζέλη, Νίκη Μαστροδήμου, Κώστα Νικολόπουλο και Βασιλική Δάβαρη.

Θα ήταν νομίζω παράλειψη να μην ευχαριστήσω το Τμήμα Επιστήμης Υπολογιστών του Πανεπιστημίου Κρήτης για τον αναγκαίο εργαστηριακό εξοπλισμό που μου προσέφερε καθώς και όλους τους Καθηγητές μου για τη γνώση με την οποία με εφοδίασαν.

Ολοκληρώνοντας, τη μεγαλύτερη ευγνωμοσύνη μου την αξίζουν οι γονείς μου, Μιχάλης και Ερασμία και ο αδερφός μου Θεμιστοκλής. Οι θυσίες της οικογένειάς μου επέτρεψαν την όλη πνευματική μου πρόοδο και καλλιέργεια και η αμέριστη συμπαράστασή της με στήριξε στη δύσκολη αυτή περίοδο της ζωής μου. Μεγάλη και ιδιαίτερη ευγνωμοσύνη χρωστάω επίσης και στη Στέλλα Μαράκη, η οποία με βοήθησε και με στήριξε σημαντικά από τη στιγμή που οι δρόμοι μας ενώθηκαν. Σ' ευχαριστώ Στέλλα, για τη συμπαράστασή σου, για την κατανόηση που έδειξες, και για όλες τις στιγμές που μοιράστηκα μαζί σου. Θα σε έχω πάντα μέσα στην καρδιά μου !

Περιεχόμενα

Περίληψη	i
Abstract	iii
Ευχαριστίες	v
1 Εισαγωγή	1
1.1 Επεξεργασία εικόνων και πρόσωπα	1
1.2 Ανίχνευση	2
1.3 Αναγνώριση - Ταυτοποίηση	3
1.4 Ανάλυση έκφρασης	4
1.5 Αναπαράσταση - Μοντελοποίηση	5
1.6 Παρακολούθηση	6
1.7 Χρήση παρακολούθησης προσώπων	8
2 Σχετικές Εργασίες	11
3 Προτεινόμενη Μέθοδος	19
4 Έυρεση παραμέτρων κίνησης και 3-Δ διανύσματος κατεύθυνσης	27
4.1 Βασικές εξισώσεις κίνησης	27
4.2 Κίνηση επίπεδων κομματιών (planar patches)	29
4.3 Προσδιορισμός των οκτώ καθαρών παραμέτρων	32
4.4 Υπολογισμός των πραγματικών παραμέτρων κίνησης	35
4.5 Βασικά συμπεράσματα	40
5 Παρακολούθηση σημείων	43
5.1 Σημασία των σημείων	43
5.2 Αντιστοίχιση μπλοκ (block matching)	46
5.3 Πρόβλεψη	48
5.4 Διόρθωση	49
5.4.1 Αναζήτηση ματιών	49
5.4.2 Αναζήτηση μύτης και στόματος	55

5.5 Παρατηρήσεις	60
6 Πειραματικά αποτελέσματα	63
6.1 Εκτίμηση παραμέτρων κίνησης και γεωμετρίας στον 3-Δ χώρο	63
6.2 Παρακολούθηση προσώπων με βάση τα σημεία	69
6.3 Συνδυασμός των δύο σταδίων - Η ολοκληρωμένη μεθοδολογία	73
7 Συμπεράσματα - Επίλογος	83
7.1 Ανασκόπηση	83
7.2 Συμπεράσματα	84
7.3 Μελλοντικές επεκτάσεις	85
A Singular Value Decomposition	87
B Αποδείξεις θεωρημάτων	89
B.1 Θεώρημα 1	89
B.2 Θεώρημα 2	98
B.3 Θεώρημα 3	104
Βιβλιογραφία	107

Κατάλογος Σχημάτων

1.1	Οι έξι βαθμοί ελευθερίας.	8
3.1	Επιλεγμένα σημεία προς παρακολούθηση.	22
3.2	Σχηματική αναπαράσταση προτεινόμενης μεθόδου.	24
3.3	Διάνυσμα προσανατολισμού.	25
4.1	Βασική γεωμετρία της εκτίμησης της 3-Δ κίνησης.	28
4.2	Νέο σύστημα συντεταγμένων.	38
5.1	Παραλληλότητα ευθ. τμημάτων κατά την προοπτική προβολή.	45
5.2	Ορισμός μπλοκ.	46
5.3	Ομαδοποίηση μπλοκ.	48
5.4	Εύρεση της περιοχής αναζήτησης.	50
5.5	Αποστάσεις που αφορούν τα μάτια.	53
5.6	Αναζήτηση ματιών.	55
5.7	Αναζήτηση μύτης.	59
5.8	Αναζήτηση στόματος.	59
6.1	Σημεία πάνω σε επίπεδη επιφάνεια στον 3-Δ χώρο.	64
6.2	Καμπύλες μεταβολής σωστού και ευρεθέντος διανύσματος N στον 3-Δ χώρο (πρώτη προσομοίωση).	67
6.3	Καμπύλες μεταβολής σωστού και ευρεθέντος διανύσματος N στον 3-Δ χώρο (δεύτερη προσομοίωση).	68
6.4	Αποτελέσματα παρακολούθησης για την ακολουθία “erik”.	70
6.5	Αποτελέσματα παρακολούθησης για την ακολουθία “foreman”.	72
6.6	Αποτελέσματα παρακολούθησης και εκτίμησης πόζας για την ακολουθία “kutsurak”.	77
6.7	Αποτελέσματα παρακολούθησης και εκτίμησης πόζας για την ακολουθία “kutsurak” (συνέχεια).	79
B.1	Επίπεδο που τέμνει σφαίρα και ελλειψοειδές στον 3-Δ χώρο.	91
B.2	Επίπεδο που τέμνει ελλειψοειδές στον 3-Δ χώρο.	92
B.3	Επίπεδο που τέμνει με δύο τρόπους ελλειψοειδές στον 3-Δ χώρο.	99

Κεφάλαιο 1

Εισαγωγή

1.1 Επεξεργασία εικόνων και πρόσωπα

Ο τομέας της Μηχανικής Όρασης και Ρομποτικής, παρόλο που περιλαμβάνει ένα ευρύ φάσμα προς μελέτη κι έρευνα, παρατηρούμε ότι έχει επίκεντρο τον άνθρωπο. Κάτι που είναι απολύτως λογικό και μπορεί εύκολα να εξηγηθεί. Ο άνθρωπος έχει έναν τρόπο να αντιλαμβάνεται τα πράγματα τον οποίο θέλει να μεταδώσει-“επιβάλλει” και στον υπολογιστή για να μπορεί να τον καταλαβαίνει. Έτσι, θέλει ο υπολογιστής να “βλέπει” όπως αυτός και να επεξεργάζεται τις εικόνες όπως αυτός. Προσπαθεί δηλαδή να κάνει τον υπολογιστή να αντιλαμβάνεται μία εικόνα όπως θα την αντιλαμβανόταν κι αυτός ο ίδιος. Ακόμα προσπάθησε να κάνει τους υπολογιστές να κινούνται όπως κι αυτός ξεκινώντας από απλά βήματα για να καταλήξουμε στα σημερινά ανδροειδές ρομπότ, που είναι ικανά να τον μιμηθούν σε μεγάλο βαθμό.

Όταν ένας άνθρωπος βλέπει έναν άλλον, το πρώτο πράγμα που προσέχει σ’ αυτόν είναι σχεδόν πάντοτε το πρόσωπό του. Από το πρόσωπο μπορεί πιο εύκολα να θυμηθεί αν τον έχει ξαναδεί, αν τον γνωρίζει ή να καταλάβει τη διάθεσή του από τις εκφράσεις του. Μεγάλο ρόλο στην όποια έκφραση παίζουν τα μάτια, η μύτη και το στόμα του ανθρώπου, κι αν όχι το καθένα ξεχωριστά τότε σίγουρα όλα μαζί μας βοηθάνε να σχηματίσουμε μία πρώτη άποψη.

Είναι ευνόητο λοιπόν το ανθρώπινο πρόσωπο να παίζει ένα σημαντικό ρόλο και στην ανάλυση ψηφιακών εικόνων και να έχει αποτελέσει, να αποτελεί τώρα αλλά και στο μέλλον σημαντικό θέμα έρευνας στα πεδία της ψηφιακής επεξεργασίας εικόνων, της μηχανικής όρασης καθώς και της ψυχολογίας. Συγκεκριμένα η έρευνα αυτή εστιάζει στην ανίχνευση (detection), αναγνώριση (recognition), ταυτοποίηση (authentication), παρακολούθηση (tracking) προσώπων, αναγνώριση-ανάλυση έκφρασης προσώπων (expression analysis-recognition), παρακολούθηση βλέμματος (gaze tracking), αναπαράσταση-μοντελοποίηση προσώπων (face synthesis, animation). Θα εξηγήσουμε παρακάτω με λίγα λόγια τι σημαίνει το καθένα από αυτά.

1.2 Ανίχνευση

Με τον όρο ανίχνευση προσώπων εννοούμε την εύρεση ενός προσώπου σε μια εικόνα και ταυτόχρονα τον προσδιορισμό της θέσης του, αν και εφόσον υπάρχει. Σαν πρώτο βήμα ένας ανιχνευτής πρέπει λοιπόν να προσδιορίσει αν υπάρχει ή όχι πρόσωπο στην εικόνα. Στη συνέχεια, αν το πρώτο βήμα είναι θετικό, πρέπει να εντοπιστεί αυτό με όσο το δυνατόν περισσότερη ακρίβεια. Μπορούμε να ξεχωρίσουμε γενικά δύο κατηγορίες εικόνων : αυτές που παίρνονται υπό ελεγχόμενες συνθήκες, όπως είναι π.χ. φωτογραφίες σε αστυνομικά τμήματα ή σε εργαστήρια φωτογραφίας, και αυτές που δεν προϋποθέτουν κανέναν απολύτως περιορισμό, όπως μπορούμε να βρούμε από κάποια φωτογραφική μηχανή ή κάποιο βίντεο. Στην πρώτη περίπτωση η ανίχνευση καθίσταται εύκολη υπόθεση. Για παράδειγμα, υπάρχει μόνο ένα πρόσωπο στην εικόνα και μάλιστα σχεδόν κεντρικά, το πρόσωπο κοιτάει το φακό, το σκηνικό είναι ομοιογενές, δεν υπάρχουν επικαλύψεις και δε φοράει γυαλιά. Στη δεύτερη κατηγορία, όμως, τα πράγματα δεν είναι τόσο εύκολα. Πολλοί παράγοντες μπορούν να επηρεάσουν και να δυσκολέψουν το πρόβλημα της ανίχνευσης. Η ύπαρξη μαλλιών μπροστά στο πρόσωπο, το μακιγιάζ, τα γυαλιά, το μουστάκι και το μούσι είναι ορισμένοι παράγοντες που καλύπτουν και αλλοιώνουν το πρόσωπο. Επίσης συναντάμε μεταβολές στην κλίμακα και τον προσανατολισμό του προσώπου που κάνουν το έργο δυσκολότερο. Τέλος, υπάρχουν και παράγοντες που οφείλονται σε τεχνικούς λόγους, όπως χαμηλή ποιότητα εικόνας, θόρυβος ή σκιές κατά τη λήψη. Το πρόβλημα της ανίχνευσης προσώπων είναι ακόμα και σήμερα αρκετά δυσεπίλυτο. Η ανίχνευση όμως γίνεται ολοένα και πιο σημαντική σε πολλές εφαρμογές. Μετά την επιτυχημένη ανίχνευση θα μπορούσε να ακολουθήσει η αναγνώριση του προσώπου μέσα από μία βάση δεδομένων προσώπων, για να προσδιοριστεί η ταυτότητά του, να ανιχνευτούν τα επιμέρους χαρακτηριστικά του, όπως είναι τα μάτια, η μύτη, το στόμα, ή να βρεθεί η θέση και η κατεύθυνσή του στον τρισδιάστατο χώρο. Θα μπορούσε επίσης να ξεκινήσει μία παρακολούθηση του προσώπου ή και των επιμέρους χαρακτηριστικών του στο χρόνο.

Η ανίχνευση προσώπων έχει γνωρίσει αρκετή προσοχή τα τελευταία χρόνια για το λόγο ότι αποτελεί προαπαιτούμενο στάδιο για την αναγνώριση. Οι μελέτες όμως επικεντρώνονταν κυρίως στην αναγνώριση, όπου βέβαια για να προχωρήσει η αναγνώριση πρέπει πρώτα να ανιχνευτεί το πρόσωπο. Σ' αυτές τις περιπτώσεις, για να απλοποιηθεί η ανίχνευση χρησιμοποιούνται εικόνες με αρκετούς περιορισμούς κατά τη λήψη τους έτσι ώστε οι ανιχνευτές να μην έχουν ιδιαίτερη δυσκολία να εντοπίζουν τα πρόσωπα. Μία πιο αναλυτική επισκόπηση τεχνικών που αφορούν την ανίχνευση μπορεί να βρεθεί στο [90]. Αυτές οι μέθοδοι μπορούν περίπου να χωριστούν σε τρεις γενικές κατηγορίες : τοπική ανίχνευση χαρακτηριστικών προσώπου, αντιστοίχιση με φόρμα (template matching) και image invariants. Στην πρώτη κατηγορία, χρησιμοποιούνται χαμηλού επιπέδου αλγόριθμοι μηχανικής όρασης σε συνδυασμό με στατιστικά μοντέλα του ανθρώπινου προσώπου για να ανιχνευτούν χαρακτηριστικά προσώπου όπως είναι τα μάτια, η μύτη, το στόμα,

όπως στα [19], [97], [44], [100]. Στη δεύτερη περίπτωση, χρησιμοποιούνται διάφορες φόρμες συσχέτισης ώστε να ανιχνευτούν τοπικά υπό-χαρακτηριστικά που θεωρούνται σταθερά (rigid) (ιδιοχώροι [59]) ή μη (deformable templates [94], [1]) στην εμφάνιση. Στην τελευταία περίπτωση, σχήματα image invariants υποθέτουν ότι υπάρχουν ιδιαίτερες χωρικές συγγένειες οι οποίες είναι γενικές και πιθανώς μοναδικές για όλα τα πρότυπα προσώπων ακόμα και υπό διαφορετικές συνθήκες εμφάνισης στην εικόνα [69]. Επίσης υπάρχουν και μέθοδοι που, αντί να ακολουθούν ένα σύνολο από κανόνες σχεδιασμένους από ανθρώπους, βασίζονται στην τεχνητή νοημοσύνη. Τέτοιες μέθοδοι χρησιμοποιούν τα νευρωνικά δίκτυα [67], [48], [25], [26], [17], [1] τα οποία και έχουν το πλεονέκτημα να μαθαίνουν τους κανόνες από μία δεδομένη συλλογή αντιπροσωπευτικών παραδειγμάτων. Τέλος, πρόσφατες έρευνες έχουν χρησιμοποιήσει τα λεγόμενα κυματίδια (wavelets), επιτυγχάνοντας πολύ καλά αποτελέσματα στην ανίχνευση [28], [30], [27], [1].

1.3 Αναγνώριση - Ταυτοποίηση

Αναγνώριση ενός προσώπου είναι, όπως λέει και η λέξη, να αναγνωριστεί ο άνθρωπος του οποίου το πρόσωπο υπάρχει σε μία εικόνα. Για να γίνει αυτό πρέπει φυσικά ο άνθρωπος αυτός να είναι γνωστός και να έχει καταχωρηθεί η ταυτότητά του σε μία βάση δεδομένων. Με άλλα λόγια, το πρόσωπο του ανθρώπου αυτού ή/και διάφορα χαρακτηριστικά που θα βοηθήσουν στη διαδικασία πρέπει να υπάρχουν για να μπορεί να γίνει σύγκριση. Η ταυτοποίηση είναι ένα παρακλάδι της αναγνώρισης. Ταυτοποίηση σημαίνει να επιβεβαιωθεί κάποιος που ισχυρίζεται ότι είναι αυτός. Η διαφορά της αναγνώρισης από την ταυτοποίηση είναι ότι στη μεν πρώτη περίπτωση πρέπει να αναζητηθεί μέσα σε μία βάση δεδομένων το πρόσωπο προς αναγνώριση, ενώ στη δεύτερη περίπτωση ήδη έχουμε την πληροφορία ότι το πρόσωπο είναι κάποιος και καλούμαστε να το επιβεβαιώσουμε. Η αναγνώριση/ταυτοποίηση γίνεται ολοένα και πιο σημαντική στη σημερινή εποχή, όπου η ασφάλειά μας κλονίζεται μέρα με τη μέρα. Περιπτώσεις που απαιτούν υψηλή ασφάλεια, όπως η είσοδος σε απαγορευμένες περιοχές ή κτίρια, και χρηματοοικονομικές συναλλαγές μέσω μηχανημάτων αυτόματης ανάληψης χρημάτων, θα μπορούσαν να χρησιμοποιήσουν τέτοια συστήματα για να βελτιώσουν την ασφάλεια που προσφέρουν και να αποτρέπουν κακοποιά στοιχεία να κάνουν εγκληματικές απόπειρες. Υπάρχουν βέβαια πολλές βιομετρικές μέθοδοι αναγνώρισης, αλλά αυτή του προσώπου έχει προσελκύσει το μεγαλύτερο ενδιαφέρον των ερευνητών για λόγους αποτελεσματικότητας, φιλικότητας προς το χρήστη και γενικότερης αποδοχής από το κοινό.

Όπως και στην ανίχνευση, έτσι και στην αναγνώριση, υπάρχουν πολλοί παράγοντες που επηρεάζουν και καθιστούν το όλο έργο πιο δύσκολο (αναφερθήκαμε σ' αυτούς γενικά στην υποενότητα 1.2). Τα τελευταία χρόνια έχει σημειωθεί αξιοσημείωτη πρόοδος στην αναγνώριση, μιας και όλο σχεδόν το βάρος των ερευνητών έχει πέσει πάνω σ' αυτήν. Ειδικά κάτω από σταθερές συνθήκες, όπως είναι μικρές αποκλίσεις στο φωτισμό,

στην έκφραση και την πόζα του προσώπου. Μία επισκόπηση των μεθόδων αναγνώρισης μπορεί να αναζητηθεί στα [93], [61]. Γενικά, οι μέθοδοι που χρησιμοποιούνται είναι γεωμετρική αντιστοίχιση χαρακτηριστικών (geometrical features matching) και αντιστοίχιση με φόρμα (template matching). Στην πρώτη κατηγορία, εξάγονται κάποιες γεωμετρικές μετρήσεις για διακριτικά χαρακτηριστικά προσώπου, όπως είναι μάτια, μύτη, στόμα [9]. Στη δεύτερη κατηγορία, η εικόνα του προσώπου αναπαρίσταται ως ένας διδιάστατος πίνακας τιμών φωτεινότητας και συγκρίνεται με μία ή περισσότερες φόρμες που αντιπροσωπεύουν ένα ολόκληρο πρόσωπο. Οι πρώτες μέθοδοι αντιστοίχισης με φόρμα βασίζονται πάνω στη συσχέτιση. Τα τελευταία χρόνια χρησιμοποιήθηκε και η μέθοδος της ανάλυσης σε κύριες συνιστώσες (Principal Component Analysis, PCA) ή αλλιώς γνωστή σαν μέθοδος Karhunen-Loeve για το σκοπό της αναγνώρισης [46], [75], [87]. Τέλος, στην αναγνώριση έχουν χρησιμοποιηθεί και νευρωνικά δίκτυα [14], [60], καθώς και πιο πρόσφατα τα κυματίδια (wavelets) [31], [29].

1.4 Ανάλυση έκφρασης

Το πρόσωπο είναι σίγουρα το πιο σημαντικό μέσο έκφρασης που διαθέτει ένας άνθρωπος. Με μια ματιά στο πρόσωπο ενός ανθρώπου μπορούμε πολύ εύκολα να καταλάβουμε αν είναι ευτυχισμένος-χαρούμενος, θυμωμένος, φοβισμένος, θλιμμένος ή έκπληκτος. Έτσι μπορούμε να προσαρμόσουμε τη συμπεριφορά μας απέναντί του, ώστε να μην τον φέρουμε σε δύσκολη θέση και να αποφευχθούν παρεξηγήσεις. Ο τομέας αυτός, λοιπόν, έχει σκοπό να καταστήσει τους υπολογιστές ικανούς να αναγνωρίζουν μέσω των εκφράσεων την ψυχολογική κατάσταση του ανθρώπου-χρήστη. Όλο αυτό μπορεί να χρησιμοποιηθεί σε περιβάλλοντα διεπαφής, για να γίνει η όλη επικοινωνία ανθρώπου και υπολογιστή πιο φιλική, θερμή και άμεση, με αποτέλεσμα την αποφυγή πολλών παρεξηγήσεων που αφορούν τους υπολογιστές. Δίδεται έτσι μία “ανθρώπινη πλευρά” σε ένα μηχάνημα, κάτι που θα το κάνει να ξεχωρίσει από τα μηχανήματα-ρομπότ που κατασκευάζονται και προγραμματίζονται για να εκτελούν μία και μόνο εργασία. Άμεσες επακολουθήσεις της ανάλυσης έκφρασης είναι η παρακολούθηση χειλιών (lip tracking) και η παρακολούθηση βλέμματος (gaze tracking). Στην πρώτη, παρακολουθούνται τα χείλια του ανθρώπου με σκοπό την αναγνώριση ομιλίας. Αν επιτευχθεί κάτι τέτοιο καταλαβαίνουμε ότι η χρήση των υπολογιστών θα μεταφερθεί σε μία άλλη “εποχή”, αφού ουσιαστικά συσκευές εισόδου, όπως είναι το πληκτρολόγιο και το ποντίκι, αν δεν θα εξαφανιστούν, σίγουρα θα περιοριστούν στο ελάχιστο. Στη δεύτερη, παρακολουθείται το βλέμμα του ανθρώπου και έτσι ο υπολογιστής θα μπορεί να ξέρει κατά πόσο έχει την πλήρη προσοχή του χρήστη ή κατά πόσο του προκαλεί κάποιο συναίσθημα (π.χ. θυμό) ή τον φέρνει σε κάποια κατάσταση (π.χ. ανία). Επίσης η παρακολούθηση του βλέμματος μπορεί να χρησιμοποιηθεί και σαν περιβάλλον διεπαφής αντικαθιστώντας το ποντίκι.

Το πρόβλημα της ανάλυσης έκφρασης είναι ένα δύσκολο πρόβλημα, καθώς η όποια ανθρώπινη έκφραση συνοδεύεται και από κάποια, απλή ή σύνθετη, κίνηση του

κεφαλιού και κατά συνέπεια του προσώπου. Γι' αυτό και οι πρώτες ερευνητικές απόπειρες περιορίζονταν την αναγνώριση έκφρασης με την προϋπόθεση το πρόσωπο να είναι ακίνητο ή περίπου ακίνητο. Τα μοντέλα που χρησιμοποιούνται στην αναγνώριση εκφράσεων προσώπου ποικίλουν στην ποσότητα γεωμετρικής πληροφορίας περί σχήματος και κίνησης προσώπου που περιέχουν. Στο ένα άκρο υπάρχουν προσεγγίσεις που χρησιμοποιούν μοντέλα που βασίζονται στη φυσική υπόσταση των προσώπων περιλαμβάνοντας το δέρμα και τους μύες [76], [20]. Ένα ελαφρώς πιο αδύναμο μοντέλο χρησιμοποιεί ελαστικές φόρμες για να αναπαραστήσει τα σχήματα που παίρνουν τα χαρακτηριστικά προσώπου [101]. Στο άλλο άκρο, η αναγνώριση των εκφράσεων γίνεται χρησιμοποιώντας στατιστικές ιδιότητες της οπτικής ροής μαζί με πολύ αδύναμα μοντέλα του σχήματος προσώπου [95]. Στο ενδιάμεσο υπάρχουν βέβαια και μέθοδοι μεταξύ των προσεγγίσεων βασιζόμενων σε φόρμες και στην οπτική ροή, όπως είναι οι [96], [7]. Άλλες μέθοδοι χρησιμοποιούν επιπλέον βοηθήματα για να καταστήσουν τον υπολογιστή ικανό να αναγνωρίσει ορισμένες εκφράσεις, όπως είναι ειδικά γυαλιά [68]. Στην πιο απλή περίπτωση η αναγνώριση εκφράσεων γίνεται χρησιμοποιώντας τεχνικές συσχέτισης εικόνων [5]. Τέλος, μέθοδοι εκμάθησης έχουν χρησιμοποιηθεί επιτυχώς περιλαμβάνοντας στατιστική εκμάθηση [15] και νευρωνικά δίκτυα [66], [38]. Όσον αφορά την παρακολούθηση βλέμματος, πάντα προηγείται μία ανίχνευση των ματιών και στη συνέχεια διαπιστώνεται η κατεύθυνση που κοιτάει το πρόσωπο χρησιμοποιώντας γεωμετρικά χαρακτηριστικά [6], [58] ή νευρωνικά δίκτυα [72].

1.5 Αναπαράσταση - Μοντελοποίηση

Η αναπαράσταση ενός προσώπου, η οποία συνήθως ξεφεύγει από τις δύο διαστάσεις και πάει στις τρεις, καθώς και η μοντελοποίηση-κωδικοποίηση του προσώπου βρίσκουν πάρα πολύ μεγάλη εφαρμογή σε τηλεσυνδιασκέψεις, βίντεο-τηλέφωνα, συμπίεση εικόνων, αναγνώριση προσώπων, εικονικά πρόσωπα, διεπαφή ανθρώπου-υπολογιστή και γραφική. Όταν κάποιος ή κάτι μας τραβάει το ενδιαφέρον στρέφουμε την προσοχή μας προς τον άνθρωπο ή το αντικείμενο που έχει την προσοχή μας. Για να γίνει αυτό δεν φτάνει μόνο να στρέψουμε τα μάτια μας προς τη συγκεκριμένη κατεύθυνση, πράγμα που ούτως ή άλλως καταντάει κουραστικό μετά από λίγη ώρα. Έτσι συνήθως στρέφουμε όλο το κεφάλι μας, και μαζί το πρόσωπο. Όλα αυτά μας φαίνονται αυτονόητα στις τρεις διαστάσεις που διέπουν τον κόσμο μας (κατά πολλούς οι διαστάσεις είναι τέσσερις, μετρώντας μαζί και το χρόνο), αλλά δεν καλύπτονται από τις δύο διαστάσεις που μας προσφέρει μία εικόνα. Θεωρώντας το απλό παράδειγμα μιας τηλε-εικονο-συνομιλίας μεταξύ δύο ατόμων μπορούμε εύκολα να καταλάβουμε πόσο αναγκαία γίνεται μία αναπαράσταση-μοντελοποίηση των προσώπων. Αν ο ένας συνομιλητής κοιτάζει πάντα την κάμερα, ο άλλος έχει την εντύπωση πως ο πρώτος τον κοιτάζει συνέχεια. Αν πάλι ο πρώτος θελήσει να κοιτάξει τον συνομιλητή του που εμφανίζεται σε ένα παράθυρο στην οθόνη του, ο δεύτερος θεωρεί ότι δεν έχει την προσοχή του πρώτου, καθώς ο πρώτος εμφανίζεται να

κοιτάει αλλού. Το πρόβλημα αυτό καλείται να λυθεί με βάση μία αναπαράσταση των προσώπων στις τρεις διαστάσεις, όπου πλέον οι συνομιλητές μπορούν να έχουν αίσθηση για το που κοιτάζει ο καθένας. Περαιτέρω, με τη μοντελοποίηση μπορούμε να επιτύχουμε καλύτερη κωδικοποίηση των προσώπων κατά τη μετάδοσή τους και να κάνουμε έτσι το όλο σύστημα πιο γρήγορο και άρα πιο ρεαλιστικό.

Η αναπαράσταση συνδέεται άμεσα με την ανεύρεση του προσανατολισμού του προσώπου. Τελευταία προτάθηκαν διάφορες τεχνικές για την εκτίμηση της κατεύθυνσης στην οποία κοιτάει ένα πρόσωπο. Έτσι, έχει μοντελοποιηθεί ο προσανατολισμός σαν ένας γραμμικός συνδυασμός από ανομοιότητες μεταξύ περιοχών προσώπων και διάφορων μοντέλων προσώπων [86]. Επίσης υπάρχουν αλγόριθμοι που εκμεταλλεύονται τις γεωμετρικές συμμετρίες του τυπικού προσώπου [32], [37], όπως αυτές αποδεικνύονται και ορίζονται στα [21], [99]. Για την ανάκτηση της τρισδιάστατης υψής από τις δύο διαστάσεις μίας εικόνας προτάθηκαν αλγόριθμοι που περιλαμβάνουν ανακατασκευή σχήματος από φωτοσκίαση (shape from shading) [3], σύνθεση από κίνηση [4], [16] καθώς και τα ιδιοχαρακτηριστικά, που βρίσκουν και ιδιαίτερη εφαρμογή στην κωδικοποίηση προσώπου, [42], [70], [87], [57], [55], [92]. Τέλος, κωδικοποίηση προσώπου επιτυγχάνεται και μέσω μοντελοποίησης με πολύγωνα στον τρισδιάστατο χώρο [2], [12], [73].

1.6 Παρακολούθηση

Ο στόχος γενικά στην παρακολούθηση κίνησης είναι η επιστροφή πληροφορίας η οποία χαρακτηρίζει τη διαμόρφωση ή, πιο γενικά, την κατάσταση ενός αντικειμένου-στόχου¹ από τα δεδομένα που εμπεριέχονται σε μία ακολουθία εικόνων. Οι εικόνες μπορεί να είναι μονόχρωμες ή έγχρωμες, απλές (monocular) ή στερεοσκοπικές (stereo). Η κατάσταση η οποία θα υπολογίζεται μπορεί να περιλαμβάνει τη θέση του στόχου, την κατεύθυνσή του, το σχήμα του, την ταχύτητά του ή μια ποικιλία άλλων γεωμετρικών ή φωτομετρικών ιδιοτήτων. Παρακολούθηση προσώπων σημαίνει, λοιπόν, όλα τα παραπάνω όταν το αντικείμενο του ενδιαφέροντος είναι ένα πρόσωπο.

Κάτω από ιδανικές συνθήκες ένα σύστημα παρακολούθησης κίνησης θα πρέπει να είναι σε θέση να επιστρέψει την πλήρη κατάσταση του αντικειμένου-στόχου, αλλά ιδανικές συνθήκες σπάνια συναντάει κανείς στον πραγματικό κόσμο. Ακόμα και εμείς οι άνθρωποι, με το υψηλά ανεπτυγμένο σύστημα όρασής μας, δεν είμαστε σε θέση να παρακολουθούμε με μεγάλη ακρίβεια αντικείμενα κάτω από αντίξοες συνθήκες, όπως όταν τα αντικείμενα είναι στην οπτική μας περιφέρεια, όταν κινούνται πολύ γρήγορα, όταν καλύπτονται μερικώς ή ολοκληρωτικά, ή όταν ο φωτισμός δεν είναι επαρκής. Παρόλα αυτά ο άνθρωπος είναι σε θέση να κάνει δύο πράγματα : να κερδίσει μερική γνώση της κατάστασης ενός αντικειμένου το οποίο δεν μπορεί να παρακολουθηθεί με ακρίβεια, και να ξαναβρίσκει ένα αντικείμενο που προσωρινά δεν ήταν ορατό. Τέτοια

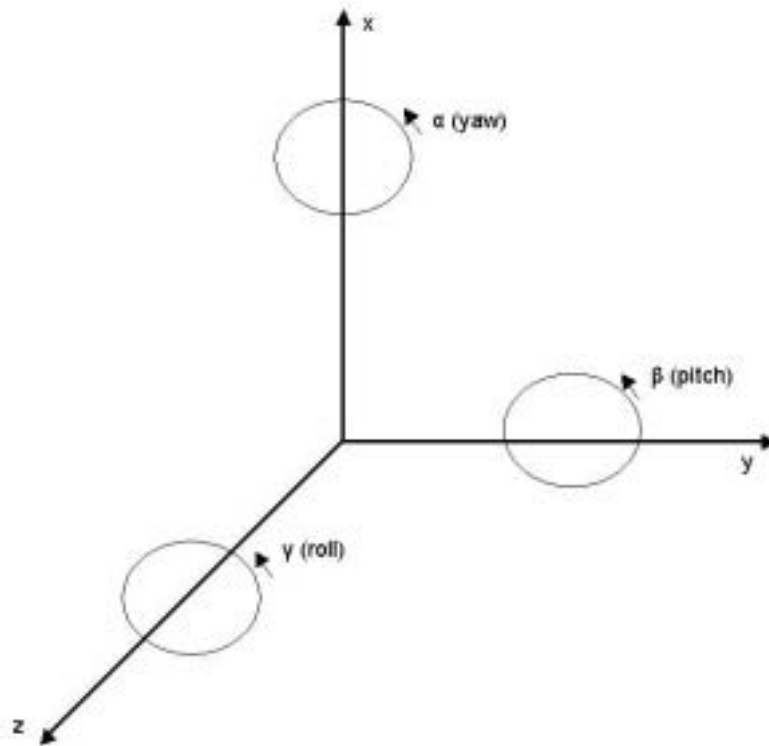
¹Με τον όρο *αντικείμενο* εννοούμε οτιδήποτε από απλά χαρακτηριστικά, όπως ακμές, ως και πολύπλοκα τρισδιάστατα αντικείμενα.

παραδείγματα βρίσκουμε καθημερινά μπροστά μας : οι αθλητές μπορούν να πιάσουν, να χτυπήσουν ή να κλοτηήσουν μια μπάλα γνωρίζοντας κατά προσέγγιση τη θέση της ακόμα και αν η κατεύθυνσή της αλλάζει συνεχώς και μάλιστα πολύ γρήγορα, και οι οδηγοί συνεχίζουν να παρακολουθούν το μπροστινό τους όχημα ακόμα και αν γυρίσουν το βλέμμα τους στους καθρέπτες για να ελέγξουν τι γίνεται πίσω τους.

Η παρακολούθηση είναι διαφορετική από την ανίχνευση καθ' όσον χρησιμοποιεί τη χρονική συσχέτιση για να εντοπίζει το αντικείμενο του ενδιαφέροντος σε μια ακολουθία εικόνων (video sequence), αντί να το ανιχνεύει σε κάθε εικόνα ανεξάρτητα. Η χρονική πληροφορία περιορίζει αρκετά την περιοχή του ψαξίματος και έτσι μπορούμε ακόμα και να πετύχουμε παρακολούθηση σε πραγματικό χρόνο (real-time tracking). Η παρακολούθηση ορίζεται σαν τον εντοπισμό και την παρακολούθηση του προσώπου από τη στιγμή που αυτό αρχικοποιείται μέχρι τη στιγμή που θα εξαφανιστεί μόνιμα ή δεν θα χρειάζεται άλλο. Η ανίχνευση είναι απαραίτητη. Αυτό γιατί για να ξεκινήσει μία παρακολούθηση πρέπει να υπάρχει κάποιος στόχος. Το στάδιο της ανίχνευσης, λοιπόν, είναι ένα προαπαιτούμενο βήμα της παρακολούθησης, καθ' όσον αυτό είναι που θα μας δώσει την πληροφορία για το αν υπάρχει κάποιος στόχος και που βρίσκεται, για να μπορέσει να αναλάβει μετά η παρακολούθηση το έργο της. Η ανίχνευση αποτελεί την αρχικοποίηση της παρακολούθησης. Βέβαια, το στάδιο αυτό δεν γίνεται πάντα αυτόματα, όπως συμβαίνει σε πολλές περιπτώσεις, είτε για λόγους απλοποίησης, είτε γιατί απαιτείται μία περισσότερο πολύπλοκη αρχικοποίηση απ' όσον είναι σε θέση να δώσει η ανίχνευση. Επίσης σε κάποια συστήματα αξιώνεται μία αλληλεπίδραση με τον άνθρωπο-χρήστη, είτε για λόγους απλότητας, είτε γιατί ο χρήστης πρέπει να πάρει την απόφαση για το πότε θα ξεκινήσει η παρακολούθηση.

Η παρακολούθηση προσώπων στις περισσότερες περιπτώσεις συνδέεται και με την ανάκτηση της πόζας, δηλαδή του προσανατολισμού του προσώπου. Με άλλα λόγια πρέπει να βρεθούν μερικοί ή όλοι από τους έξι συνολικά βαθμούς ελευθερίας (6 Degrees of Freedom, DOF), όπως φαίνονται στο σχήμα 1.1. Οι βαθμοί ελευθερίας είναι η μεταφορά στους άξονες και η περιστροφή γύρω από τους άξονες x, y, z, συνολικά έξι παράμετροι. Παρόλο που αυτό δεν είναι απολύτως αναγκαίο, η ανεύρεσή τους είναι απαραίτητη για την πλήρη περιγραφή της κίνησης.

Η παρακολούθηση έχει να αντιμετωπίσει όλες τις κοινές δυσκολίες που συναντιούνται στη μηχανική όραση. Παράγοντες όπως οι αποκλίσεις του φωτισμού και ο θόρυβος είναι διαδεδομένοι σε ψηφιακές εικόνες. Προβλήματα δημιουργεί και η ανώμαλη κίνηση που μπορεί να κάνει ένα αντικείμενο. Όταν κάποιο αντικείμενο βρίσκεται κοντά στο φακό φαίνεται να κινείται γρηγορότερα απ' όσον κάποιο άλλο που βρίσκεται πιο μακριά. Επίσης υπάρχει και ο παράγοντας σύγχυσης, που σημαίνει ότι άλλα αντικείμενα, είτε του ίδιου τύπου, είτε διαφορετικά βρίσκονται στο σκηνικό. Έτσι, αν το αντικείμενο προς παρακολούθηση είναι ένα πρόσωπο, το σύστημα παρακολούθησης μπορεί να μπερδευτεί αν εμφανιστούν κι άλλα πρόσωπα μέσα στην εικόνα. Το ίδιο μπορεί να συμβεί εάν μεταξύ του αντικειμένου-στόχου παρεμβληθεί κάτι άλλο, π.χ. κάποιο άλλο πρόσωπο ή



Σχήμα 1.1: Οι έξι βαθμοί ελευθερίας : $[x \ y \ z \ \alpha \ \beta \ \gamma]^T$ (3-Δ μεταφορά και περιστροφή).

χέρι στην περίπτωσή μας. Ακόμα, η μετακίνηση ή η περιστροφή του στόχου εκτός πλάνου μπορούν να δυσκολέψουν το σύστημα παρακολούθησης με απρόβλεπτες συνέπειες. Τέλος, η απαίτηση που έχουν πολλές εφαρμογές για παρακολούθηση σε πραγματικό χρόνο επιφέρει ένα ακόμα παραπάνω πρόβλημα.

1.7 Χρήση παρακολούθησης προσώπων

Η παρακολούθηση προσώπων μπορεί να χρησιμεύσει σαν προγενέστερο βήμα για άλλες μεθόδους ανάλυσης της εικόνας που περιέχει πρόσωπο, όπως είναι ανάλυση έκφρασης προσώπων (face expression analysis), παρακολούθηση βλέμματος (gaze tracking), δυναμική κίνηση ομοιωμάτων (face animation) και αναγνώριση ομιλίας (lip-reading). Είναι σε θέση να διευκολύνει πάρα πολύ το έργο των προαναφερθέντων μεθόδων, καθώς τις τροφοδοτεί ανά πάσα στιγμή με χρήσιμες πληροφορίες, που διαφορετικά θα αναγκάζονταν να ανακτήσουν μόνες τους. Επίσης βρίσκει εφαρμογή στην αναγνώριση προσώπων (face recognition), διότι μπορεί να προσφέρει στο σύστημα αναγνώρισης μία πληθώρα εικόνων του προσώπου, σε διαφορετικές γωνίες και κλίμακες, το οποίο μπορεί να χρησιμοποιήσει πολλές από αυτές για επιβεβαίωση ή να διαλέξει από αυτές τις πιο χρήσιμες γι' αυτό. Είναι επίσης γεγονός ότι σε ένα σύστημα αναγνώρισης το οποίο προορίζεται για λειτουργία σε

ποικίλες τοποθεσίες, π.χ. για ενίσχυση της ασφάλειας, δεν μπορούμε να αναγκάσουμε το χρήστη να πάρει μία συγκεκριμένη πόζα μπροστά στην κάμερα του συστήματος για να πετύχει η αναγνώριση. Είναι δύσκολο να επιτευχθεί κάτι τέτοιο και προπάντων είναι μη φιλικό προς το χρήστη. Η παρακολούθηση μπορεί να δώσει κι εκεί λύση.

Μία άλλη, διαφορετική εφαρμογή της παρακολούθησης βρίσκουμε σε θέματα κωδικοποίησης και μετάδοσης. Παρακολουθείται ένα αντικείμενο σε μία ακολουθία εικόνων (video), έτσι ώστε να κωδικοποιηθεί. Αυτό μπορεί να γίνει κωδικοποιώντας το αντικείμενο, όπως στην περίπτωση μας είναι το πρόσωπο, με περισσότερα μπιτ (δυφία, bits) μιας και μας ενδιαφέρει, ενώ χρησιμοποιούμε λιγότερα δυφία για το υπόλοιπο (φόντο, background). Έτσι πετυχαίνουμε υψηλότερη συμπίεση που μας ευνοεί στην μετάδοση του βίντεο χωρίς όμως να χάσουμε σε ποιότητα [64].

Η παρακολούθηση προσώπων βρίσκει ιδιαίτερη χρήση σε βίντεο-τηλέφωνα και τηλεσυνδιασκέψεις (video-conferencing). Τα περισσότερα συστήματα τηλεσυνδιασκέψεων αδυνατούν να εξασφαλίσουν οπτική επαφή (eye contact) μεταξύ των συνομιλητών, με την έννοια ότι δεν μπορούν να προσδιορίσουν πού ή τι κοιτάνε οι συνομιλητές. Αυτό συμβαίνει διότι για να κοιταχτούν οι συνομιλητές ουσιαστικά κοιτάζουν την απεικόνιση του άλλου στην οθόνη και όχι την κάμερα, η οποία βρίσκεται συνήθως πάνω, κάτω ή κάπου δίπλα στην οθόνη. Αν δεν κοιτάζουν μέσα στην κάμερα είναι δύσκολο να αντιληφτεί κανείς ότι έχουμε οπτική επαφή μαζί τους. Από την άλλη μεριά, αν κοιτάζουν πάντα μέσα στην κάμερα, εμείς θα αντιλαμβανόμαστε ότι μας κοιτάζουν πάντοτε, άσχετα με το αν μετακινούμαστε ή όχι. Ένα διάσημο τέτοιο παράδειγμα είναι ο πίνακας Τζοκόντα του Λεονάρντο ντα Βίντσι, τα μάτια της οποίας φαίνεται να μας ακολουθούν παντού. Παρακολουθώντας λοιπόν το πρόσωπο και υπολογίζοντας την κατεύθυνση στην οποία κοιτάει (την πόζα του), η οποία είναι σχετική με την κάμερα, μπορούμε να την γνωστοποιήσουμε στους άλλους συνομιλητές ανασυνθέτοντας σωστά το πρόσωπο στην οθόνη τους [34].

Τέλος, η παρακολούθηση μπορεί να είναι και απαραίτητη σε περιβάλλοντα διεπαφής ανθρώπου-υπολογιστή ή αντιληπτικά περιβάλλοντα, γιατί σίγουρα μπορεί να κάνει τη χρήση του υπολογιστή πιο εύκολη και πιο φιλική προς το χρήστη. Η απελευθέρωση του ενός χεριού που σχεδόν συνέχεια αναγκάζεται να προσφύγει στο ποντίκι, ο έλεγχος του κέρσορα σε προγράμματα ψυχαγωγίας όπως είναι παιχνίδια και προγράμματα ζωγραφικής, δίνουν μία ακόμα ελευθερία στο χρήστη και αυξάνουν την αποδοτικότητα και τη διασκέδασή του. Επιπλέον, εφοδιάζουν και τον υπολογιστή με μία ικανότητα αίσθησης, αντίληψης και όρασης κοντά στα πρότυπα του ανθρώπου.

Από τα παραπάνω είναι προφανές ότι η παρακολούθηση προσώπων είναι αρκετά χρήσιμη σε (και για) πολλές άλλες εφαρμογές. Τα συστήματα παρακολούθησης έχουν κεντρίσει ιδιαίτερα το ενδιαφέρον των ερευνητών. Ειδικά τα τελευταία χρόνια με την τεχνολογική εξέλιξη της μικροηλεκτρονικής και των συστημάτων όρασης καθίσταται πια εφικτή η υψηλή απόδοση των συστημάτων αυτών, ενώ το κόστος κτήσης και λειτουργίας τους παραμένει σε χαμηλά επίπεδα. Πολλές και διαφορετικές προσεγγίσεις

έχουν προταθεί, από τις οποίες τις σπουδαιότερες θα παρουσιάσουμε και θα αναλύσουμε στο επόμενο κεφάλαιο.

Κεφάλαιο 2

Σχετικές Εργασίες

Η παρακολούθηση προσώπων έχει γνωρίσει μεγάλη άνθιση τα τελευταία χρόνια στον ερευνητικό τομέα, με αποτέλεσμα να έχουν δημοσιευθεί μία πληθώρα μεθόδων και προτάσεων. Το κεφάλαιο αυτό δεν είναι δυνατόν να τις καλύψει όλες, αλλά γίνεται μία προσπάθεια να γνωρίσουμε στον αναγνώστη τις πιο βασικές και ουσιώδεις.

Μπορούμε να ξεχωρίσουμε τις υπάρχουσες μεθόδους ανάλογα με το τι ακριβώς παρακολουθείται. Έτσι έχουμε μεθόδους που παρακολουθούν μία ομάδα από εικονοστοιχεία (pixels) με βάση το χρώμα ή την κίνηση. Επίσης παρακολουθούνται ακμές, π.χ. το περίγραμμα του προσώπου σε σχέση με το φόντο. Πιο πολύπλοκες μέθοδοι παρακολουθούν συγκεκριμένα χαρακτηριστικά προσώπου, τα οποία μπορεί και να περιορίζονται γεωμετρικά ή χρησιμοποιούν φόρμες, σταθερές ή ελαστικές, τις οποίες και προσπαθούν να ταιριάξουν στο πρόσωπο. Ακόμα, έχει χρησιμοποιηθεί και η οπτική ροή για να παρακολουθούνται πρόσωπα. Ξεχωρίζουμε επίσης τους δισδιάστατους παρακολουθητές και τους τρισδιάστατους. Οι πρώτοι, όπως φανερώνει και το όνομά τους, παρακολουθούν το πρόσωπο στις δύο διαστάσεις, επομένως μπορούν να δώσουν πληροφορίες μόνο για τη μεταφορά του προσώπου στους δύο άξονες, x και y , της εικόνας. Αντίθετα, οι τρισδιάστατοι παρακολουθητές έχουν τη δυνατότητα να δώσουν πληροφορίες σχετικές και με την περιστροφή του προσώπου. Υπολογίζουν μερικούς ή όλους από τους έξι βαθμούς ελευθερίας (βλ. και σχήμα 1.1). Είναι σε θέση να ανακτήσουν τρισδιάστατες παραμέτρους του προσώπου παρακολουθώντας σημαντικά σημεία, χαρακτηριστικά ή μικρά δισδιάστατα κομμάτια. Ένας άλλος τρόπος να ανακτηθεί η τρισδιάστατη δομή, η εστιακή απόσταση και ο προσανατολισμός του προσώπου από τη δισδιάστατη παρακολούθηση είναι η χρήση των φίλτρων Kalman. Για το σκοπό αυτό χρησιμοποιούνται και τρισδιάστατα μοντέλα προσώπων βασισμένα σε στατιστικές μελέτες. Παρακάτω θα κατηγοριοποιήσουμε ορισμένες μεθόδους και θα τις παρουσιάσουμε πιο αναλυτικά.

Οι Rao και Mersereau [65] προτείνουν ένα σύστημα παρακολούθησης ενός προσώπου βασισμένο στη στατιστική μοντελοποίηση του χρώματος (statistical color modeling) του ανθρώπινου δέρματος και στις ελαστικές φόρμες (deformable templates). Συγκεκριμένα προτείνουν μία δομή δισδιάστατου μοντέλου Hidden Markov (HM), η οποία

αποκομίζεται εμπεδώνοντας καταστάσεις εντός περιοχών μίας ελαστικής φόρμας (state-embedded deformable template, SEDT). Με αυτόν τον τρόπο, κάθε περιοχή της ελαστικής φόρμας συνεπάγει μία κατανομή πιθανότητας προερχόμενη από παρατηρήσεις. Αυτή η δομή επιτρέπει τον υπολογισμό της πιθανότητας $P[\text{εικόνα}|\text{φόρμα}]$. Η φόρμα που μεγιστοποιεί την πιθανότητα αυτή παρέχει την καταλληλότερη τμηματοποίηση της εικόνας. Οι Huang και Chen [39] κατάφεραν να επεκτείνουν τον προηγούμενο αλγόριθμο ώστε να παρακολουθεί πολλαπλά πρόσωπα, μέρος των οποίων μπορεί να επικαλύπτεται κιόλας. Αυτό επιτυγχάνεται εφαρμόζοντας περιορισμούς στη μεταβολή της ταχύτητας και της κλίμακας των προσώπων. Υποθέτουν ότι κατά τη διάρκεια της επικάλυψης, η ταχύτητα και το μέγεθος του προσώπου που επικαλύπτεται δεν αλλάζουν σημαντικά. Τα μοντέλα HM εξασφαλίζουν ένα δυνατό στατιστικό πλαίσιο εργασίας για την ανάλυση μονοδιάστατων τυχαίων διαδικασιών, αλλά η χρήση τους στις δύο διαστάσεις δεν είναι τόσο βολική. Αποδεικνύεται ότι έχουν εκθετική πολυπλοκότητα [47] και γι' αυτό το λόγο χρησιμοποιούνται ψευδο-2-Δ μοντέλα HM, τα οποία με τη σειρά τους δεν ενσωματώνουν περιορισμούς σχήματος, καθότι κάθε γραμμή αναλύεται ανεξάρτητα.

Στην ίδια κατεύθυνση κινούνται και οι Yang και Waibel [98], δημοσιεύοντας μία μέθοδο παρακολούθησης προσώπων σε πραγματικό χρόνο βασισμένη σε κανονικοποίηση χρώματος. Παρουσιάζουν ένα στοχαστικό μοντέλο που χαρακτηρίζει την κατανομή του χρώματος του ανθρώπινου δέρματος. Η πληροφορία που εξάγεται από το μοντέλο αυτό αποδεικνύεται αρκετή για την παρακολούθηση ενός προσώπου σε διαφορετικές πόζες και όψεις. Επίσης, το μοντέλο προσαρμόζεται σε διαφορετικά άτομα και διαφορετικές συνθήκες φωτισμού σε πραγματικό χρόνο. Ακόμα, χρησιμοποιούν ένα μοντέλο κίνησης για το χειρισμό της κίνησης του προσώπου. Πρόκειται για ένα απλό μοντέλο, το οποίο χρησιμοποιείται για την εκτίμηση της δισδιάστατης μεταφοράς του προσώπου στην εικόνα με βάση την προηγούμενη ταχύτητα, έτσι ώστε να περιορίζεται η περιοχή αναζήτησης και να είναι εφικτή η λειτουργία του συστήματος σε πραγματικό χρόνο.

Ο Bradski [8] μετατρέπει τον αλγόριθμο mean shift ([24]), ο οποίος αποτελεί μία σθεναρή μη παραμετρική τεχνική για την ανάβαση κλίσεων πυκνοτήτων (density gradients) και ουσιαστικά εντοπίζει τις επικρατούσες τιμές κατανομών πιθανοτήτων. Ο αλγόριθμος αυτός ποτέ δεν προορίζονταν σαν αλγόριθμος παρακολούθησης, αλλά ο συγγραφέας αποδεικνύει το αντίθετο. Η κατανομή πιθανότητας εδώ είναι τα χρώματα της εικόνας, κάτι που επιτυγχάνεται χρησιμοποιώντας ιστογράμματα χρώματος. Οι κατανομές του χρώματος που προέρχονται από μία ακολουθία βίντεο αλλάζουν με το χρόνο, οπότε ο αλγόριθμος mean shift πρέπει να προσαρμόζεται δυναμικά στην κατανομή πιθανότητας την οποία και παρακολουθεί. Το όλο σχήμα το ονόμασε συνεχώς προσαρμόσιμο αλγόριθμο mean shift (continuously adaptive mean shift algorithm, CAMSHIFT). Το σύστημα αυτό παρακολουθεί τη μεταφορά στους άξονες x και y , καθώς και το εμβαδόν της περιοχής που περιέχει χρώμα ανθρώπινου προσώπου. Το εμβαδόν αυτό θεωρείται ανάλογο της απόστασης του προσώπου από το φακό και υποστηρίζεται ότι έτσι είναι δυνατόν και η παρακολούθηση της μεταφοράς στον άξονα z . Επίσης, παρακολουθείται και η

περιστροφή του προσώπου ως προς τον z άξονα (roll), σύμφωνα με την κλίση που έχει η προαναφερθέντα περιοχή. Έχουμε, λοιπόν, ανάκτηση μόνο των τεσσάρων από τους έξι βαθμούς ελευθερίας (βλ. και σχήμα 1.1).

Μία πιο πολύπλοκη πρόταση συναντάμε στις δημοσιεύσεις των Colmenarez, Lopez και Huang [52], [13]. Το σύστημα παρακολούθησής τους αποτελείται από τρεις μονάδες : (α) 3-Δ μοντελοποίηση του προσώπου, (β) 2Δ-3Δ εκτίμηση προσανατολισμού του προσώπου και (γ) αντιστοίχιση με φόρμα βασισμένη σε σύνθεση. Δεδομένης μιας εκτίμησης του προσανατολισμού του προσώπου, το 3-Δ μοντέλο εξασφαλίζει μία συνθετική όψη του προσώπου, από την οποία κατασκευάζονται φόρμες. Τα χαρακτηριστικά του προσώπου ανιχνεύονται με βάση αυτές τις φόρμες με αντιστοίχιση. Τελικά, χρησιμοποιούνται οι θέσεις των χαρακτηριστικών στις δύο διαστάσεις καθώς και οι αντίστοιχες θέσεις τους στις τρεις διαστάσεις του μοντέλου για την εκτίμηση του νέου προσανατολισμού. Για την πρόβλεψη χρησιμοποιούνται φίλτρα Kalman με βάση τις προηγούμενες εικόνες. Η διαδικασία αυτή εφαρμόζεται επαναληπτικά στην ίδια εικόνα και έτσι επιτυγχάνεται βελτίωση της εκτίμησης του προσανατολισμού. Επειδή ακριβώς ο αλγόριθμος χρησιμοποιεί ανάλυση από σύνθεση, μέσω του 3-Δ μοντέλου, το σύστημα είναι ικανό να ξεπερνάει πολλά συνηθισμένα προβλήματα, όπως είναι η συσσώρευση λάθους σε μεγάλες ακολουθίες, οι εναλλασσόμενες συνθήκες φωτισμού και οι επικαλύψεις. Όμως, σε τέτοιου είδους προσεγγίσεις, όπως είναι η ανάλυση από σύνθεση, η αντίληψη της σκηνής περιορίζεται από την ικανότητα του μοντέλου να αναπαραστήσει τη σκηνή.

Στην κατηγορία των 3-Δ μοντέλων για την παρακολούθηση ανήκει και η πρόταση των DeCarlo, Metaxa [16]. Επιστρατεύεται η οπτική ροή, η οποία παρέχει έναν περιορισμό στην κίνηση του ελαστικού μοντέλου. Επίσης, εξασφαλίζεται η ικανοποίηση του περιορισμού όταν συνδυάζεται με την πληροφορία των ακμών, η οποία βοηθάει στην καταπολέμηση της συσσώρευσης λάθους. Η επιβολή του περιορισμού χαλαρώνεται χρησιμοποιώντας ένα φίλτρο Kalman, το οποίο επιτρέπει την ελεγχόμενη παράβαση του περιορισμού και η οποία γίνεται αναγκαία λόγω του θορύβου που υπάρχει στην πληροφορία της οπτικής ροής, ενώ επιτρέπει ταυτόχρονα τον πιο αποτελεσματικό συνδυασμό των πληροφοριών οπτικής ροής και ακμών. Το 3-Δ ελαστικό μοντέλο του προσώπου επιτρέπει την ανάκτηση της πόζας.

Μία διαφορετική προσέγγιση του θέματος συναντάμε στο σύστημα των Maurer και Malsburg [54]. Το σύστημα αυτό παρακολουθεί συγκεκριμένα χαρακτηριστικά σημεία του προσώπου. Κάθε χαρακτηριστικό περιγράφεται με ένα διάνυσμα. Η κάθε συνιστώσα του διανύσματος είναι η απόκριση ενός κύματος Gabor, συγκεκριμένης συχνότητας και προσανατολισμού, που εξάγεται σε κάποιο σημείο. Ένα σύνολο από τέτοια διανύσματα από διάφορα σημεία σχηματίζουν ένα γράφο, ο οποίος περιγράφει το πρόσωπο στην εικόνα. Για την παρακολούθηση υπολογίζεται η μετατόπιση καθενός χαρακτηριστικού ξεχωριστά σε δύο συνεχόμενες εικόνες, με μία μέθοδο που αναπτύχθηκε για την εκτίμηση ανομοιότητας σε στερεοσκοπικές εικόνες βασισμένη στα [23], [77]. Επίσης, στην εργασία αυτή γίνεται εκτίμηση της πόζας υποθέτοντας ότι οι κόμβοι του γράφου,

δηλαδή τα χαρακτηριστικά σημεία, βρίσκονται περίπου πάνω σε μία επίπεδη επιφάνεια και υπολογίζοντας τον αφινικό μετασχηματισμό του γράφου από την εικόνα 1 (αρχικοποίηση) ως την εικόνα n (εικόνα ζητούμενης εκτίμησης πόζας). Σημειώνεται ότι κατά την αρχικοποίηση το πρόσωπο πρέπει υποχρεωτικά να κοιτάει το φακό.

Το πρόβλημα της εκτίμησης της πόζας με χρήση 3-Δ μοντέλων έχει ως αποτέλεσμα να υπάρχουν συνήθως παραπάνω μετρήσεις απ' ό,τι χρειάζονται. Η λύση δίδεται, γενικά, χρησιμοποιώντας όλες τις μετρήσεις και κάποιου είδους κριτήριο ελαχίστων τετραγώνων για να βρεθεί η καλύτερη. Τέτοιου είδους προσεγγίσεις παραπλανούνται όμως εύκολα από τις λανθασμένες μετρήσεις. Μία λανθασμένη επιλογή εκτίμησης σε μία εικόνα μπορεί πολύ εύκολα να οδηγήσει σε λάθος, το οποίο είναι μη αναστρέψιμο, στην επόμενη. Ένας εναλλακτικός τρόπος προσέγγισης είναι να χρησιμοποιηθεί ένα υποσύνολο των δεδομένων για να εκτιμηθεί η πόζα, όπως μας προτείνουν οι Gee και Cipolla [33]. Ένα ελάχιστο υποσύνολο των δεδομένων εφοδιάζει τον εκτιμητή της πόζας και μία μέθοδο σθεναρής εκτίμησης (robust regression) διαλέγει την καλύτερη εκτίμηση. Συμπεράσματα που εξάγονται με βάση μπαεσιανών κανόνων (bayesian inference) στο στάδιο της εκτίμησης χρησιμοποιούνται για την επανεξέταση των μετρήσεων, που πάρθηκαν από μία εικόνα, με εκτιμήσεις από προηγούμενες εικόνες, εξαλείφοντας την ανάγκη για επί πλέον φιλτράρισμα των εκτιμήσεων. Όσο υπάρχει ένα τέτοιο καλό και ακριβές υποσύνολο μετρήσεων, μπορούν να αγνοηθούν τα υπόλοιπα δεδομένα και έτσι μεγάλα λάθη δεν θα έχουν επίδραση στην παρακολούθηση. Η υπόθεση κλειδί είναι να βρεθεί αυτό το υποσύνολο. Συνήθως χρησιμοποιείται για το σκοπό αυτό ένα κριτήριο ομαδοποίησης ή πλειοψηφίας : αν η εκτίμηση που “προτείνει” ένα υποσύνολο συμφωνεί με μια σημαντική αναλογία των άλλων μετρήσεων, τότε αυτή η εκτίμηση είναι πολύ πιθανών να είναι σωστή. Εδώ επιστρατεύτηκε ο αλγόριθμος RANSAC (RANdom SAMple Consensus) [22]. Το αποτέλεσμα είναι ένα σύστημα που επιτελεί αρκετά καλά το δύσκολο έργο της παρακολούθησης του ανθρώπινου προσώπου, ακόμα και όταν το πρόσωπο εξαφανίζεται μερικώς.

Οι Sobottka και Pitas [71] χρησιμοποιούν ελαστικές καμπύλες (snakes) για την παρακολούθηση του περιγράμματος του προσώπου. Η ανίχνευση του προσώπου γίνεται με βάση το χρώμα και το σχήμα. Τα ανιχνευθέντα πρόσωπα επιβεβαιώνονται με την ανίχνευση χαρακτηριστικών προσώπου στο εσωτερικό. Η παρακολούθηση του περιγράμματος γίνεται με μία ελαστική καμπύλη. Οι εσωτερικές δυνάμεις, που επηρεάζουν την καμπύλη, οδηγούν σε περιορισμούς ομαλότητας στο περίγραμμα, ενώ οι εξωτερικές προσελκύουν το περίγραμμα σε ακμές. Το περίγραμμα παρακολουθείται τοποθετώντας την ελαστική καμπύλη από την εικόνα i στην εικόνα $i + 1$, και η καλύτερη τοποθέτηση της καμπύλης εξασφαλίζεται ελαχιστοποιώντας την ενέργεια. Τα χαρακτηριστικά προσώπου παρακολουθούνται με αντιστοίχιση μπλοκ (block matching).

Σημαντική είναι η συμβολή του Kentaro Toyama, στο πεδίο αυτό της έρευνας. Ξεκίνησε μαζί με το Hager αναπτύσσοντας μία θεωρία που ονόμασε Incremental Focus of Attention (IFA) [80]. Πρόκειται για ένα πλαίσιο εργασίας το οποίο προσθέτει σθεναρότητα

στους περισσότερους υπάρχοντες τότε αλγόριθμους παρακολούθησης χαρακτηριστικών προσώπου, που λειτουργούσαν σε πραγματικό χρόνο. Η ουσία του IFA είναι μία δομή, η οποία, δεδομένης μιας περιοχής μίας εικόνας προς αναζήτηση, εστιάζει αποτελεσματικά την προσοχή του συστήματος παρακολούθησης σε ένα μικρό σύνολο σχηματισμών, το οποίο περιέχει και το σχηματισμό των χαρακτηριστικών-στόχων (που θέλουμε να παρακολουθήσουμε). Όταν οπτικές διαταραχές γίνονται αιτία για να χάσει το σύστημα το στόχο, το πλαίσιο εργασίας IFA εξασφαλίζει και τη δυνατότητα να αρχικοποιηθεί ξανά ο παρακολουθητής. Εφόσον ένα σύστημα βασισμένο σε IFA αποτελείται από πολλαπλούς παρακολουθητές, διαφορετικής ακρίβειας, η αποτυχία ενός παρακολουθητή σημαίνει ότι ένας άλλος, λιγότερο ακριβής, θα αναλάβει. Με αυτόν τον τρόπο, η πληροφορία που εξασφαλίζεται από συστήματα παρακολούθησης βασισμένα σε IFA εκφυλίζεται καθώς διάφορες υποθέσεις για το περιβάλλον παραβιάζονται. Στη συνέχεια [81] δημοσίευσαν ένα άρθρο στο οποίο εξηγούν τη σθεναρότητα που παρέχουν προ-αποτυχίας (ante-failure) και μετά-αποτυχίας συστήματα¹ (post-failure). Τονίζουν τη δύναμη της σθεναρότητας των μετά-αποτυχίας συστημάτων σε προβλήματα τεχνητής νοημοσύνης (TN) εξετάζοντας προσεκτικά ένα παράδειγμα παρακολούθησης κίνησης. Στο [79] ουσιαστικά υλοποιεί τη θεωρία IFA χρησιμοποιώντας έξι επίπεδα, δηλαδή έξι συστήματα παρακολούθησης διαφορετικής ακρίβειας, από ένα σύστημα επιλογής εικονοστοιχείων με βάση το χρώμα του ανθρώπινου δέρματος (επίπεδο 1), μέχρι ένα σύστημα παρακολούθησης χαρακτηριστικών προσώπου (επίπεδο 6). Το αποτέλεσμα είναι ένα πλαίσιο που παρακολουθεί πρόσωπα και εκτιμάει την πόζα ανακτώντας και τους έξι βαθμούς ελευθερίας (επίπεδο 6). Καθώς οι συνθήκες φθείρονται, η παρακολούθηση συνεχίζεται αλλά με μερική εκτίμηση της πόζας (στο επίπεδο 3 ανακτάται η 2-Δ μεταφορά στην εικόνα, χωρίς όμως καμία πληροφορία σχετικά με τον προσανατολισμό). Τέλος, όταν η παρακολούθηση αποτυγχάνει, το σύστημα αρχικοποιείται ξανά (επίπεδο 1). Στο [78] βρίσκει χρήση το προηγούμενο πλαίσιο για την παρακολούθηση προσώπων, απαρτίζοντας μέρος μίας εφαρμογής που αντικαθιστά το ποντίκι του υπολογιστή και ο κέρσορας μετακινείται ανάλογα με το πού κοιτάει ο χρήστης πάνω στην οθόνη. Σε μια πιο πρόσφατη δημοσίευσή του μαζί με τους Gemmell, Zitnick, Kang και Seitz [34] προτείνει ένα σύστημα παρακολούθησης προσώπων και ματιών για τηλε-συνδιασκέψεις. Τα πρόσωπα και τα μάτια των συνομιλητών παρακολουθούνται για να αντικατασταθούν στην κάθε οθόνη με συνθετικά 3-Δ μοντέλα, έτσι ώστε να μπορεί καθένας να ξέρει προς τα που κοιτάνε οι συνομιλητές του, σε ποιον απευθύνονται κάθε φορά και αν έχουν την προσοχή τους.

Οι Valente και Dugelay [89], [88] προτείνουν έναν οπτικό κύκλο ανάδρασης (visual feedback loop) για παρακολούθηση προσώπων. Αντί να γίνεται προσπάθεια να προσαρμόζεται ένα περίπου ρεαλιστικό τεχνητό 3-Δ πρόσωπο σε κάποιο άτομο, κατασκευάζεται από ακριβή δεδομένα απόστασης μία συγκεκριμένη υφή προσώπου για να χρησιμοποιη-

¹Προ-αποτυχίας σθεναρά συστήματα αντιστέκονται στην αποτυχία, ενώ μετά-αποτυχίας συστήματα ενσωματώνουν την ικανότητα ανάκτησης αφού αποτύχουν.

θεί στη μοντελοποίηση, η ρεαλιστικότητα της οποίας επιτρέπει στις μονάδες ανάλυσης και σύνθεσης να συνεργαστούν στο επίπεδο της εικόνας χρησιμοποιώντας άμεσα 2-Δ πρότυπα, τα οποία συντέθηκαν από το μοντέλο του προσώπου. Σε αντίθεση με άλλους κύκλους ανάδρασης, που συναντώνται στη βιβλιογραφία, εδώ δεν χειρίζονται αποκλειστικά τα πολύπλοκα 3-Δ γεωμετρικά δεδομένα του μοντέλου του προσώπου, έτσι ώστε να επιτυγχάνεται χειρισμός σε πραγματικό χρόνο. Η κυρίως προσφορά συνίσταται σε ένα ολοκληρωμένο πλαίσιο παρακολούθησης προσώπων και εκτίμησης πόζας με λίγες υποθέσεις για τη σταθερή κίνηση του προσώπου (rigid face motion), επιτρέποντας μεγάλες περιστροφές εκτός του πλαισίου της εικόνας και χωρίς σηματοδοτήσεις ή ειδικό μακιγιάζ στο πρόσωπο του χρήστη για να καθιστούν ευκολότερη την παρακολούθηση. Το πλαίσιο που προτείνεται τροφοδοτεί τη διαδικασία παρακολούθησης των χαρακτηριστικών προσώπου με τα συνθετικά πρότυπα ελεγχόμενο από ένα εκτεταμένο φίλτρο Kalman (extended Kalman filter, EKF). Εντός του πλαισίου αυτού, παρουσιάζονται πρωτότυπες και αποτελεσματικές γεωμετρικές και φωτομετρικές τεχνικές μοντελοποίησης και μία επαναδιατύπωση ενός αλγορίθμου αντιστοίχισης μπλοκ, ώστε να γίνεται εφικτή η αντιστοίχιση των συνθετικών προτύπων σε πραγματικές εικόνες και να αποφεύγονται περιοχές του φόντου. Το όλο πλαίσιο χρησιμοποιείται επιτυχώς για την παρακολούθηση και ανάκτηση της τρισδιάστατης θέσης και του προσανατολισμού του προσώπου και για τη δημιουργία ενός τεχνητού περιβάλλοντος τηλε-συνδιασκέψεων.

Στην ίδια κατεύθυνση κινήθηκαν νωρίτερα και οι Jebara, Pentland [43]. Μία προσέγγιση κλειστού κύκλου (closed loop) προτείνεται, η οποία αξιοποιεί τη δομή από κίνηση (structure from motion) για να παράγει ένα 3-Δ μοντέλο ενός προσώπου και στη συνέχεια επανατροφοδοτεί στο σύστημα την εκτίμηση της δομής για να περιορίσει την παρακολούθηση χαρακτηριστικών στην επόμενη εικόνα. Το σύστημα αρχικοποιείται χρησιμοποιώντας ταξινόμηση του χρώματος του ανθρώπινου δέρματος, εκμετάλλευση συμμετριών, τρισδιάστατη παραμόρφωση και ιδιοπρόσωπα (eigenfaces) για να ανιχνεύσει κάποιο πρόσωπο. Οι τροχιές των χαρακτηριστικών υπολογίζονται μέσω παρακολούθησης βασισμένη σε συσχέτιση αθροισμάτων τετραγωνικών διαφορών (sum squared differences, SSD). Ταυτόχρονα οι τροχιές επεξεργάζονται με ένα επεκταμένο φίλτρο Kalman (extended Kalman filter, EKF) για τη σταθερή ανάκτηση της τρισδιάστατης δομής, της γεωμετρίας της κάμερας και της πόζας του προσώπου. Επίσης, προσαρμόστηκε μία εκτίμηση με βάρη στο φίλτρο (weighted estimation). Επιπλέον, η εκτίμηση της δομής περιορίζεται χρησιμοποιώντας παραμετρικά μοντέλα δομών προσώπων (ιδιοπρόσωπα, eigenheads). Η εκτίμηση της τρισδιάστατης κατάστασης και κίνησης του προσώπου από το φίλτρο Kalman προβλέπει την τροχιά των χαρακτηριστικών, η οποία περιορίζει την περιοχή αναζήτησης για την επόμενη εικόνα.

Στην πρόταση των Gong, Psarrou, Katsouli και Palavouzi [35] το πρόσωπο αναπαρίσταται ως μία έλλειψη στον τρισδιάστατο χώρο. Η χρήση έλλειψης παρουσιάζει τα εξής πλεονεκτήματα : πρώτον, το ανθρώπινο πρόσωπο μπορεί να μοντελοποιηθεί με μία έλλειψη με αρκετή ακρίβεια. Δεύτερον, η έλλειψη, όπως και τα σημεία χαρακτηριστικών

προσώπων, διατηρείται κάτω από προοπτικούς (perspective) ή προβολικούς (projective) μετασχηματισμούς. Τρίτον, οι ελλείψεις εμπεριέχουν συμπαγή γενικές πληροφορίες του προσώπου και θεωρούνται πιο σθεναρές ενάντια στο θόρυβο. Τέταρτον, οι αντιστοιχίες μεταξύ των δύο και τριών διαστάσεων μπορούν να επιτευχθούν πιο εύκολα. Η παρακολούθηση απαιτεί την συνεχή ανανέωση των παραμέτρων της έλλειψης, καθώς το πρόσωπο κινείται στο χρόνο. Για το σκοπό αυτό χρησιμοποιούνται φίλτρα Kalman, τα οποία μοντελοποιούν τη δυναμική των παραμέτρων αυτών και προβλέπουν τη μελλοντική κίνηση της έλλειψης, ώστε να εξασφαλίζεται περισσότερο ακριβή εκτίμηση της θέσης της στο χρόνο.

Την ίδια προσέγγιση ακολουθούν και οι Ji, Hu στο [45]. Υποθέτουν ότι το σχήμα ενός τρισδιάστατου προσώπου μπορεί να προσεγγιστεί με μία έλλειψη, της οποίας οι λόγοι των διαστάσεων είναι γνωστοί. Πρώτα εφαρμόζεται μία ανίχνευση έλλειψης για να εντοπιστεί το πρόσωπο στην εικόνα και στη συνέχεια εκτιμούνται η θέση και ο προσανατολισμός του προσώπου στον τρισδιάστατο χώρο. Η ανίχνευση του προσώπου διευκολύνεται από τις φυσιολογικές ιδιότητες που έχουν τα μάτια κάτω από ειδικό υπέρυθρο φωτισμό και από γεωμετρικούς περιορισμούς. Η έλλειψη παρακολουθείται χρησιμοποιώντας φίλτρα Kalman.

Η ιδέα της μοντελοποίησης του προσώπου με υφή τοποθετημένη πάνω σε έναν κύλινδρο προτάθηκε από τους La Cascia, Sclaroff και Athitsos [11]. Η παρακολούθηση διατυπώνεται ως ένα πρόβλημα ταιριάσματος της υφής, που είναι τοποθετημένη πάνω στον κύλινδρο, με την δισδιάστατη εικόνα. Για να λυθούν προβλήματα που προκύπτουν λόγω της απόκλισης φωτισμού και της κίνησης του προσώπου, μοντελοποιείται σαν γραμμικός μετασχηματισμός το λάθος, που προκύπτει από το ταιρίασμα της υφής με φόρμες που παραμορφώνονται (warping templates) και ορθογωνικές φόρμες φωτισμού (orthogonal illumination templates), οι οποίες εφαρμόζονται πάνω στην εικόνα. Γρήγορη και σταθερή παρακολούθηση προκύπτει μέσω ελαχιστοποίησης του λάθους, χρησιμοποιώντας κανονικοποιημένα ελάχιστα τετράγωνα με βάρη (regularized weighted least-squares). Η κανονικοποίηση τείνει να περιορίσει πιθανές αμφιβολίες στις φόρμες παραμόρφωσης και φωτισμού και καθιστά ικανή τη σταθερή παρακολούθηση σε εκτεταμένες ακολουθίες. Το σύστημα αυτό δεν απαιτεί μία ακριβή αρχική τοποθέτηση του μοντέλου. Αρχικοποιείται αυτόματα χρησιμοποιώντας έναν απλό δισδιάστατο ανιχνευτή, με μοναδική υπόθεση ότι το πρόσωπο κοιτάει το φακό στην πρώτη εικόνα.

Οι Wang και Wang [91] θεωρούν ότι χαρακτηριστικά προσώπου, όπως τα μάτια και το στόμα, αναπαρίστανται ως κοιλάδες σε εικόνες φωτεινότητας και συνεπώς μπορούν να χρησιμοποιηθούν ειδικά φίλτρα που ανιχνεύουν κοιλάδες για τον εντοπισμό τους. Τα χαρακτηριστικά αυτά παρακολουθούνται με βάση αντιστοίχιση προτύπων από εικόνα σε εικόνα, αλλά όταν η απλή αυτή μέθοδος ξεπεράσει ένα προκαθορισμένο κατώφλι, και συνεπώς θα οδηγήσει σε λάθος που θα αυξάνεται και δεν θα είναι αναστρέψιμο, χρησιμοποιούνται οι ανιχνευτές κοιλάδων ώστε να ξανά-αρχικοποιηθεί το σύστημα. Η πρώτη μέθοδος είναι πολύ γρήγορη, αλλά οδηγείται εύκολα σε λάθος, ενώ η δεύτερη

είναι υπολογιστικά ακριβή, οπότε χρησιμοποιείται μόνο στις περιπτώσεις που η πρώτη αποτυγχάνει.

Τέλος, υπάρχουν και σχετικές εργασίες, οι οποίες επιστρατεύουν για την επίλυση του προβλήματος της παρακολούθησης τα νευρωνικά δίκτυα. Ενδεικτικά αναφέρουμε τα [40], [41], που χρησιμοποίησαν γι' αυτό το σκοπό ένα νευρωνικό δίκτυο του τύπου multilayer perceptron, εκπαιδευμένο με backpropagation. Το δίκτυο ανιχνεύει και παρακολουθεί πρόσωπα με βάση το χρώμα και το σχήμα τους. Επίσης, εφαρμόστηκαν νευρωνικά δίκτυα και για την εκτίμηση του προσανατολισμού του προσώπου [36], [63]. Τα νευρωνικά δίκτυα χρησιμοποιούνται σ' αυτήν την περίπτωση για να κατασκευαστεί μία αντιστοιχία μεταξύ των δισδιάστατων προσώπων και τους τρισδιάστατους προσανατολισμούς των προσώπων. Οι προσεγγίσεις αυτές βασίζονται στην εκμάθηση μοντέλων προσανατολισμού από παραδείγματα. Είναι προφανές, ότι η εκπαίδευση των δικτύων αποτελεί μία χρονοβόρα και κουραστική διαδικασία. Επίσης, όσον αφορά τη σχέση απόδοσης με την πολυπλοκότητα της λειτουργίας τους, οι απόψεις διαιρούνται, με επικρατέστερη εκείνη που υποστηρίζει ότι δεν συμφέρουν προς το παρόν.

Κεφάλαιο 3

Προτεινόμενη Μέθοδος

Στο προηγούμενο Κεφάλαιο παρουσιάσαμε μία ανασκόπηση εργασιών σχετικά με το θέμα της παρακολούθησης προσώπων. Αρκετές από τις εργασίες αυτές χρησιμοποιούν το χρώμα για την εκπλήρωση του έργου. Από τη μια μεριά, υπάρχει μεγάλο κίνητρο να χρησιμοποιηθεί το χρώμα για να γίνει η παρακολούθηση προσώπων, γιατί το ανθρώπινο χρώμα ξεχωρίζει, στις περισσότερες περιπτώσεις, από το φόντο. Επίσης, η επεξεργασία του χρώματος είναι συνήθως πιο γρήγορη απ' ό,τι η επεξεργασία οποιουδήποτε άλλου χαρακτηριστικού του ανθρώπινου προσώπου. Τέλος, κάτω από ορισμένες συνθήκες φωτισμού, το χρώμα είναι σταθερό ως προς τον προσανατολισμό του προσώπου. Έτσι προκύπτει ένας τρόπος να παρακολουθείται το ανθρώπινο πρόσωπο, παρακολουθώντας ουσιαστικά το ανθρώπινο χρώμα μέσα στην ακολουθία εικόνων. Από την άλλη μεριά, όμως, το χρώμα δεν είναι φυσικό φαινόμενο, αλλά αντιληπτικό, το οποίο σχετίζεται με τα φασματικά χαρακτηριστικά της ηλεκτρομαγνητικής ακτινοβολίας, στα ορατά μήκη κύματος, που πέφτει στον αμφιβληστροειδή χιτώνα του ματιού. Έτσι, η αναπαράσταση του χρώματος ενός προσώπου, που πετυχαίνουμε μέσω μίας κάμερας, επηρεάζεται από πολλούς παράγοντες, όπως τον περιβάλλοντα φωτισμό, την κίνηση του αντικειμένου κτλ. Ακόμα, διαφορετικές κάμερες παράγουν σημαντικά διαφορετικές τιμές για τα χρώματα, ακόμα και για το ίδιο άτομο κάτω από τις ίδιες συνθήκες φωτισμού. Επίσης, το χρώμα του ανθρώπινου δέρματος διαφέρει από άτομο σε άτομο. Βλέπουμε, λοιπόν, μία πληθώρα σημαντικών παραγόντων που δεν συνιστούν τη χρησιμοποίηση του χρώματος για την παρακολούθηση προσώπων. Συνήθως μαζί με το χρώμα εφαρμόζονται και περιορισμοί σχήματος, γεωμετρικοί περιορισμοί και γενικά διάφορα τεχνάσματα που θα αλληλοσυμπληρωθούν προκειμένου να πραγματοποιηθεί η παρακολούθηση.

Ακόμα, πολλές από τις μεθόδους που αναφέραμε χρησιμοποιούν πολύπλοκα τρισδιάστατα μοντέλα προσώπων για να πετύχουν την παρακολούθηση. Οι προσεγγίσεις αυτές δεν είναι όμως αρκετά ακριβείς, καθώς πολλές από αυτές απαιτούν παρεμβολή (interpolation). Επιπλέον απαιτούν ένα μεγάλο αριθμό από εικόνες διαφορετικών προσώπων, με διαφορετικούς προσανατολισμούς, διαφορετική κλίμακα και διαφορετικό φωτισμό για την εκπαίδευση. Και αυτό γιατί το πρότυπο (pattern) του προσώπου αλλάζει

με διάφορους τρόπους, λόγω της ατομικότητας, της κλίμακας και του φωτισμού. Υπάρχει επίσης και ένα ερωτηματικό σχετικά με την εγκυρότητα της βασικής υπόθεσης ότι υπάρχει μία μοναδική σχέση αιτιατού αποτελέσματος (unique causal-effect relationship) μεταξύ ορισμένων ιδιοτήτων της εικόνας του προσώπου και του τρισδιάστατου προσανατολισμού του προσώπου. Αυτή η υπόθεση δεν έχει επιβεβαιωθεί, και ακόμα και η σχέση μπορεί να μην είναι μοναδική. Ακόμα και υποθέτοντας ότι υπάρχει μία τέτοια σχέση, αυτή μπορεί να είναι εξαιρετικά πολύπλοκη και έτσι να χρειάζεται ένας αρκετά μεγάλος αριθμός εικόνων προς εκπαίδευση για να μπορεί να προσδιοριστεί.

Τέλος, κάμποσες μέθοδοι χρησιμοποιούν τεχνικές οι οποίες χαρακτηρίζονται από υψηλή πολυπλοκότητα και είναι υπολογιστικά ακριβές. Τέτοιες είναι οι ελαστικές καμπύλες (snakes) για την παρακολούθηση περιγραμμάτων, το φίλτρο Κάλμαν και η εκτεταμένη του μορφή (extended Kalman filter, EKF) για την εκτίμηση του προσανατολισμού και για την πρόβλεψη της επόμενης θέσης, και οι ανιχνευτές κοιλάδων για την παρακολούθηση χαρακτηριστικών. Τέτοιες τεχνικές σίγουρα έχουν μέλλον, καθώς η τεχνολογία των υπολογιστών συνεχώς και εξελίσσεται σε πιο υψηλά επίπεδα, επιτρέποντας να γίνονται αρκετά γρήγορα πολύπλοκοι υπολογισμοί, στο οποίο και βασίζονται κυρίως. Προς το παρόν, όμως, με βάση τα διαθέσιμα μέσα, μπορούν να χαρακτηριστούν ασύμφορες.

Για να λάβουμε μία σθεναρή και ολοκληρωτική ανάλυση των αντικειμένων και των κινήσεών τους απαιτείται υψηλού επιπέδου μοντελοποίηση. Χρησιμοποιώντας μία τέτοια προσέγγιση μπορούν να σχεδιαστούν συστήματα που να ταιριάζουν και σε μη στερεά (non-rigid, articulated) αντικείμενα. Εδώ ασχολούμαστε μόνο με στερεά (rigid) αντικείμενα, τα οποία και χαρακτηρίζουν την πλειοψηφία των εφαρμογών και διέπονται από τις εξής αρχές :

- **Μέτρηση** : Ένας αριθμός από χαρακτηριστικά-κλειδιά, συνήθως γωνίες ή ακμές, ανιχνεύονται σε κάθε εικόνα (frame) της ακολουθίας. Είναι γνωστό ότι αυτά τα χαρακτηριστικά αντιστοιχούν σε χαρακτηριστικά ενός τρισδιάστατου μοντέλου του στόχου.
- **Εκτίμηση προσανατολισμού** : Οι συντεταγμένες στην εικόνα των χαρακτηριστικών αυτών χρησιμοποιούνται για να υπολογιστεί η πόζα του αντικειμένου σχετικά με την κάμερα, δηλαδή η σχετική θέση του στόχου και η κατεύθυνσή του.
- **Φιλτράρισμα** : Το αποτέλεσμα της παραπάνω εκτίμησης φιλτράρεται στις περισσότερες περιπτώσεις έτσι ώστε να μειωθεί η επιρροή θορύβου (δηλαδή των λάθους εκτιμήσεων) του ανιχνευτή χαρακτηριστικών. Αυτό γίνεται συνήθως χάρη σε μια ομαλή υπόθεση κίνησης, το οποίο είναι λογικό δεδομένου ότι η κίνηση στερεών αντικειμένων διέπεται από αδράνεια.
- **Πρόβλεψη** : Η θέση (συντεταγμένες) του αντικειμένου προβλέπεται για την επόμενη εικόνα, χρησιμοποιώντας ξανά την υπόθεση κίνησης, και τα χαρακτηριστικά-

κλειδιά ανιχνεύονται σε μικρά παράθυρα γύρω από τις αναμενόμενες συντεταγμένες τους. Αυτό το τοπικό ψάξιμο επιτρέπει στο σύστημα παρακολούθησης να είναι γρήγορο και να ξεπερνάει τα όποια προβλήματα αντιστοίχισης μεταξύ του μοντέλου και των χαρακτηριστικών στην εικόνα.

Προτείνουμε, λοιπόν, μία τέτοιου είδους προσέγγιση για την παρακολούθηση προσώπων. Η παρακολούθηση γίνεται παρακολουθώντας ουσιαστικά σημεία χαρακτηριστικών του προσώπου, όπως είναι τα μάτια, η μύτη και το στόμα. Τα χαρακτηριστικά αυτά είναι τα πιο προφανή και αυτά χάρη στα οποία ξεχωρίζουν κυρίως τα πρόσωπα. Επίσης, η επιλογή των σημείων (βλ. και σχ. 3.1) έγινε με τέτοιο τρόπο, ώστε να προκύψει και μία άλλη, σημαντική ιδιότητα. Μπορούμε, χωρίς μεγάλο σφάλμα, να υποθέσουμε ότι τα σημεία αυτά βρίσκονται πάνω σε ένα επίπεδο. Έτσι καθίσταται και το μοντέλο μας για την εκτίμηση του προσανατολισμού πιο απλό, αφού πρόκειται περί ενός επιπέδου στον τρισδιάστατο χώρο. Ο προσανατολισμός του προσώπου είναι το κάθετο στο επίπεδο διάνυσμα. Υπάρχει σχετική βιβλιογραφία για την εκτίμηση των παραμέτρων κίνησης ενός στερεού επιπέδου κομματιού (rigid planar patch) : [82], [85], [84], [51], [50], [74], στην οποία και βασιστήκαμε. Η πρόβλεψη γίνεται με βάση τις παραμέτρους κίνησης που ανακτήθηκαν και χρησιμοποιείται επίσης και μία τεχνική αντιστοίχισης μπλοκ (block matching) για τη διόρθωση λάθους. Σημειώνουμε ότι ο όρος μοντέλο αναφέρεται μεν σε μία επίπεδη επιφάνεια, όπως αναφέραμε παραπάνω, αλλά δεν πρόκειται για τοποθέτηση υφής πάνω σε επίπεδο ή για κατασκευή ενός τεχνητού “επίπεδου”-προσώπου. Το επίπεδο χρησιμοποιείται για το μοναδικό σκοπό της εκτίμησης της κίνησης και του προσανατολισμού του προσώπου, τοποθετώντας ουσιαστικά πάνω σ’ αυτό τα σημεία που παρακολουθούνται και τίποτε άλλο. Η αντιστοίχιση μπλοκ γίνεται κυρίως μεταξύ δύο εικόνων, της εικόνας που βρίσκεται προς επεξεργασία και της ακριβώς προηγούμενης, χρησιμοποιώντας τις τιμές της φωτεινότητας μόνο. Στην τελική επιλογή συμβάλλουν ελάχιστοι γεωμετρικοί περιορισμοί, οι οποίοι είναι γενικοί και δεν εξαρτώνται αποκλειστικά από το κάθε πρόσωπο. Εδώ φαίνεται πραγματικά η απλότητα της προτεινόμενης μεθόδου.

Η στρατηγική που ακολουθείται είναι η εξής : στην εικόνα I_i γνωρίζουμε ήδη τον προσανατολισμό του προσώπου, δηλαδή το κάθετο διάνυσμα στο επίπεδο (N_i). Επίσης μας είναι γνωστές οι παράμετροι της τρισδιάστατης κίνησης που ακολούθησε το επίπεδο από την εικόνα $I_{(i-1)}$ στην εικόνα I_i , δηλαδή τα διανύσματα της τρισδιάστατης μεταφοράς $T_{(i-1) \rightarrow i}$ και περιστροφής $\Omega_{(i-1) \rightarrow i}$. Ξέρουμε και τις δισδιάστατες συντεταγμένες P_i που έχουν τα σημεία που παρακολουθούμε στην εικόνα I_i . Χρησιμοποιώντας τις προαναφερθείσες παραμέτρους κίνησης γίνεται μία πρόβλεψη των θέσεων των σημείων στην επόμενη εικόνα $I_{(i+1)}$. Υποθέτουμε δηλαδή ότι εφόσον τα σημεία κινήθηκαν έτσι προηγουμένως, θα συνεχίσουν να κινούνται κατά τον ίδιο τρόπο και βρίσκουμε μία εκτίμηση των συντεταγμένων των σημείων στην επόμενη εικόνα $P_{(i+1)}^{est}$. Για να επιβεβαιώσουμε αν όντως συνέχισαν να κινούνται έτσι ή αν έχει αλλάξει κάτι στην κίνησή τους



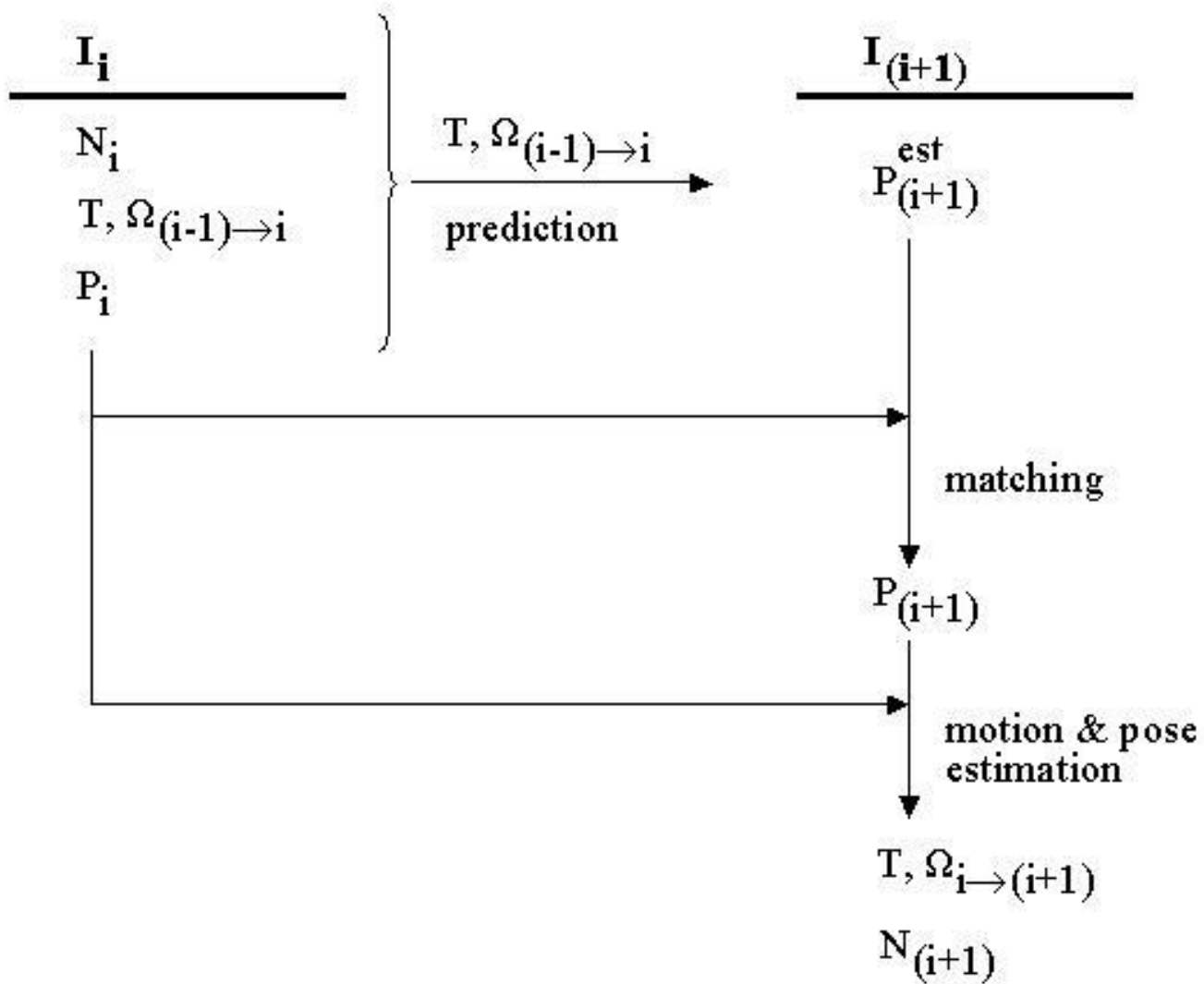
Σχήμα 3.1: Τα σημεία που επιλέχθηκαν για την παρακολούθηση προσώπων με την προτεινόμενη μέθοδο. Τέσσερα σημεία για τα μάτια, δύο για τη μύτη και δύο για το στόμα, τοποθετημένα στα άκρα των χαρακτηριστικών.

Ξεκινάμε μία διαδικασία αντιστοίχισης μπλοκ, χρησιμοποιώντας και γεωμετρικούς περιορισμούς, μεταξύ των σημείων των εικόνων I_i και $I_{(i+1)}$. Έτσι επιδιορθώνονται πιθανά λάθη και βρίσκονται οι πραγματικές συντεταγμένες των σημείων στην εικόνα $I_{(i+1)}$, $P_{(i+1)}$. Αφού λοιπόν έχουμε βρει και τις σωστές θέσεις των σημείων στην επόμενη εικόνα, χρησιμοποιούμε τις θέσεις των σημείων αυτών $P_{(i+1)}$ μαζί με τις προηγούμενες τους θέσεις P_i για να υπολογίσουμε τις πραγματικές παραμέτρους της κίνησης, που έκανε το επίπεδο από την εικόνα I_i στην $I_{(i+1)}$, δηλαδή τα διανύσματα μεταφοράς $T_{i \rightarrow (i+1)}$ και περιστροφής $\Omega_{i \rightarrow (i+1)}$, καθώς και το κάθετο διάνυσμα στο επίπεδο $N_{(i+1)}$, που μας δίνει τον προσανατολισμό του προσώπου στην εικόνα $I_{(i+1)}$. Μία σχηματική περιγραφή της μεθόδου φαίνεται στο σχήμα 3.2.

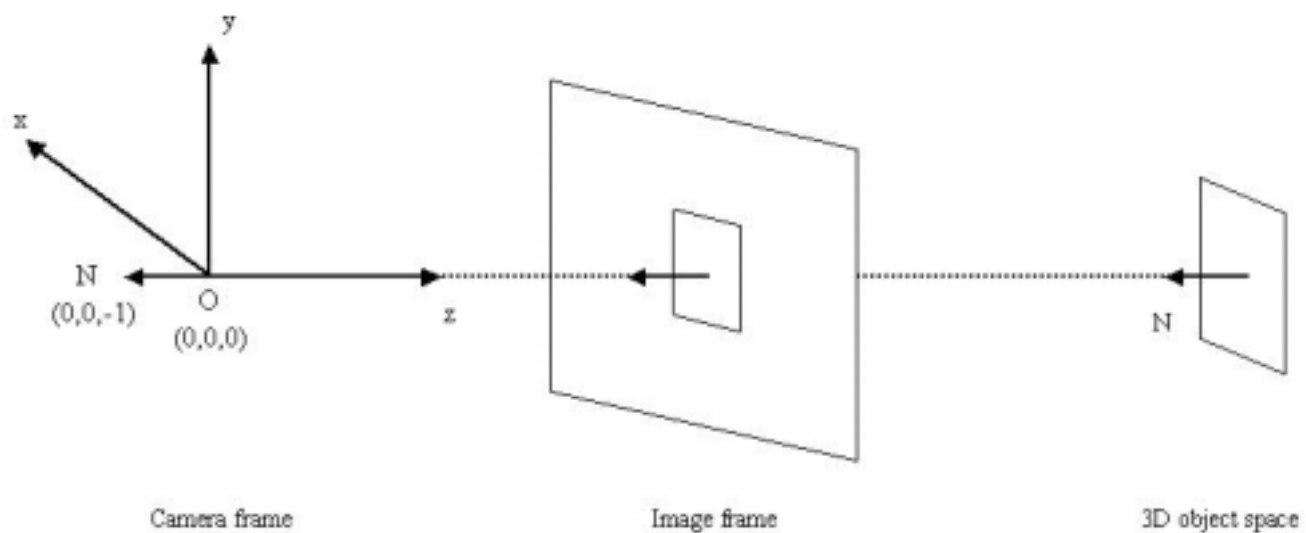
Φυσικά το όλο σύστημα χρειάζεται κάποια αρχικοποίηση. Αυτή περιλαμβάνει τον προσδιορισμό των 2- Δ συντεταγμένων που έχουν τα σημεία στις δύο πρώτες εικόνες της ακολουθίας, P_0 , P_1 καθώς και το διάνυσμα του προσανατολισμού του προσώπου

στην πρώτη εικόνα, N_0 . Η αρχικοποίηση των σημείων γίνεται από το χρήστη, αλλά θα μπορούσε πολύ εύκολα να προσαρμοστούν ελαφρά και να χρησιμοποιηθούν οι ανιχνευτές χαρακτηριστικών προσώπων που αναπτύξαμε στα [1], [27]. Όσον αφορά τον προσανατολισμό του προσώπου, είναι λίγο δύσκολο να υπολογιστεί αυτόματα καθώς η μέθοδος παρουσιάζει ένα μικρό μειονέκτημα, στο οποίο θα αναφερθούμε στο επόμενο Κεφάλαιο. Εφόσον όμως το όλο σύστημα προορίζεται κυρίως για περιβάλλοντα τηλεσυνδιασκέψεων ή για περιπτώσεις όπου κάποιος(α) “μιλάει” κυρίως στην κάμερα (π.χ. παρουσιαστές δελτίων ειδήσεων, εκφωνητές κτλ.) μπορεί εκ των προτέρων να καθοριστεί ένα σημείο, το οποίο καθώς το κοιτάζει το πρόσωπο ξέρουμε τι προσανατολισμό έχει. Το πιο άμεσο και εύκολο βέβαια είναι η αρχικοποίηση να γίνει όταν το πρόσωπο κοιτάει ίσια στο φακό, όπου βέβαια ο προσανατολισμός του δίδεται από το διάνυσμα $N = (0, 0, -1)$ (βλ. και σχήμα 3.3).

Στη συνέχεια θα εξηγήσουμε τον τρόπο με τον οποίο ανακτώνται οι παράμετροι κίνησης και το κάθετο στο επίπεδο διάνυσμα που μας δίνει τον προσανατολισμό του προσώπου στο Κεφάλαιο 4. Έπειτα θα εμβαθύνουμε στην παρακολούθηση των σημείων και στην αντιστοίχιση στο Κεφάλαιο 5. Στο Κεφάλαιο 6 θα παρουσιάσουμε και θα εξηγήσουμε πειραματικά αποτελέσματα που επιβεβαιώνουν την ορθότητα και την αποτελεσματικότητα της προσέγγισής μας, και θα κλείσουμε με βασικά συμπεράσματα και μελλοντικές επεκτάσεις στο Κεφάλαιο 7.



Σχήμα 3.2: Σχηματική αναπαράσταση προτεινόμενης μεθόδου. Διακρίνουμε τα βήματα της πρόβλεψης (prediction), της αντιστοίχισης (matching) και της εκτίμησης κίνησης και προσανατολισμού (motion & pose estimation).



Σχήμα 3.3: Το διάνυσμα προσανατολισμού προσώπου N . Όταν το πρόσωπο, που βρίσκεται στον τρισδιάστατο χώρο (3D object space) και μοντελοποιείται με ένα επίπεδο, κοιτάει ίσια το φακό, που βρίσκεται στο σημείο O του camera frame, το διάνυσμα του προσανατολισμού, δηλαδή το κάθετο διάνυσμα στο επίπεδο, είναι το $N = (0, 0, -1)$.

Κεφάλαιο 4

Έυρεση παραμέτρων κίνησης και 3-Δ διανύσματος κατεύθυνσης

Στο κεφάλαιο αυτό θα παρουσιάσουμε μία άμεση μέθοδο για την εκτίμηση των τρισδιάστατων παραμέτρων κίνησης ενός στερεού επίπεδου κομματιού (rigid planar patch) από δύο συνεχόμενες στο χρόνο εικόνες. Δοκιμάστηκε αρχικά η προσέγγιση του Longuet-Higgins [50], χωρίς όμως να έχει μεγάλη επιτυχία. Έτσι, προτιμήθηκε και υλοποιήθηκε η πρόταση των Tsai και Huang [82], [85], την οποία περιγράφει το Κεφάλαιο αυτό. Καταλήγουμε στην ενότητα 4.5 με βασικά συμπεράσματα, που αφορούν τη συγκεκριμένη προσέγγιση και εξηγούμε τη χρησιμότητά της για την προτεινόμενη μεθοδολογία μας. Πρώτα ορίζεται ένα σύνολο από οκτώ (8) καθαρές παραμέτρους. Αυτές οι παράμετροι μπορούν να προσδιοριστούν μοναδικά από τις δύο δεδομένες εικόνες, λύνοντας ένα σύνολο γραμμικών εξισώσεων. Μετά προσδιορίζονται οι πραγματικές παράμετροι της τρισδιάστατης κίνησης του επιπέδου υπολογίζοντας την Singular Value Decomposition (SVD) ενός 3×3 πίνακα που περιέχει τις αποκαλούμενες “καθαρές” παραμέτρους. Εκτός ενός παράγοντα κλίμακας για τις παραμέτρους της μεταφοράς, το πλήθος των λύσεων είναι ή μία ή δύο και εξαρτάται από την πολλαπλότητα των singular values του πίνακα.

4.1 Βασικές εξισώσεις κίνησης

Για την εκτίμηση των παραμέτρων κίνησης ενός αντικειμένου, όπως την παρουσιάζουμε εδώ, θα υποθέσουμε ότι δουλεύουμε με δύο συνεχόμενες στο χρόνο εικόνες, συγκεκριμένα τις χρονικές στιγμές t_1 και t_2 ($t_1 < t_2$). Η βασική γεωμετρία του προβλήματος απεικονίζεται στο σχήμα 4.1. Θεωρούμε ένα συγκεκριμένο σημείο P στο αντικείμενο του ενδιαφέροντός μας, δηλαδή, στην προκειμένη περίπτωση, πάνω στο πρόσωπο που παρακολουθούμε. Έστω,

$$(x, y, z) = 3 - \Delta \text{ συντεταγμένες ενός σημείου } P \text{ τη χρονική στιγμή } t_1$$

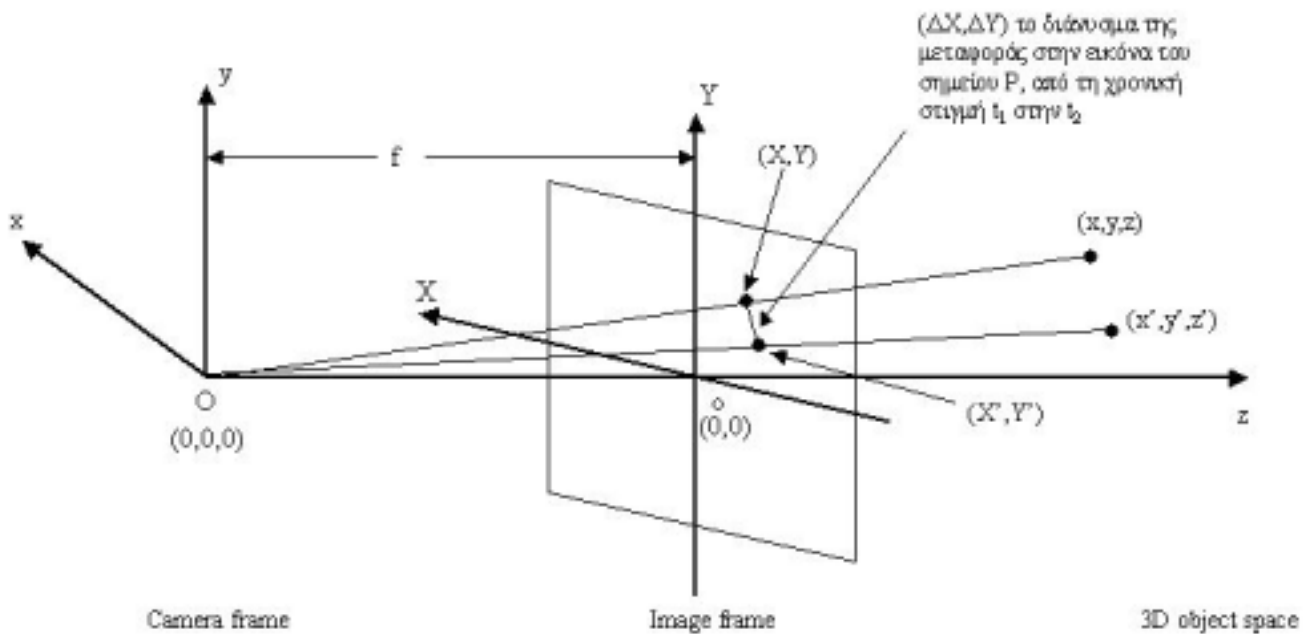
(x', y', z') = 3 – Δ συντεταγμένες του σημείου P τη χρονική στιγμή t_2

(X, Y) = 2 – Δ συντεταγμένες του σημείου P (στην εικόνα) τη χρονική στιγμή t_1

(X', Y') = 2 – Δ συντεταγμένες του σημείου P (στην εικόνα) τη χρονική στιγμή t_2

Είναι προφανές από την εικόνα 4.1 ότι προκύπτουν οι εξής εξισώσεις της προοπτικής προβολής :

$$\begin{aligned} X &= f \frac{x}{z} & , & & Y &= f \frac{y}{z} \\ X' &= f \frac{x'}{z'} & , & & Y' &= f \frac{y'}{z'} \end{aligned} \quad (4.1)$$



Σχήμα 4.1: Βασική γεωμετρία της εκτίμησης της τρισδιάστατης κίνησης.

Υποθέτοντας ότι, από τη χρονική στιγμή t_1 στη t_2 , το τρισδιάστατο αντικείμενό μας υποβλήθηκε σε μεταφορά (translation), περιστροφή (rotation) και γραμμική παραμόρφωση (linear deformation), έχουμε :

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = S \begin{bmatrix} x \\ y \\ z \end{bmatrix} + R \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} \quad (4.2)$$

όπου,

$$S = \begin{bmatrix} S_{11} & S_{12} & S_{13} \\ S_{12} & S_{22} & S_{23} \\ S_{13} & S_{23} & S_{33} \end{bmatrix} \quad (4.3)$$

$$R = \begin{bmatrix} 0 & -\varphi_3 & \varphi_2 \\ \varphi_3 & 0 & -\varphi_1 \\ -\varphi_2 & \varphi_1 & 0 \end{bmatrix} \quad (4.4)$$

$$\varphi_1 = n_1\theta \quad , \quad \varphi_2 = n_2\theta \quad , \quad \varphi_3 = n_3\theta \quad (4.5)$$

$$n_1^2 + n_2^2 + n_3^2 = 1 \quad (\text{υποθέτοντας μικρή περιστροφή}) \quad (4.6)$$

Σημειώνουμε ότι το διάνυσμα $(\Delta x, \Delta y, \Delta z)$ είναι η μεταφορά, ο πίνακας S είναι η γραμμική παραμόρφωση και ο R είναι ο πίνακας της περιστροφής. Η περιστροφή γίνεται γύρω από έναν άξονα που περνάει από την αρχή των αξόνων (O) και έχει διευθύνοντα συνημίτονα τα (n_1, n_2, n_3) . Υποθέτουμε μικρή γωνία περιστροφής θ . Τα $\varphi_1, \varphi_2, \varphi_3$, που ορίστηκαν στην (4.5), είναι οι x, y και z συνιστώσες του διανύσματος περιστροφής με μήκος θ και διευθύνοντα συνημίτονα (n_1, n_2, n_3) .

Η εξίσωση (4.2) αναπαριστά έναν αφινικό μετασχηματισμό :

$$\begin{aligned} \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} &= (S + R) \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} \Leftrightarrow \\ \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} &= \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} \end{aligned} \quad (4.7)$$

και αντιστρόφως, οποιοσδήποτε αφινικός μετασχηματισμός μπορεί να αναλυθεί στην (4.2).

4.2 Κίνηση επίπεδων κομματιών (planar patches)

Σ' αυτήν την ενότητα θα περιοριστούμε σε σημεία που ανήκουν σε ένα επίπεδο κομμάτι με εξίσωση :

$$ax + by + cz = 1 \quad (4.8)$$

τη χρονική στιγμή t_1 . Γίνεται αμέσως φανερό από την (4.1) και την (4.8) ότι :

$$z = \frac{f}{aX + bY + cf} \quad (4.9)$$

και από τις (4.1), (4.9), (4.7) ότι :

$$\begin{aligned} X' &= \frac{a_1 X + a_2 Y + a_3}{a_7 X + a_8 Y + 1} \triangleq T_1(X, Y) \\ Y' &= \frac{a_4 X + a_5 Y + a_6}{a_7 X + a_8 Y + 1} \triangleq T_2(X, Y) \end{aligned} \quad (4.10)$$

όπου :

$$\begin{aligned} a_1 &= \frac{b_{11} + a\Delta x}{b_{33} + c\Delta z} & a_2 &= \frac{b_{12} + b\Delta x}{b_{33} + c\Delta z} \\ a_3 &= \frac{(b_{13} + c\Delta x)f}{b_{33} + c\Delta z} & a_4 &= \frac{b_{21} + a\Delta y}{b_{33} + c\Delta z} \\ a_5 &= \frac{b_{22} + b\Delta y}{b_{33} + c\Delta z} & a_6 &= \frac{(b_{23} + c\Delta y)f}{b_{33} + c\Delta z} \\ a_7 &= \frac{b_{31} + a\Delta z}{(b_{33} + c\Delta z)f} & a_8 &= \frac{b_{32} + b\Delta z}{(b_{33} + c\Delta z)f} \end{aligned} \quad (4.11)$$

Εφόσον πρόκειται όμως για στερεό επίπεδο κομμάτι (rigid planar patch), η (4.11) γίνεται :

$$\begin{aligned} a_1 &= \frac{1 + a\Delta x}{1 + c\Delta z} & a_2 &= \frac{-n_3\theta + b\Delta x}{1 + c\Delta z} \\ a_3 &= \frac{(n_2\theta + c\Delta x)f}{1 + c\Delta z} & a_4 &= \frac{n_3\theta + a\Delta y}{1 + c\Delta z} \\ a_5 &= \frac{1 + b\Delta y}{1 + c\Delta z} & a_6 &= \frac{(-n_1\theta + c\Delta y)f}{1 + c\Delta z} \\ a_7 &= \frac{-n_2\theta + a\Delta z}{(1 + c\Delta z)f} & a_8 &= \frac{n_1\theta + b\Delta z}{(1 + c\Delta z)f} \end{aligned} \quad (4.12)$$

και αυτό γιατί η παραμόρφωση S δεν υφίσταται, δηλαδή η (4.3) μετατρέπεται σε :

$$S = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (4.13)$$

και με βάση την (4.13) η (4.7) γίνεται :

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = R \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} \quad (4.14)$$

που είναι η γενική εξίσωση κίνησης για ένα 3-Δ στερεό αντικείμενο.

Ο R είναι ο 3×3 ορθοκανονικός πίνακας :

$$R = \begin{bmatrix} 1 & -\varphi_3 & \varphi_2 \\ \varphi_3 & 1 & -\varphi_1 \\ -\varphi_2 & \varphi_1 & 1 \end{bmatrix} \Leftrightarrow R = \begin{bmatrix} n_1^2 + (1 - n_1^2) \cos \theta & n_1 n_2 (1 - \cos \theta) - n_3 \sin \theta & n_1 n_3 (1 - \cos \theta) + n_2 \sin \theta \\ n_1 n_2 (1 - \cos \theta) + n_3 \sin \theta & n_2^2 + (1 - n_2^2) \cos \theta & n_2 n_3 (1 - \cos \theta) - n_1 \sin \theta \\ n_1 n_3 (1 - \cos \theta) - n_2 \sin \theta & n_2 n_3 (1 - \cos \theta) + n_1 \sin \theta & n_3^2 + (1 - n_3^2) \cos \theta \end{bmatrix} \quad (4.15)$$

Η παραπάνω ισοδυναμία ισχύει, αφού υποθέτουμε μικρή περιστροφή, δηλαδή γωνία θ πολύ μικρή, οπότε $\cos \theta \approx 1$. Με βάση την υπόθεση αυτή και κάνοντας τις πράξεις, αποδεικνύεται επίσης ότι η (4.12) μπορεί να αντικατασταθεί από την :

$$\begin{aligned} a_1 &= \frac{n_1^2 + (1 - n_1^2) \cos \theta + a \Delta x}{n_3^2 + (1 - n_3^2) \cos \theta + c \Delta z} \\ a_2 &= \frac{n_1 n_2 (1 - \cos \theta) - n_3 \sin \theta + b \Delta x}{n_3^2 + (1 - n_3^2) \cos \theta + c \Delta z} \\ a_3 &= \frac{n_1 n_3 (1 - \cos \theta) + n_2 \sin \theta + c \Delta x}{n_3^2 + (1 - n_3^2) \cos \theta + c \Delta z} \\ a_4 &= \frac{n_1 n_2 (1 - \cos \theta) + n_3 \sin \theta + a \Delta y}{n_3^2 + (1 - n_3^2) \cos \theta + c \Delta z} \\ a_5 &= \frac{n_2^2 + (1 - n_2^2) \cos \theta + b \Delta y}{n_3^2 + (1 - n_3^2) \cos \theta + c \Delta z} \\ a_6 &= \frac{n_2 n_3 (1 - \cos \theta) + n_1 \sin \theta + c \Delta y}{n_3^2 + (1 - n_3^2) \cos \theta + c \Delta z} \\ a_7 &= \frac{n_1 n_3 (1 - \cos \theta) - n_2 \sin \theta + a \Delta z}{n_3^2 + (1 - n_3^2) \cos \theta + c \Delta z} \\ a_8 &= \frac{n_2 n_3 (1 - \cos \theta) + n_1 \sin \theta + b \Delta z}{n_3^2 + (1 - n_3^2) \cos \theta + c \Delta z} \end{aligned} \quad (4.16)$$

όπου για λόγους απλότητας θέσαμε $f = 1$.

Η εξίσωση (4.10) ορίζει μία αντιστοίχιση από το δισδιάστατο χώρο (X, Y) στο δισδιάστατο χώρο (X', Y') . Θα δείξουμε στην επόμενη ενότητα (4.3) ότι, για κάθε καθορισμένη αντιστοιχία μεταξύ των δύο αυτών χώρων μπορεί να υπάρχει μόνο ένα σύνολο τιμών για τις παραμέτρους a_1, \dots, a_8 , που τις αποκαλούμε καθαρές παραμέτρους.

Γεώργιος Σημαντήρης

4.3 Προσδιορισμός των οκτώ καθαρών παραμέτρων

Στην ενότητα αυτή θα εξετάσουμε προσεκτικότερα την εξίσωση (4.10) και θα καταλήξουμε με οδηγό τη θεωρία Lie σε ορισμένα σημαντικά συμπεράσματα. Για ένα συγκεκριμένο σύνολο τιμών για τις παραμέτρους (a_1, \dots, a_8) , η εξίσωση αναπαριστά έναν μετασχηματισμό, ο οποίος αντιστοιχεί το δισδιάστατο χώρο (X, Y) (τις 2-Δ συντεταγμένες της εικόνας τη χρονική στιγμή t_1) στο δισδιάστατο χώρο (X', Y') (τις 2-Δ συντεταγμένες της εικόνας τη χρονική στιγμή t_2). Ας θεωρήσουμε τη συλλογή (collection) G των μετασχηματισμών που αντιστοιχούν σε όλα τα $(a_1, \dots, a_8) \in R^8$. Θα δείξουμε ότι είναι μία συνεχή (Lie) ομάδα διάστασης οκτώ και ότι στην οποιαδήποτε δεδομένη αντιστοιχία από το (X, Y) -χώρο στον (X', Y') -χώρο αντιστοιχεί μόνο ένα σύνολο τιμών για τα (a_1, \dots, a_8) . Περαιτέρω, θα περιγράψουμε μία μέθοδο για τον προσδιορισμό των καθαρών παραμέτρων (a_1, \dots, a_8) από ένα δεδομένο ζεύγος εικόνων στις χρονικές στιγμές t_1 και t_2 .

Στην κλασική θεωρία συνεχών ομάδων (continuous group theory) είναι γνωστό [18] ότι η συλλογή G ικανοποιεί τα τέσσερα αξιώματα ομάδας, δηλαδή, κλειστότητα (το αποτέλεσμα της πράξης μεταξύ δύο στοιχείων της συλλογής είναι κι αυτό στοιχείο της συλλογής), ύπαρξη ουδέτερου στοιχείου (υπάρχει ένα στοιχείο της συλλογής το οποίο αφήνει όλα τα υπόλοιπα στοιχεία αμετάβλητα ως προς την πράξη), ύπαρξη αντιστρόφου στοιχείου (για κάθε στοιχείο της συλλογής υπάρχει ένα άλλο στοιχείο της συλλογής, η πράξη μεταξύ των οποίων έχει σαν αποτέλεσμα το ουδέτερο), και προσεταιριστικότητα $((\alpha \circ \beta) \circ \gamma = \alpha \circ (\beta \circ \gamma))^1$. Επιπλέον, οι συνθέσεις για τις συναρτήσεις $T_1(X, Y)$ και $T_2(X, Y)$ για την ομάδα παραμέτρων a_i στην (4.10) είναι συνεχής. Είναι επίσης γνωστό [18] ότι τα a_i είναι συναρτησιακά ανεξάρτητα. Όμως, δεν είναι γνωστό εάν τα a_i στην (4.10) είναι μοναδικά, δηλαδή, αν μπορεί να υπάρχουν δύο διαφορετικά σύνολα τιμών a_i, a'_i , τέτοια ώστε η (4.10) να δίνει την ίδια αντιστοίχιση $(X, Y) \rightarrow (X', Y')$. Γι' αυτόν το λόγο δεν είναι εύκολο να επαληθευτεί εάν η G είναι μία ομάδα Lie σύμφωνα με το σύγχρονο ορισμό, αφού, στο σύγχρονο ορισμό, συν των ιδιοτήτων που ικανοποιούνται από την κλασική συνεχής ομάδα, σύμφωνα με τον κλασικό ορισμό [10], [18], πρέπει να ικανοποιούνται και μερικές τοπολογικές ιδιότητες. Οι τελευταίες ιδιότητες δεν μπορούν να επαληθευτούν εύκολα εκτός αν είμαστε σίγουροι ότι οι παράμετροι ομάδας a_i είναι μοναδικές. Παρακάτω θα αποδείξουμε ότι η G είναι αυστηρά μία ομάδα Lie και ότι τα a_i είναι μοναδικά.

Υποθέτουμε ότι η G είναι όντως μία ομάδα Lie και ότι (a_1, \dots, a_8) είναι μία αναπαράσταση συντεταγμένων της ομάδας G . Το ουδέτερο στοιχείο της ομάδας είναι προφανώς το $e = (1, 0, 0, 0, 1, 0, 0, 0)$. Τότε οι τελεστές της άλγεβρας Lie σχετικοί με την ομάδα G δίδονται από :

¹Όπου το "ο" συμβολίζει την πράξη.

$$X_j = \left. \frac{\partial T_1}{\partial a_j} \right|_{g=\epsilon} \frac{\partial}{\partial X} + \left. \frac{\partial T_2}{\partial a_j} \right|_{g=\epsilon} \frac{\partial}{\partial Y} \quad (4.17)$$

όπου g χρησιμοποιείται να αναπαραστήσει ένα μέλος της G . Από τη (4.10) παίρνουμε άμεσα :

$$\begin{aligned} X_1 &= X \frac{\partial}{\partial X} \\ X_2 &= Y \frac{\partial}{\partial X} \\ X_3 &= \frac{\partial}{\partial X} \\ X_4 &= X \frac{\partial}{\partial Y} \\ X_5 &= Y \frac{\partial}{\partial Y} \\ X_6 &= \frac{\partial}{\partial Y} \\ X_7 &= -X^2 \frac{\partial}{\partial X} - XY \frac{\partial}{\partial Y} \\ X_8 &= -XY \frac{\partial}{\partial X} - Y^2 \frac{\partial}{\partial Y} \end{aligned} \quad (4.18)$$

Ας θεωρήσουμε το σύνολο των διανυσματικών πεδίων στο διαφορήσιμο σύνολο (X, Y) όπως δίδεται από τη (4.18). Μπορεί εύκολα να επαληθευτεί ότι κανένα από τα X_j δεν μπορεί να εκφραστεί σαν ένας γραμμικός συνδυασμός των άλλων, δηλαδή τα $X_j : j = 1, \dots, 8$ είναι γραμμικά ανεξάρτητα. Επιπλέον, για οποιαδήποτε i, j

$$[X_i, X_j] \triangleq X_i X_j - X_j X_i = \sum_k c_{ij}^k X_k \quad (4.19)$$

όπου τα c_{ij}^k είναι σταθερές. Από αυτές τις δύο ιδιότητες του $\{X_i\}$ συμπεραίνουμε, λόγω του αντιστρόφου του δευτέρου θεμελιώδους θεωρήματος του Lie ([49], [62], [10], [18]), ότι υπάρχει μία μοναδική ομάδα Lie μετασχηματισμού τάξεως οκτώ, της οποίας το $\{X_i\}$ αποτελεί την άλγεβρα Lie. Στη συνέχεια θα αποδείξουμε ότι η G είναι ομάδα.

Από τα [53], [49], [62] μπορούμε να δημιουργήσουμε τις πεπερασμένες εξισώσεις μιας ομάδας Lie μετασχηματισμού με τα λ_i να είναι οι κανονικές συντεταγμένες δευτέρας φύσεως ως εξής :

$$\begin{aligned} X' &= \exp(\lambda_8 X_8) \exp(\lambda_7 X_7) \cdots \exp(\lambda_1 X_1) X \\ Y' &= \exp(\lambda_8 X_8) \exp(\lambda_7 X_7) \cdots \exp(\lambda_1 X_1) Y \end{aligned} \quad (4.20)$$

Αποδεικνύεται στο [83], ότι η (4.20) ισοδυναμεί με :

$$\begin{aligned} X' &= \frac{a_1 X + a_2 Y + a_3}{a_7 X + a_8 Y + 1} \\ Y' &= \frac{a_4 X + a_5 Y + a_6}{a_7 X + a_8 Y + 1} \end{aligned} \quad (4.21)$$

όπου,

$$\begin{aligned} a_1 &= e^{\lambda_1} \lambda \\ a_2 &= \lambda_2 \lambda \\ a_3 &= (\lambda_3 e^{\lambda_1} + \lambda_6 \lambda_2) \lambda \\ a_4 &= e^{\lambda_1} \lambda_4 e^{\lambda_5} \lambda \\ a_5 &= e^{\lambda_5} (1 + \lambda_2 \lambda_4) \lambda \\ a_6 &= [\lambda_6 e^{\lambda_5} + \lambda_4 e^{\lambda_5} (\lambda_3 e^{\lambda_1} + \lambda_6 \lambda_2)] \lambda \\ a_7 &= -(\lambda_7 e^{\lambda_1} + \lambda_8 e^{\lambda_1} \lambda_4 e^{\lambda_5}) \lambda \\ a_8 &= -[\lambda_7 \lambda_2 + e^{\lambda_5} \lambda_8 (1 + \lambda_2 \lambda_4)] \lambda \\ \lambda &= [1 - \lambda_7 (\lambda_3 e^{\lambda_1} + \lambda_6 \lambda_2) - \lambda_6 e^{\lambda_5} \lambda_8 - \lambda_4 e^{\lambda_5} \lambda_8 (\lambda_3 e^{\lambda_1} + \lambda_6 \lambda_2)]^{-1} \end{aligned} \quad (4.22)$$

Συγκρίνοντας τώρα τις (4.10) και (4.21) βλέπουμε ότι η συλλογή G είναι όντως μία ομάδα μετασχηματισμού Lie, και εφόσον τα λ_i είναι κανονικές συντεταγμένες, αυτά είναι μοναδικά, οπότε, από την (4.22) προκύπτει ότι τα a_1, \dots, a_8 είναι επίσης μοναδικά.

Έστω f μία οποιαδήποτε συνάρτηση ορισμένη στο δισδιάστατο χώρο (X, Y) (στην περίπτωση μας, η f είναι οι τιμές της φωτεινότητας των εικονοστοιχείων), τότε από τη θεωρία Lie [49], [62], [10], [18], [56], [53] έχουμε :

$$\Delta f = \sum_{i=1}^8 \beta_i X_i f \quad (4.23)$$

όπου,

$$\begin{aligned} \beta_i &= a_i - e_i \\ e_i &= \text{η } i\text{-συνιστώσα της ομάδας παραμέτρων του ουδέτερου στοιχείου} \\ \Delta f &= f(X', Y') - f(X, Y) = \text{η διαφορά των εικόνων} \end{aligned}$$

(Η εξυπακουόμενη υπόθεση εδώ είναι ότι οι τιμές της φωτεινότητας των εικονοστοιχείων, στις εικόνες που αντιστοιχούν στο ίδιο φυσικό αντικείμενο, είναι ίδιες.) Προφανώς :

$$\begin{aligned} a_1 &= \beta_1 + 1 \\ a_5 &= \beta_5 + 1 \\ a_i &= \beta_i, \quad i = 2, 3, 4, 6, 7, 8. \end{aligned}$$

Η εξίσωση (4.23) χρησιμοποιείται για τον προσδιορισμό των β_i , οπότε και για των a_i . Διαλέγουμε, λοιπόν, οκτώ ή περισσότερα σημεία (X, Y) , υπολογίζουμε για κάθε σημείο τα Δf και $X_i, i = 1, \dots, 8$ και αντικαθιστούμε στην (4.23) για να πάρουμε οκτώ ή περισσότερες εξισώσεις, οι οποίες είναι γραμμικές στα οκτώ άγνωστα β_i . Μετά υπολογίζουμε τη λύση με ελάχιστα τετράγωνα.

Στη συγκεκριμένη περίπτωση, μπορούμε απ' ευθείας από τη (4.21) να υπολογίσουμε τα a_i . Κάνοντας τις πράξεις, από τη (4.21) προκύπτει :

$$\left. \begin{aligned} a_1 X + a_2 Y + a_3 - a_7 X X' - a_8 Y X' &= X' \\ a_4 X + a_5 Y + a_6 - a_7 X Y' - a_8 Y Y' &= Y' \end{aligned} \right\} \quad (4.24)$$

Τα (X, Y) και (X', Y') μας είναι γνωστά. Το (4.24) είναι ένα γραμμικό σύστημα με οκτώ αγνώστους. Χρειάζονται λοιπόν οκτώ εξισώσεις για να βρεθεί η μοναδική λύση του συστήματος, δηλαδή τέσσερα σημεία. Εμείς παρακολουθούμε οκτώ σημεία P_i , των οποίων ξέρουμε τις 2- Δ συντεταγμένες και στην εικόνα $t_1 : P_i(X_i, Y_i)$ και στην εικόνα $t_2 : P'_i(X'_i, Y'_i)$, οπότε προκύπτει ένα σύστημα δεκαέξι εξισώσεων για οκτώ αγνώστους, του οποίου τη μοναδική λύση (a_1, \dots, a_8) την παίρνουμε μέσω Singular Value Decomposition (SVD).

4.4 Υπολογισμός των πραγματικών παραμέτρων κίνησης

Στην προηγούμενη ενότητα (4.3) δείξαμε, χρησιμοποιώντας τη θεωρία Lie (Lie group theory), πως, δεδομένων δύο προοπτικών όψεων στις χρονικές στιγμές t_1 και t_2 , οι οκτώ καθαρές παράμετροι a_1, \dots, a_8 είναι μοναδικές και μπορούν να εκτιμηθούν λύνοντας ένα σύστημα γραμμικών εξισώσεων. Αφού, λοιπόν, προσδιορίσαμε τις καθαρές παραμέτρους θα δείξουμε στην ενότητα αυτή πώς υπολογίζονται οι πραγματικές παράμετροι κίνησης.

Εφόσον μιλάμε για ένα στερεό επίπεδο κομμάτι (rigid planar patch) στον τρισδιάστατο χώρο, θα χρησιμοποιήσουμε τα γεγονότα ότι αυτό είναι αμετάβλητο ως προς το σχήμα και τη μάζα του και ότι αυτό μπορεί να προσανατολιστεί σε δύο πιθανές κατευθύνσεις, έτσι ώστε να κόψει έναν κύκλο σε ένα ελλειψοειδές, για να αποδείξουμε ότι το πλήθος των πιθανών λύσεων για τις παραμέτρους κίνησης ποτέ δεν ξεπερνά τις δύο. Θα βρούμε τις παραμέτρους κίνησης εκτός ενός παράγοντα κλίμακας για τις παραμέτρους μεταφοράς. Το πλήθος των λύσεων εξαρτάται από την πολλαπλότητα των singular values του παρακάτω πίνακα (4.25), ο οποίος αποτελείται από τις οκτώ καθαρές παραμέτρους a_i :

$$A = \begin{bmatrix} a_1 & a_2 & a_3 \\ a_4 & a_5 & a_6 \\ a_7 & a_8 & 1 \end{bmatrix} \quad (4.25)$$

Η SVD (Singular Value Decomposition) του A δίδεται από :

Γεώργιος Σημαντήρης

$$A = U \begin{bmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \lambda_3 \end{bmatrix} V^T = U \Lambda V^T \quad (4.26)$$

όπου τα λ_i είναι τα singular values του A , και οι U, V είναι 3×3 ορθοκανονικοί πίνακες.

Έστω

$$k = n_3^2 + (1 - n_3^2) \cos \theta + c \Delta z \quad (4.27)$$

τότε η (4.25), με βάση τις (4.16), (4.15) και (4.27), γίνεται :

$$A = k^{-1} \left\{ R + \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} \begin{bmatrix} a & b & c \end{bmatrix} \right\} \Leftrightarrow \quad (4.28)$$

$$kA = R + \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} \begin{bmatrix} a & b & c \end{bmatrix} \quad (4.29)$$

Από τις (4.8) και (4.14) βλέπουμε ότι :

$$\begin{aligned} \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} &= R \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} \begin{bmatrix} a & b & c \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} \Leftrightarrow \\ \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} &= \left\{ R + \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} \begin{bmatrix} a & b & c \end{bmatrix} \right\} \begin{bmatrix} x \\ y \\ z \end{bmatrix} \stackrel{(4.29)}{\Leftrightarrow} \\ \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} &= kA \begin{bmatrix} x \\ y \\ z \end{bmatrix} \end{aligned} \quad (4.30)$$

Μετασηματίζοντας το αρχικό σύστημα συντεταγμένων με τον ορθοκανονικό πίνακα V της (4.26), όπως απεικονίζεται στο σχήμα 4.2, όπου το (x_n, y_n, z_n) , είναι το νέο σύστημα συντεταγμένων μετά το μετασηματισμό, έχουμε :

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = V \begin{bmatrix} x_n \\ y_n \\ z_n \end{bmatrix} \quad (4.31)$$

και

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = V \begin{bmatrix} x'_n \\ y'_n \\ z'_n \end{bmatrix} \quad (4.32)$$

Αντικαθιστώντας τις (4.31) και (4.32) στην (4.30), έχουμε :

$$V \begin{bmatrix} x'_n \\ y'_n \\ z'_n \end{bmatrix} = kAV \begin{bmatrix} x_n \\ y_n \\ z_n \end{bmatrix} \quad (4.33)$$

Παίρνοντας τις Ευκλείδειες νόρμες των διανυσμάτων και των δύο μελών της (4.33), λαμβάνουμε :

$$\begin{bmatrix} x'_n & y'_n & z'_n \end{bmatrix} V^T V \begin{bmatrix} x'_n \\ y'_n \\ z'_n \end{bmatrix} = k^2 \begin{bmatrix} x_n & y_n & z_n \end{bmatrix} V^T A^T AV \begin{bmatrix} x_n \\ y_n \\ z_n \end{bmatrix} \quad (4.34)$$

Εφόσον ο πίνακας V είναι ορθοκανονικός, μπορούμε να αντικαταστήσουμε το γινόμενο VV^T στο πρώτο μέλος της (4.34) με τον μοναδιαίο πίνακα I . Αντικαθιστούμε επίσης την (4.26) στην (4.34) και έχουμε :

$$x_n'^2 + y_n'^2 + z_n'^2 = k^2 \begin{bmatrix} x_n & y_n & z_n \end{bmatrix} V^T V \Lambda U^T U \Lambda V^T V \begin{bmatrix} x_n \\ y_n \\ z_n \end{bmatrix} \quad (4.35)$$

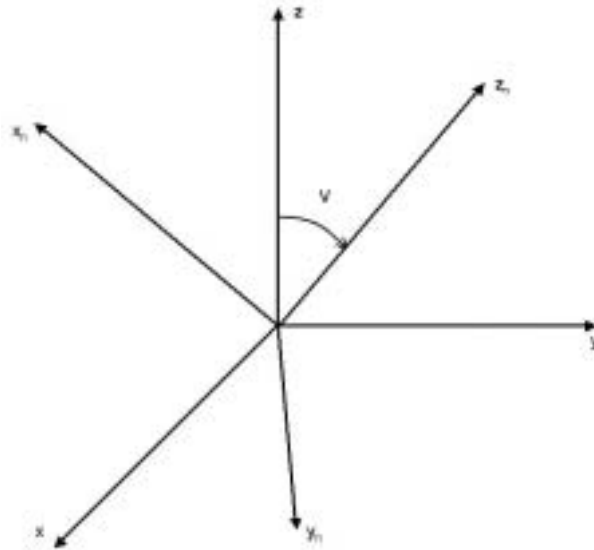
Ο πίνακας U είναι επίσης ορθοκανονικός, οπότε $U^T U = I$, και η (4.35) γίνεται :

$$x_n'^2 + y_n'^2 + z_n'^2 = k^2 (\lambda_1^2 x_n^2 + \lambda_2^2 y_n^2 + \lambda_3^2 z_n^2) \quad (4.36)$$

η οποία αποτελεί την εξίσωση-κλειδί που θα μας οδηγήσει στη λύση του προβλήματος μοναδικότητας, όπως θα δούμε παρακάτω.

Στη συνέχεια θα εκθέσουμε τρία θεωρήματα σχετικά με τη μοναδικότητα και τον υπολογισμό των παραμέτρων κίνησης, δεδομένων των καθαρών παραμέτρων, και των χαρακτηρισμών της κίνησης στον 3-Δ χώρο των αντικειμένων (3-D object space) για τις διάφορες πολλαπλότητες των singular values του πίνακα A . Η αποδείξεις τους είναι σημαντικές για την κατανόηση του προβλήματος και προκύπτουν από τη βασική βιβλιογραφία ([85]). Παραθέτονται στο Παράρτημα (βλ. Κεφάλαιο B).

Γεώργιος Σημαντήρης



Σχήμα 4.2: Το νέο σύστημα συντεταγμένων (x_n, y_n, z_n) , όπως προκύπτει μετασχηματίζοντας το παλιό (x, y, z) με τον ορθοκανονικό πίνακα V .

Θεώρημα 1

Αν η πολλαπλότητα των singular values του πίνακα A είναι δύο, π.χ. $\lambda_1 = \lambda_2 \neq \lambda_3$, τότε η λύση για τις παραμέτρους κίνησης και γεωμετρίας (προσανατολισμός) είναι μοναδική, εκτός ενός κοινού παράγοντα κλίμακας για τις παραμέτρους μεταφοράς και γεωμετρίας, και

$$R = \lambda_1^{-1} A - \left(\frac{\lambda_3}{\lambda_1} - s \right) U_3 V_3^T,$$

$$\begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} = w^{-1} \left(\frac{\lambda_3}{\lambda_1} - s \right) U_3 \text{ και}$$

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = w V_3$$

όπου

$$s = \det(U)\det(V),$$

w ένας παράγοντας κλίμακας,

a, b και c οι παράμετροι της (4.8) (εξίσωση επιπέδου τη χρονική στιγμή t_1),

U_3 και V_3 οι τρίτες στήλες των πινάκων U και V αντίστοιχα.

(4.37)

Επιπλέον, μία αναγκαία και επαρκής συνθήκη, ώστε η πολλαπλότητα των singular values του πίνακα A να είναι δύο, είναι ότι η κίνηση μπορεί να πραγματοποιηθεί περιστρέφοντας το αντικείμενο γύρω από την αρχή των αξόνων και μετά μεταφέροντάς το διαμήκως του κάθετου διανύσματος της επίπεδης επιφάνειας (N).

Θεώρημα 2

Αν τα singular values του A είναι όλα διαφορετικά, π.χ. $\lambda_1 > \lambda_2 > \lambda_3$, τότε υπάρχουν ακριβώς δύο λύσεις για τις παραμέτρους γεωμετρίας και κίνησης, εκτός ενός παράγοντα κλίμακας για τις παραμέτρους μεταφοράς και γεωμετρίας, και

$$R = U \begin{bmatrix} \alpha & 0 & \beta \\ 0 & 1 & 0 \\ -s\beta & 0 & s\alpha \end{bmatrix} V^T$$

$$\begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} = w^{-1} \left[-\beta U_1 + \left(\frac{\lambda_3}{\lambda_2} - s\alpha \right) U_3 \right]$$

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = w [\delta V_1 + V_3]$$

όπου

$$\delta = \pm \left(\frac{\lambda_1^2 - \lambda_2^2}{\lambda_2^2 - \lambda_3^2} \right)^{\frac{1}{2}}$$

$$\alpha = \frac{\lambda_1 + s\lambda_3\delta^2}{\lambda_2(1 + \delta^2)}$$

$$\beta = \pm \sqrt{1 - \alpha^2}$$

$$s = \det(U)\det(V)$$

Σε κάθε μία από τις δύο λύσεις, $\text{sgn}(\beta) = -\text{sgn}(\delta)$.

Επιπλέον, μία αναγκαία και επαρκής συνθήκη, ώστε ο πίνακας A να έχει διαφορετικά singular values, είναι ότι η κίνηση μπορεί να πραγματοποιηθεί περιστρέφοντας το αντικείμενο γύρω από την αρχή των αξόνων και μετά μεταφέροντάς το διαμήκως μίας κατεύθυνσης διαφορετικής του κάθετου διανύσματος της επίπεδης επιφάνειας (N) τη χρονική στιγμή t_2 .

Θεώρημα 3

Η ικανή και αναγκαία συνθήκη, ώστε η πολλαπλότητα των singular values του πίνακα A να είναι τρία, δηλαδή $\lambda_1 = \lambda_2 = \lambda_3$, είναι ότι η κίνηση αποτελείται μόνο από

Γεώργιος Σημαντήρης

μία περιστροφή γύρω από έναν άξονα, που περνάει απ' την αρχή των αξόνων. Δηλαδή το διάνυσμα της μεταφοράς $(\Delta x, \Delta y, \Delta z)$ είναι μηδέν. Ακόμα, ο πίνακας περιστροφής είναι μοναδικός και $R = \lambda_1^{-1} A$. Η επιφάνεια μπορεί να βρίσκεται οπουδήποτε.

4.5 Βασικά συμπεράσματα

Στο Κεφάλαιο αυτό αναφερθήκαμε γενικά στις εξισώσεις κίνησης που διέπουν ένα αντικείμενο στον 3-Δ χώρο. Περιοριστήκαμε σε στερεά αντικείμενα που είναι επίπεδα και προσδιορίσαμε τις οκτώ καθαρές παραμέτρους που χαρακτηρίζουν την κίνησή τους. Από αυτές τις καθαρές παραμέτρους υπολογίσαμε τελικά τις πραγματικές παραμέτρους της κίνησης, καθώς και το κάθετο διάνυσμα στο επίπεδο, που φανερώνει τον προσανατολισμό του επιπέδου, χρησιμοποιώντας τρία βασικά θεωρήματα, οι αποδείξεις των οποίων μπορούν να βρεθούν στο Παράρτημα, Κεφάλαιο Β. Στην ενότητα αυτή θα καταλήξουμε σε κάποια συμπεράσματα και θα περιγράψουμε τον τρόπο με τον οποίο χρησιμοποιήθηκαν τα θεωρήματα στην πράξη.

Πρώτα απ' όλα θα πρέπει να σημειώσουμε ότι το κάθετο διάνυσμα N στο επίπεδο, δηλαδή ο προσανατολισμός του αντικειμένου και κατά συνέπεια του προσώπου που παρακολουθούμε, βρίσκεται μεν, αλλά εξαρτάται από έναν παράγοντα κλίμακας δε, που δεν μπορούμε να ανακτήσουμε. Το ίδιο συμβαίνει και με το διάνυσμα της μεταφοράς T . Στην περίπτωση του N αυτό δεν μας πειράζει, καθότι ο άγνωστος παράγοντας κλίμακας επηρεάζει μόνο το μέτρο του διανύσματος και όχι την κατεύθυνσή του. Όταν μιλάμε για προσανατολισμό μας ενδιαφέρει αποκλειστικά μόνο η κατεύθυνση του διανύσματος. Για λόγους συνέπειας μετατρέπουμε το διάνυσμα N σε μοναδιαίο ($\|N\| = 1$) και έτσι δεν εξαρτόμαστε πλέον από τον άγνωστο παράγοντα, ενώ η κατεύθυνση του διανύσματος παραμένει αναλλοίωτη. Δεν μπορούμε να πούμε όμως το ίδιο και για το διάνυσμα της μεταφοράς T . Μη έχοντας, λοιπόν, τη δυνατότητα να βρούμε ή να απαλείψουμε με κάποιον τρόπο τον άγνωστο παράγοντα κλίμακας, είμαστε αναγκασμένοι να περιοριστούμε στην κατεύθυνση του T καθώς και στις συνιστώσες του, που προκύπτουν από τις εξισώσεις θέτοντας $w = 1$ (όπου w ο παράγοντας κλίμακας). Αυτό δεν επηρεάζει σε καμία περίπτωση την κίνηση του αντικειμένου στις δύο διαστάσεις και τη μέθοδο πρόβλεψης που θα χρησιμοποιήσουμε στο επόμενο Κεφάλαιο, διότι οι συντεταγμένες των σημείων που παρακολουθούνται στην εικόνα προκύπτουν κατευθείαν από τις προηγούμενες με βάση τις οκτώ καθαρές παραμέτρους, όπως φαίνεται και στην (4.10), πράγματα που μας είναι γνωστά. Τέλος, το διάνυσμα N , όπως προκύπτει από τα θεωρήματα αφορά τον προσανατολισμό του επιπέδου τη χρονική στιγμή t_1 . Εμείς ναι μεν θέλουμε το διάνυσμα αυτό, αλλά τη χρονική στιγμή t_2 , κάτι που λύνεται εύκολα περιστρέφοντας απλά το N με τον πίνακα περιστροφής R που βρήκαμε.

Στο θεώρημα 1 (βλ. και Β.1) είδαμε ότι οι παράμετροι κίνησης και γεωμετρίας είναι μοναδικές (πέραν του παράγοντα κλίμακας) αν τα singular values του πίνακα A , που περιέχει τις οκτώ καθαρές παραμέτρους, έχουν πολλαπλότητα δύο. Αυτό σημαίνει δύο

από αυτά είναι ίσα και το τρίτο διαφορετικό. Είναι η μόνη περίπτωση που μπορούμε να ανακτήσουμε τα διανύσματα T (μεταφορά), N (προσανατολισμός) και τον πίνακα R (περιστροφή) χωρίς προβλήματα. Δεν συμβαίνει όμως το ίδιο και με τα δύο άλλα θεωρήματα.

Στο τρίτο θεώρημα (βλ. και B.3), όπου τα singular values του πίνακα A είναι όλα ίσα, το πρόβλημα έγκειται στο διάνυσμα N . Ο προσανατολισμός του επιπέδου παραμένει θεωρητικά άγνωστος, καθώς, με βάση την απόδειξη, το επίπεδο μπορεί να βρίσκεται οπουδήποτε στο χώρο. Η δυσκολία αυτή λύνεται εν μέρει πρακτικά. Εφόσον είμαστε σε θέση να ανακτήσουμε τον πίνακα περιστροφής R , η μοναδικότητα του οποίου έχει αποδειχθεί, και αφού ξέρουμε τον προηγούμενο προσανατολισμό, δηλαδή το προηγούμενο N , αρκεί να περιστρέψουμε το N με τον πίνακα R . Το προηγούμενο N το ξέρουμε πάντοτε από τις προηγούμενες αναλύσεις της κίνησης στις προηγούμενες εικόνες. Όταν όμως το πρόσωπο που παρακολουθούμε ξεκινήσει μόνο με περιστροφή, δηλαδή δεν υπάρχει καθόλου μεταφορά στις δύο πρώτες εικόνες της ακολουθίας, τότε δεν μπορούμε να ανακτήσουμε τον προσανατολισμό του προσώπου. Το πρόβλημα αυτό είναι σοβαρό. Για το λόγο αυτό, κατά την αρχικοποίηση, πρέπει αναγκαστικά να προσδιορίζεται και ο προσανατολισμός του προσώπου στην πρώτη εικόνα της ακολουθίας.

Πρόβλημα, όμως, συναντάμε και στην περίπτωση που τα singular values του πίνακα A είναι όλα διαφορετικά μεταξύ τους (θεώρημα 2, βλ. και B.2). Εκεί οι λύσεις για τις παραμέτρους κίνησης και γεωμετρίας είναι δύο και θα πρέπει να αποφανθούμε ποια από τις δύο είναι η σωστή. Για να πάρουμε αυτήν την απόφαση στηριζόμαστε στην υπόθεση ότι η κίνηση που κάνει το πρόσωπο από εικόνα σε εικόνα γενικά δεν είναι μεγάλη κάτι που έχει ιδιαίτερη σημασία για την περιστροφή. Ο πίνακας R , όπως ορίστηκε στην (4.15), ισχύει για μικρή περιστροφή, δηλαδή η γωνία της περιστροφής θ είναι αρκετά μικρή. Αυτό σημαίνει ότι η διαγώνιος του θα αποτελείται από στοιχεία περίπου ίσα με 1. Λόγω των δύο πιθανών λύσεων, έχουμε και δύο πίνακες περιστροφής. Ο κάθε πίνακας ορίζει την περιστροφή που έχει εκτελέσει το επίπεδο στον 3-Δ χώρο, δηλαδή το πρόσωπο, όπως φαίνεται από τη μία εικόνα στην επόμενη. Όπως τονίζουμε και στην απόδειξη (βλ. Παράρτημα B.2) υπάρχουν δύο τρόποι να τοποθετήσουμε το επίπεδο, έτσι ώστε η τομή του με το ελλειψοειδές να αποτελεί έναν κύκλο. Στους δύο αυτούς τρόπους τοποθέτησης φτάνουμε, πέραν της μεταφοράς, και μέσω των δύο πινάκων περιστροφής που αναφέραμε. Ο ένας μόνο θα ορίζει την περιστροφή που μας ενδιαφέρει, η οποία είπαμε ότι είναι αρκετά μικρή. Εξετάζοντας, λοιπόν, τη διαγώνιο του καθενός μπορούμε να αποφανθούμε και για τη σωστή λύση. Η περιστροφή όμως μπορεί να είναι το ίδιο μικρή και στις δύο λύσεις, θα έχει όμως σίγουρα διαφορετική κατεύθυνση. Συγκρίνοντας, λοιπόν, τους πίνακες R δεν μπορούμε να καταλήξουμε σε απόφαση. Στην περίπτωση αυτή αποφαινόμαστε με βάση την τρίτη συνιστώσα του διανύσματος N , που βρίσκουμε σε κάθε λύση. Η συνιστώσα c του $N = (a, b, c)$ θα πρέπει να έχει αρνητικό πρόσημο με βάση τη γεωμετρία του σχήματος 4.1, άρα διαλέγουμε εκείνη τη λύση, για την οποία $c < 0$. Συνδυάζοντας τις δύο παραπάνω παρατηρήσεις είμαστε σε θέση να επιλέξουμε

τελικά τη μοναδική λύση, που θα μας δώσει τις παραμέτρους κίνησης και γεωμετρίας.

Κεφάλαιο 5

Παρακολούθηση σημείων

Στο κεφάλαιο αυτό θα αναφερθούμε στην παρακολούθηση των σημείων από εικόνα σε εικόνα. Είδαμε παραπάνω (Κεφ. 3) ότι παρακολουθούνται οκτώ (8) σημεία που βρίσκονται στις άκρες χαρακτηριστικών του προσώπου (μάτια, μύτη και στόμα). Τα χαρακτηριστικά αυτά ουσιαστικά ορίζουν το πρόσωπο, οπότε παρακολουθώντας τα σημεία παρακολουθούμε και τα χαρακτηριστικά και κατά συνέπεια το πρόσωπο. Ένα σημαντικό πλεονέκτημα της παρακολούθησης προσώπων με βάση την παρακολούθηση χαρακτηριστικών είναι ότι μπορούμε να προχωρήσουμε άμεσα σε άλλες εφαρμογές που αφορούν το πρόσωπο, όπως είναι η ανάλυση-αναγνώριση της έκφρασης, η παρακολούθηση του βλέμματος και η αναγνώριση της ομιλίας. Επίσης είπαμε ότι η παρακολούθηση γίνεται με βάση μία πρόβλεψη και μία διόρθωση, που θα εξηγήσουμε στο Κεφάλαιο αυτό.

5.1 Σημασία των σημείων

Στην εικόνα 3.1 του Κεφαλαίου 3 αναφερθήκαμε σύντομα στα σημεία που επιλέχθηκαν για το σκοπό της παρακολούθησης. Επισημάναμε το γεγονός, ότι, έτσι όπως τοποθετήθηκαν τα σημεία, μπορούμε να υποθέσουμε ότι βρίσκονται πάνω σε ένα επίπεδο στον 3-Δ χώρο με αποτέλεσμα να μπορέσουμε στο Κεφάλαιο 4 να ανακτήσουμε τις παραμέτρους κίνησης και γεωμετρίας.

Ο υπολογιστής, σαν μηχανήμα που είναι, δεν έχει τη δυνατότητα αντίληψης, όπως την έχουμε εμείς οι άνθρωποι. Είναι εύκολο για κάποιον άνθρωπο, βλέποντας ένα πρόσωπο, να ξεχωρίσει τα μάτια, τη μύτη και το στόμα και να παρακολουθήσει πάνω στα χαρακτηριστικά αυτά κάποια σημεία, ιδιαίτερα σαν κι αυτά της εικόνας 3.1, που βρίσκονται στα άκρα των προαναφερθέντων χαρακτηριστικών και δύσκολα τα μπερδεύει με άλλα. Για να παρακολουθήσει όμως ο υπολογιστής τα σημεία αυτά, πρέπει να είναι ικανός να εξάγει κάποιες πληροφορίες από τη θέση και το γύρω περιβάλλον του κάθε σημείου. Η αναζήτηση των πληροφοριών αυτών από εικόνα σε εικόνα του επιτρέπει να παρακολουθεί τα σημεία. Χρησιμοποιώντας τα συγκεκριμένα σημεία ο υπολογιστής έχει τη δυνατότητα αυτή, με βάση την αντιστοίχιση μπλοκ (block matching), όπως θα

περιγράψουμε παρακάτω. Αυτό είναι δυνατό διότι τα σημεία υποδηλώνουν κάποια πρότυπα (patterns).

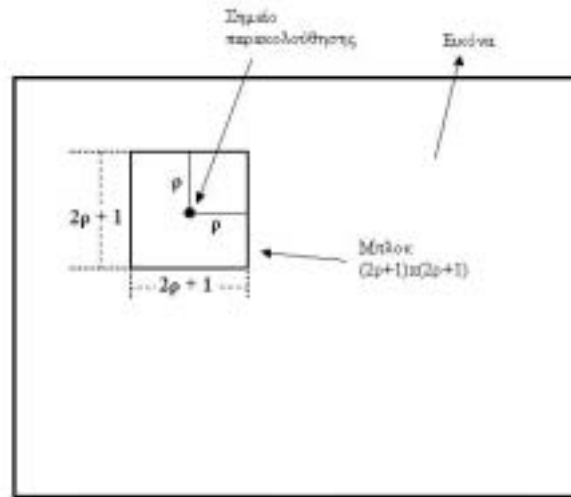
Μεγάλο ρόλο όμως στην παρακολούθηση παίζουν και κάποιοι συνδυασμοί των σημείων. Παίρνοντας τα τέσσερα σημεία που αντιστοιχούν στα μάτια και ενώνοντάς τα σχηματίζουμε ένα ευθύγραμμο τμήμα. Όμοια, ένα άλλο ευθύγραμμο τμήμα προκύπτει από την ένωση των δύο σημείων της μύτης, και ένα τρίτο ενώνοντας τα δύο σημεία του στόματος. Αν προσέξουμε γενικά τα πρόσωπα θα δούμε ότι διέπονται από κάποια γεωμετρία. Στην εικόνα 5.1 παρατηρούμε ότι τα ευθύγραμμα τμήματα που ορίσαμε προηγουμένως είναι παράλληλα μεταξύ τους, άσχετα με τη θέση και τον προσανατολισμό του προσώπου. Βέβαια, εφόσον μιλάμε για ψηφιακές εικόνες και άρα για το διακριτό χώρο, η παραλληλότητα ορίζεται με ακρίβεια εικονοστοιχείων. Τα ευθύγραμμα τμήματα σχεδιάζονται χρησιμοποιώντας τον αλγόριθμο *Midpoint* και η παραλληλότητά τους μπορεί να “διαταρράσεται” ελάχιστα πάνω στην εικόνα. Άλλο ένα κίνητρο, λοιπόν, να επιλέξουμε τα συγκεκριμένα σημεία είναι ότι λόγω της προοπτικής προβολής, της μορφολογίας του προσώπου καθώς και της θέσης του σχετικά με την κάμερα διατηρείται η παραλληλότητα των ευθειών και αλλάζουν απλώς τα μήκη των ευθυγράμμων τμημάτων. Θα δούμε παρακάτω πως αυτό, αλλά και οι λόγοι των ευθυγράμμων τμημάτων, μας βοηθάνε να ορίσουμε αρκετά απλούς γεωμετρικούς περιορισμούς για να εξασφαλίσουμε την επιτυχία της παρακολούθησης. Αυτό προϋποθέτει, βέβαια, ότι και κατά την αρχικοποίηση τα σημεία θα τοποθετούνται κατάλληλα, ώστε να σχηματίζουν παράλληλες ευθείες, με την έννοια της παραλληλότητας που αναφέραμε παραπάνω.



Σχήμα 5.1: Στις παραπάνω εικόνες βλέπουμε πως διατηρείται η παραλληλότητα των ευθυγράμμων τμημάτων που ορίστηκαν κατά την προοπτική προβολή, ανεξάρτητα από τον προσανατολισμό και τη μεταφορά του προσώπου.

5.2 Αντιστοίχιση μπλοκ (block matching)

Το μπλοκ (block) ορίζεται με κέντρο ένα σημείο, από αυτά που παρακολουθούμε, και ένα τετράγωνο παραθύρο, η κάθε πλευρά του οποίου απέχει απόσταση ρ από το σημείο αυτό. Η πλευρά του μπλοκ είναι δηλαδή $2\rho + 1$, όπως φαίνεται και στο σχήμα 5.2. Το μπλοκ αποτελείται από όλα τα εικονοστοιχεία εντός του παραθύρου. Πειραματισμοί με διάφορα μεγέθη μπλοκ έδειξαν ότι καταλληλότερο μπλοκ είναι αυτό με $\rho = 4$, δηλαδή πλευρά 9. Έχοντας, λοιπόν, ορίσει το μπλοκ 9×9 που χρησιμοποιούμε, θα εξηγήσουμε τι σημαίνει στην περίπτωση μας αντιστοίχιση μπλοκ.



Σχήμα 5.2: Ο ορισμός του μπλοκ για συγκεκριμένο σημείο. Αποτελείται από όλα τα εικονοστοιχεία που βρίσκονται εντός τετραγωνικού παραθύρου πλευράς $2\rho + 1$ με κέντρο το σημείο παρακολούθησης.

Θεωρούμε δύο μπλοκ T και M . Το πρώτο αποτελεί το πρότυπο (Template) ενώ το δεύτερο είναι αυτό που προσπαθούμε να ταιριάξουμε (Match). Η αντιστοίχιση των δύο αυτών μπλοκ γίνεται με βάση μία μορφή συσχέτισης (correlation). Ο τύπος της συσχέτισης δίνεται από την :

$$cor = \frac{covTM}{\sqrt{covTT \cdot covMM}} \quad (5.1)$$

Τα $covTM$, $covTT$ και $covMM$ ορίζουν συνδιασπορές (covariances) που αφορούν αντίστοιχα και τα δύο μπλοκ, το μπλοκ πρότυπο και το μπλοκ του ταιριάσματος :

$$covTM = \frac{1}{N-1} \sum_{i=0}^{N-1} ((T(i) - \bar{T})(M(i) - \bar{M})) \quad (5.2)$$

$$covTT = \frac{1}{N-1} \sum_{i=0}^{N-1} ((T(i) - \bar{T})^2) \quad (5.3)$$

$$covMM = \frac{1}{N-1} \sum_{i=0}^{N-1} \left((M(i) - \overline{M})^2 \right) \quad (5.4)$$

όπου \overline{T} και \overline{M} οι μέσες τιμές των δύο μπλοκ αντίστοιχα :

$$\overline{T} = \frac{1}{N} \sum_{i=0}^{N-1} T(i) \quad (5.5)$$

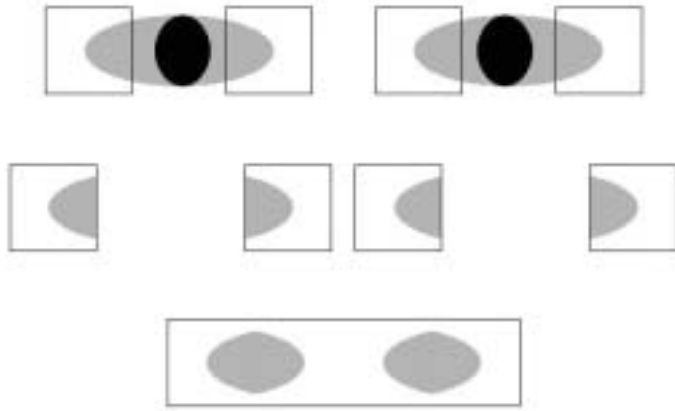
$$\overline{M} = \frac{1}{N} \sum_{i=0}^{N-1} M(i) \quad (5.6)$$

και N είναι το πλήθος των εικονοστοιχείων του κάθε μπλοκ. Στην περίπτωση μας, για μπλοκ 9×9 , $N = 81$. Η συσχέτιση αυτή (5.1), όπως μπορούμε να διαπιστώσουμε κι από τους τύπους των συντελεστών που την αποτελούν (5.2), (5.3), (5.4), είναι κανονικοποιημένη και δίνει τιμές από -1 μέχρι $+1$. Η φυσική σημασία των τιμών αυτών είναι ότι τα δύο μπλοκ είναι εντελώς αντίθετα (-1), δεν έχουν καμία απολύτως σχέση (0), ταυτίζονται ($+1$).

Όλα τα παραπάνω ισχύουν σε περίπτωση που έχουμε ένα μπλοκ ως πρότυπο και ένα για να ταιριάζουμε. Πειράματα έδειξαν ότι η αντιστοίχιση χρησιμοποιώντας ένα ένα τα μπλοκ δεν έχει καλά αποτελέσματα. Γίνεται φανερό ότι ένα μπλοκ από μόνο του είναι πολύ αδύναμο με αποτέλεσμα να συμβαίνουν λάθη, που δύσκολα διορθώνονται. Επιλέξαμε να συνδυάσουμε τα μπλοκ των σημείων που αφορούν το ίδιο χαρακτηριστικό για να αποκτήσουν περισσότερη ισχύ, πράγμα που επιβεβαιώθηκε στην πράξη. Η ένωση, λοιπόν, των μπλοκ που προκύπτουν από τα τέσσερα σημεία των ματιών δίνει έναν τέτοιο συνδυασμό. Όμοια, ένας δεύτερος προκύπτει ενώνοντας τα μπλοκ των δύο σημείων της μύτης και ένας τρίτος ενώνοντας τα μπλοκ των δύο σημείων του στόματος. Η φυσική σημασία των συνδυασμών αυτών φαίνεται στο σχήμα 5.3, όπου βλέπουμε σε ένα συνθετικό παράδειγμα ματιών τι γίνεται κατά την ομαδοποίηση των μπλοκ. Τα μπλοκ-πρότυπα που σχηματίζονται για τα μάτια από μόνα τους είναι αδύναμα, αλλά η ομαδοποίηση έχει σαν αποτέλεσμα ένα πιο δυνατό πρότυπο. Το ίδιο ακριβώς συμβαίνει και για τα μπλοκ της μύτης και του στόματος.

Χρησιμοποιώντας συνδυασμούς από μπλοκ, είναι φυσικό να μεταβληθούν και οι τύποι που ορίζουν τη συσχέτιση. Όταν τα μπλοκ που θέλουμε να χρησιμοποιήσουμε είναι περισσότερα, ουσιαστικά θεωρούμε σαν το πρότυπο την ένωση όλων των μπλοκ. Αυτό μεταφράζεται στην προσπάθεια να αντιστοιχίσουμε το πρότυπο, που σχηματίζεται ενώνοντας τα μπλοκ, με ένα ανάλογο ταίριασμα φυσικά. Η σχέση (5.1) μένει ως έχει, αλλά οι υπόλοιπες μετατρέπονται σε :

$$covTM = \frac{1}{k(N-1)} \sum_{i=0}^{k(N-1)} \left((T'(i) - \overline{T}')(M'(i) - \overline{M}') \right) \quad (5.7)$$



Σχήμα 5.3: Διακρίνουμε πώς από τα μάτια (πάνω) προκύπτουν τα 4 μπλοκ που αντιστοιχούν στα 4 σημεία που παρακολουθούνται (κέντρο) και στη συνέχεια η ομαδοποίηση αυτών για να σχηματίσουν το νέο πρότυπο (κάτω) που είναι πιο ισχυρό.

$$cov_{TT} = \frac{1}{k(N-1)} \sum_{i=0}^{k(N-1)} \left((T'(i) - \bar{T}')^2 \right) \quad (5.8)$$

$$cov_{MM} = \frac{1}{k(N-1)} \sum_{i=0}^{k(N-1)} \left((M'(i) - \bar{M}')^2 \right) \quad (5.9)$$

όπου

$$\bar{T}' = \frac{1}{kN} \sum_{i=0}^{k(N-1)} T'(i) \quad (5.10)$$

$$\bar{M}' = \frac{1}{kN} \sum_{i=0}^{k(N-1)} M'(i) \quad (5.11)$$

με T' και M' να είναι τα νέα πρότυπα που προκύπτουν από την ένωση των μπλοκ-προτύπων και την ένωση των μπλοκ-ταιριασμάτων αντίστοιχα, \bar{T}' και \bar{M}' οι νέες μέσες τιμές, και k είναι ο αριθμός των μπλοκ που χρησιμοποιούμε.

5.3 Πρόβλεψη

Η αντιστοίχιση προσπαθεί να εντοπίσει τα σημεία που παρακολουθούνται από τη μία εικόνα στην άλλη, χρησιμοποιώντας τη συσχέτιση και το συνδυασμό μπλοκ, όπως ορίστηκαν παραπάνω. Γίνεται κατανοητό ότι το ψάξιμο σε όλη την εικόνα είναι υπολογιστικά ακριβό και καθυστερεί πολύ την εκτέλεση του αλγορίθμου. Για το λόγο αυτό προσπαθούμε να περιορίσουμε την αναζήτηση στην επόμενη εικόνα, ώστε

να κερδίσουμε και σε κόστος υπολογισμού, αλλά και σε χρόνο εκτέλεσης. Αυτό το πετυχαίνουμε με την πρόβλεψη. Γνωρίζοντας, λοιπόν, τις θέσεις των σημείων στην προηγούμενη εικόνα $I^{prev} = I_{i-1}$ και ξέροντας την κίνηση που αυτά εκτελέσανε από την I_{i-2} στην I_{i-1} (όπως περιγράψαμε στο Κεφάλαιο 4) υποθέτουμε ότι αυτά συνεχίζουν να κινούνται με τον ίδιο τρόπο, οπότε μία απλή εφαρμογή των τύπων (4.10) θα μας δώσει μία εκτίμηση των θέσεων των σημείων στην εικόνα προς επεξεργασία I_i . Φυσικά, αυτό δεν σημαίνει ότι τα σημεία όντως κινήθηκαν έτσι, αλλά μας δίνει τη δυνατότητα να ορίσουμε μία περιοχή όπου θα αναζητήσουμε τα σημεία στην επόμενη εικόνα, ώστε να επιταχύνουμε την αναζήτηση. Η κίνηση που εκτελεί ένα πρόσωπο είναι ομαλή, οπότε δεν αναμένεται να προκύψουν σημαντικά προβλήματα κατά την εκτίμηση. Όταν αλλάζει όμως η κίνηση θα πέφτουμε σε λάθος θέσεις. Για το λόγο αυτό είμαστε αναγκασμένοι να κρατάμε μία ικανή περιοχή αναζήτησης, παρόλο που με την πρόβλεψη πέφτουμε συνήθως πολύ κοντά στις νέες θέσεις των σημείων. Εντός της περιοχής αναζήτησης θα προσπαθήσουμε στη συνέχεια να διορθώσουμε τυχόν λάθη, δηλαδή να βρούμε τις σωστές νέες θέσεις των σημείων χρησιμοποιώντας την αντιστοίχιση και τους γεωμετρικούς περιορισμούς.

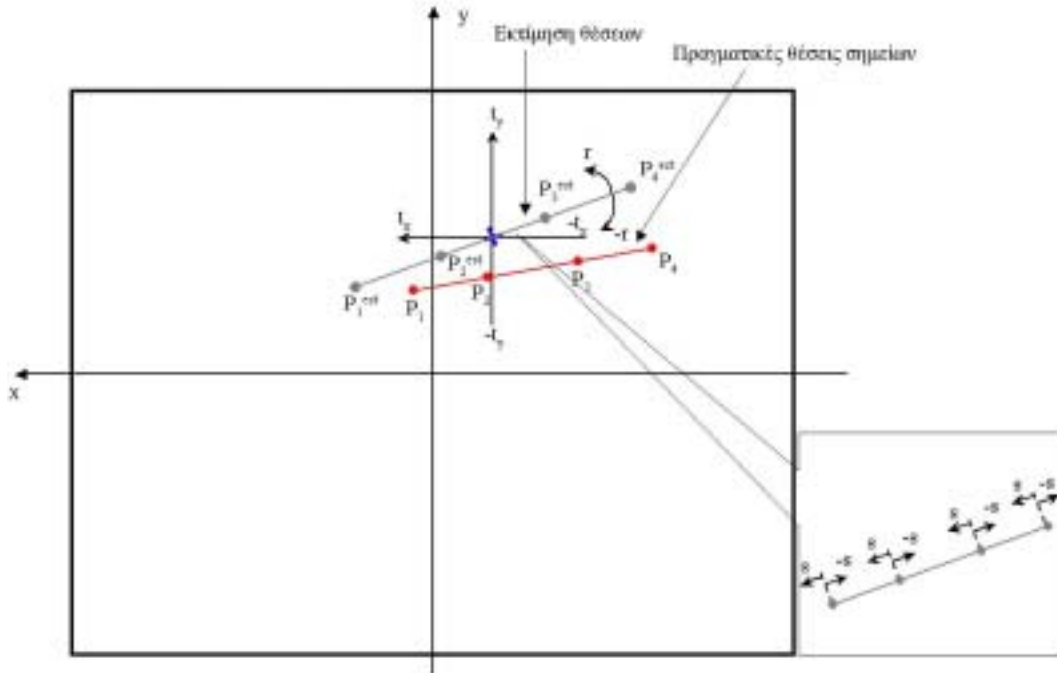
5.4 Διόρθωση

Στην ενότητα αυτή θα δούμε πως διαμορφώνεται η περιοχή αναζήτησης και πραγματοποιείται η αντιστοίχιση. Βασικό ρόλο παίζουν τα ευθύγραμμα τμήματα που περιγράψαμε παραπάνω. Ένα ευθύγραμμο τμήμα από τη μία εικόνα στην άλλη μπορεί να υποστεί σε μεταφορά $t(x, y)$ (στους δύο άξονες x και y) και περιστροφή r γύρω από το κέντρο του. Τα σημεία που βρίσκονται πάνω στο ευθύγραμμο τμήμα μπορεί να έχουν μεταφερθεί σε άλλες θέσεις, πάνω στο ευθύγραμμο τμήμα πάντα, προκαλώντας ουσιαστικά μία κλιμάκωση s . Πρώτα θα βρεθούν τα σημεία που αφορούν τα μάτια. Αυτά θα μας βοηθήσουν στη συνέχεια να βρούμε τα σημεία της μύτης και, τέλος, του στόματος.

5.4.1 Αναζήτηση ματιών

Έχοντας εκτιμήσει τις νέες θέσεις των τεσσάρων σημείων που αφορούν τα μάτια και, εφόσον μιλάμε για κίνηση στερεών αντικειμένων, θα σχηματιστεί και στην επόμενη εικόνα ένα ευθύγραμμο τμήμα e' . Η περιοχή αναζήτησης προκύπτει με βάση τέσσερις παραμέτρους, όπως αναφέραμε και παραπάνω. Το ευθύγραμμο τμήμα μεταφέρεται από $-t_x$ ως $+t_x$ στον οριζόντιο άξονα της εικόνας και από $-t_y$ ως $+t_y$ στον κάθετο άξονα. Σε κάθε νέα θέση περιστρέφεται από $-r$ ως $+r$ γύρω από το κέντρο του. Τα σημεία μπορεί να έχουν μετακινηθεί πάνω στο ευθύγραμμο τμήμα λόγω της 3-Δ κίνησης του προσώπου και της προοπτικής προβολής που μας δίνει τη 2-Δ εικόνα. Αυτό προσομοιώνεται με έναν παράγοντα κλίμακας s στον οποίο υπόκεινται όλα τα σημεία του ευθύγραμμου

τμήματος, και ουσιαστικά πρόκειται για μία μεταφορά από $-s$ μέχρι $+s$ των σημείων πάνω στο ευθύγραμμο τμήμα. Στο σχήμα 5.4 απεικονίζεται πώς από την εκτίμηση φτάνουμε στην περιοχή αναζήτησης.



Σχήμα 5.4: Το ευθύγραμμο τμήμα που δημιουργείται από τις εκτιμήσεις των θέσεων P_i^{est} αποτελεί τη βάση για να προκύψει η περιοχή αναζήτησης. Οι παράμετροι μεταφοράς (t_x , t_y), περιστροφής (r) καθώς και ο παράγοντας κλίμακας (s), που αφορά τη μετακίνηση των σημείων πάνω στο ευθύγραμμο τμήμα, συνεισφέρουν στην εύρεση, εντός της περιοχής αναζήτησης, των σωστών συντεταγμένων των σημείων που παρακολουθούμε P_i .

Ο αλγόριθμος για την αναζήτηση των ματιών απεικονίζεται στον πίνακα 5.1. Η αντιστοίχιση εφαρμόζεται για $k = 4$ μπλοκ, εφόσον πρόκειται για τέσσερα σημεία, σύμφωνα με τις σχέσεις (5.1), (5.7), (5.8), (5.9). Παρατηρούμε ότι τελικά η περιοχή αναζήτησης δεν είναι όλα τα ευθύγραμμα τμήματα ϵ'_j , που προκύπτουν από τη μεταφορά και την περιστροφή του ευθύγραμμου τμήματος ϵ' , αλλά τα σημεία SP_i , τα οποία προέρχονται από την εφαρμογή της κλιμάκωσης s στα σημεία RP_i και βρίσκονται φυσικά πάνω στα ευθύγραμμα τμήματα ϵ'_j . Αυτό έχει σαν αποτέλεσμα μία αρκετά μικρή αλλά και ικανή περιοχή αναζήτησης (βλ. και σχήμα 5.6).

1	Εκτίμηση των συντεταγμένων των τεσσάρων σημείων των ματιών P_i^{est} , $i = 1, \dots, 4$
2	Εύρεση κέντρου C ευθύγραμμου τμήματος $P_1^{est} P_4^{est}$ (για να γίνει η μεταφορά με βάση το κέντρο)
3	Μέγιστη συσχέτιση $maxCor = -2$
4	\forall μεταφορά t_x στον x άξονα {
5	$\vdots \forall$ μεταφορά t_y στον y άξονα {
6	$\vdots \vdots$ Τα νέα σημεία που προκύπτουν είναι TP_i , $i = 1, \dots, 4$
7	$\vdots \vdots$ Εύρεση κέντρου TC ευθύγραμμου τμήματος $TP_1 TP_4$
8	$\vdots \vdots$ \forall περιστροφή r γύρω από το κέντρο TC {
9	$\vdots \vdots \vdots$ Τα νέα σημεία που προκύπτουν είναι RP_i , $i = 1, \dots, 4$
10	$\vdots \vdots \vdots$ \forall μεταφορά s του σημείου RP_1 στην ευθεία $RP_1 RP_4$ {
11	$\vdots \vdots \vdots \vdots$ \forall μεταφορά s του σημείου RP_2 στην ευθεία $RP_1 RP_4$ {
12	$\vdots \vdots \vdots \vdots \vdots$ \forall μεταφορά s του σημείου RP_3 στην ευθεία $RP_1 RP_4$ {
13	$\vdots \vdots \vdots \vdots \vdots \vdots$ \forall μεταφορά s του σημείου RP_4 στην ευθεία $RP_1 RP_4$ {
14	$\vdots \vdots \vdots \vdots \vdots \vdots \vdots$ Τα νέα σημεία που προκύπτουν είναι SP_i , $i = 1, \dots, 4$
15	$\vdots \vdots \vdots \vdots \vdots \vdots \vdots \vdots$ Υπολογισμός συσχέτισης cor πολλαπλών μπλοκ με κέντρα $\vdots \vdots \vdots \vdots \vdots \vdots \vdots \vdots$ τα σημεία SP_i , $i = 1, \dots, 4$
16	$\vdots \vdots \vdots \vdots \vdots \vdots \vdots \vdots$ $An\ cor \geq maxCor$ {
17	$\vdots \vdots \vdots \vdots \vdots \vdots \vdots \vdots$ \vdots Αν ικανοποιούνται οι γεωμετρικοί περιορισμοί {
18	$\vdots \vdots \vdots \vdots \vdots \vdots \vdots \vdots$ $\vdots \vdots$ $maxCor = cor$
19	$\vdots \vdots \vdots \vdots \vdots \vdots \vdots \vdots$ $\vdots \vdots$ Κράτα τα σημεία SP_i , $i = 1, \dots, 4$ ως σωστά
	$\vdots \vdots \vdots \vdots \vdots \vdots \vdots \vdots$ }
	$\vdots \vdots \vdots \vdots \vdots \vdots \vdots \vdots$ }
	$\vdots \vdots \vdots \vdots \vdots \vdots \vdots$ }
	$\vdots \vdots \vdots \vdots \vdots$ }
	$\vdots \vdots \vdots \vdots$ }
	$\vdots \vdots \vdots$ }
	$\vdots \vdots$ }
	\vdots }
	}

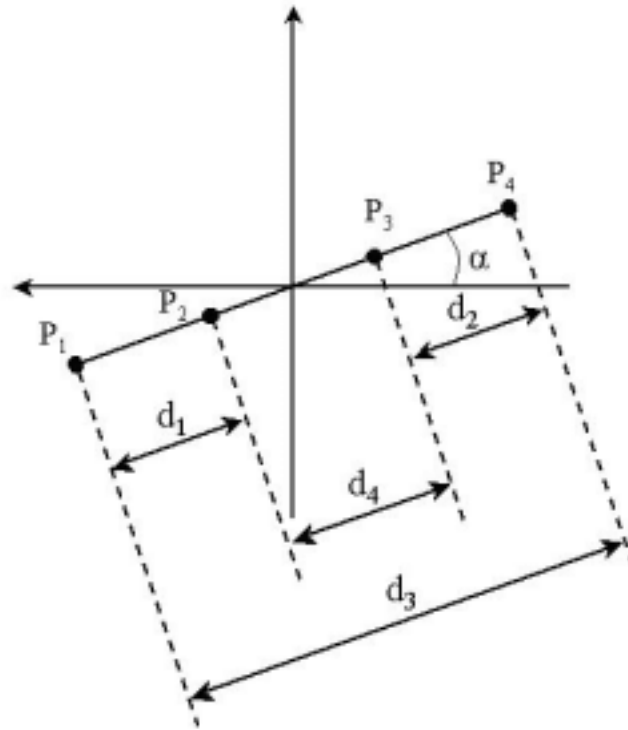
Πίνακας 5.1: Ο αλγόριθμος αναζήτησης ματιών. Στα βήματα 4 και 5 γίνεται η μεταφορά του ευθ. τμήματος. Στο βήμα 8 το νέο ευθ. τμήμα περιστρέφεται γύρω από το κέντρο του. Τα βήματα 10 – 13 προσομοιώνουν την κλιμάκωση. Στο βήμα 15 υπολογίζεται η συσχέτιση για $k = 4$ μπλοκ. Τέλος, στο βήμα 17 ελέγχονται οι γεωμετρικοί περιορισμοί.

Γεώργιος Σημαντήρης

Μένει να διευκρινίσουμε τους γεωμετρικούς περιορισμούς, οι οποίοι προέκυψαν από ανθρωπομορφικούς παράγοντες και πειραματικές επιβεβαιώσεις. Στο σχήμα 5.5 απεικονίζονται τα τέσσερα σημεία που αντιστοιχούν στα μάτια. Ορίζουμε τις εξής αποστάσεις στο δισδιάστατο χώρο :

- $d_1 = d(P_1, P_2)$: Η ευκλείδεια απόσταση των δύο σημείων του αριστερού ματιού (όπως βλέπουμε την εικόνα) P_1 και P_2 .
- $d_2 = d(P_3, P_4)$: Η ευκλείδεια απόσταση των δύο σημείων του δεξιού ματιού (όπως βλέπουμε την εικόνα) P_3 και P_4 .
- $d_3 = d(P_1, P_4)$: Η ευκλείδεια απόσταση των δύο εξωτερικών σημείων των ματιών P_1 και P_4 .
- $d_4 = d(P_2, P_3)$: Η ευκλείδεια απόσταση των δύο εσωτερικών σημείων των ματιών P_2 και P_3 .

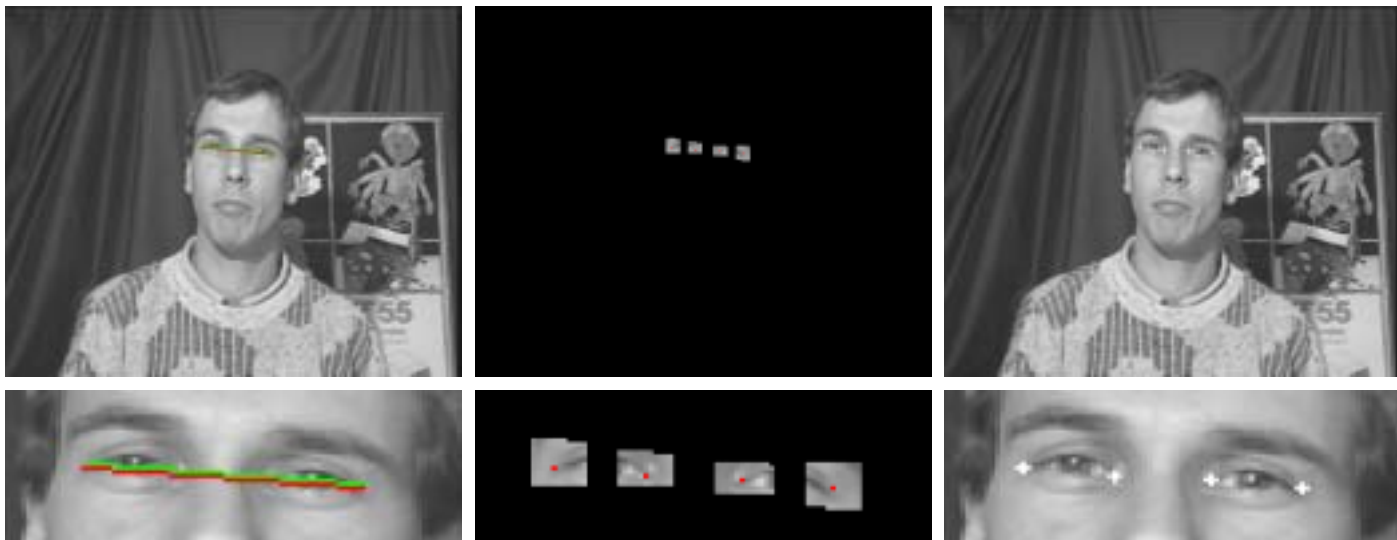
Επίσης ορίζουμε την κλίση α του ευθύγραμμου τμήματος, που προκύπτει από τα σημεία των ματιών, σε σχέση με τον οριζόντιο άξονα. Υπολογίζουμε τα παραπάνω μεγέθη για τα σημεία SP_i , που βρίσκονται υπό εξέταση, καθώς και για τα σημεία P_i^{prev} , τα οποία είναι τα σωστά σημεία που βρέθηκαν στην προηγούμενη εικόνα I^{prev} . Θεωρούμε ότι η κλίση α του ευθύγραμμου τμήματος των ματιών δεν πρέπει να μεταβάλλεται περισσότερο από $7.7^\circ (\approx \tan^{-1}(0.135))$ από εικόνα σε εικόνα και δεν πρέπει να ξεπερνάει τις $24.2^\circ (\approx \tan^{-1}(0.45))$ συνολικά στην εικόνα. Στηριζόμαστε στην υπόθεση ότι το πρόσωπο εκτελεί σχετικά μικρές περιστροφές και συνήθως κοιτάει το φακό. Αν το πρόσωπο κινηθεί προς το φακό ή απομακρυνθεί από αυτόν θα αλλάξουν μονάχα τα μήκη των ευθυγράμμων τμημάτων, αλλά οι λόγοι θα παραμείνουν περίπου σταθεροί, οπότε θεωρούμε ότι ο λόγος $\frac{d_3}{d_4}$ δεν πρέπει να μεταβάλλεται παραπάνω από 0.5 από εικόνα σε εικόνα. Λόγω τις κλιμάκωσης που μπορεί να υποστεί το πρόσωπο οι αποστάσεις θα μεγαλώσουν ή θα μικρύνουν, αλλά και πάλι η κίνηση είναι μικρή και ομαλή, οπότε δεν αναμένεται οι αποστάσεις d_1 και d_2 να μεταβληθούν παραπάνω από 1.0 από εικόνα σε εικόνα. Κατά κανόνα, τα δύο μάτια είναι περίπου ίσα μεταξύ τους αλλά και ίσα με το τμήμα ανάμεσά τους. Επίσης, τα τμήματα αυτά φαίνεται να αποτελούν περίπου το 40% της ολικής απόστασης μεταξύ των ματιών. Μία συνοπτική περιγραφή των κανόνων αυτών φαίνεται στον πίνακα 5.2. Εφαρμόζοντας τον αλγόριθμο και τους γεωμετρικούς περιορισμούς φτάνουμε στο τελικό αποτέλεσμα, όπως μπορούμε να δούμε στο σχήμα 5.6.



Σχήμα 5.5: Τα τέσσερα σημεία των ματιών (P_i) και τα γεωμετρικά μεγέθη που ορίζονται με βάση αυτά : οι αποστάσεις d_i και η κλίση α .

$ \alpha - \alpha^{prev} \leq 0.135$
$ \alpha \leq 0.45$
$\left \frac{d_3}{d_4} - \frac{d_3^{prev}}{d_4^{prev}} \right \leq 0.5$
$ d_1 - d_1^{prev} \leq 1.0$
$ d_2 - d_2^{prev} \leq 1.0$
$\frac{d_1}{d_2} \leq 1.1$ και $\frac{d_2}{d_1} \leq 1.1$
$\frac{d_1}{d_4} \leq 1.1$ και $\frac{d_4}{d_1} \leq 1.1$
$\frac{d_2}{d_4} \leq 1.1$ και $\frac{d_4}{d_2} \leq 1.1$
$d_1 \leq 0.4d_3$
$d_2 \leq 0.4d_3$
$d_4 \leq 0.4d_3$

Πίνακας 5.2: Οι γεωμετρικοί περιορισμοί για τα μάτια, όπως προκύπτουν από ανθρωπομορφικούς παράγοντες, υποθέσεις κίνησης και πειράματα.



Σχήμα 5.6: Εικόνα πάνω αριστερά : Η εκτίμηση (πράσινη γραμμή) και η τελική θέση (κόκκινη γραμμή) του ευθύγραμμου τμήματος, που προκύπτει από τον αλγόριθμο. Εικόνα πάνω κέντρο : Η περιοχή αναζήτησης που προκύπτει (με κόκκινο τα τελικά σημεία). Εικόνα πάνω δεξιά : Τα τελικά σημεία, όπως προέκυψαν από τη μέθοδο. Οι κάτω εικόνες αποτελούν μία μεγέθυνση των αντίστοιχων επάνω εικόνων για καλύτερη οπτική απεικόνιση.

5.4.2 Αναζήτηση μύτης και στόματος

Η αναζήτηση των σημείων της μύτης και του στόματος γίνεται με βάση την ίδια διαδικασία που περιγράψαμε για τα μάτια (5.4.1). Η μόνη διαφορά είναι ότι πρόκειται πλέον για δύο (2) σημεία, οπότε η συσχέτιση γίνεται για $k = 2$ μπλοκ. Λαμβάνουμε όμως υπόψη και το γεγονός της παραλληλίας των ευθυγράμμων τμημάτων, στο οποίο αναφερθήκαμε στο σχήμα 5.1, για να περιορίσουμε ακόμα περισσότερο την περιοχή αναζήτησης. Εφόσον τα ευθύγραμμα τμήματα θα πρέπει να είναι παράλληλα και έχοντας βρει τα σωστά σημεία των ματιών κρατάμε την κλίση α του ευθύγραμμου τμήματος των ματιών, που προκύπτει από τα σωστά σημεία, και απαιτούμε τα ευθύγραμμα τμήματα της μύτης και του στόματος να έχουν την ίδια κλίση α . Αφαιρείται, λοιπόν, από τον αλγόριθμο ο βρόχος της περιστροφής. Ο αλγόριθμος που αφορά τη μύτη και το στόμα είναι ίδιος, αφού αφορά δύο σημεία κάθε φορά και απεικονίζεται στον πίνακα 5.3.

1	Εκτίμηση των συντεταγμένων των δύο σημείων είτε της μύτης, είτε του στόματος P_i^{est} , $i = 1, 2$
2	Εύρεση κέντρου C ευθύγραμμου τμήματος $P_1^{est}P_2^{est}$ (για να γίνει η μεταφορά με βάση το κέντρο)
3	Εύρεση της κλίσης α του ευθύγραμμου τμήματος των ματιών
4	Μέγιστη συσχέτιση $maxCor = -2$
5	\forall μεταφορά t_x στον x άξονα {
6	$\vdots \forall$ μεταφορά t_y στον y άξονα {
7	$\vdots \vdots$ Τα νέα σημεία που προκύπτουν είναι TP_i , $i = 1, 2$
8	$\vdots \vdots$ Εύρεση κέντρου TC ευθύγραμμου τμήματος TP_1TP_4
9	$\vdots \vdots$ Περιστροφή $r = \tan^{-1}(\alpha)$ γύρω από το κέντρο TC $\vdots \vdots$ ώστε η κλίση του ευθύγραμμου τμήματος να είναι ίση $\vdots \vdots$ με την κλίση του ευθ. τμήματος των ματιών
10	$\vdots \vdots$ Τα νέα σημεία που προκύπτουν είναι RP_i , $i = 1, 2$
11	$\vdots \vdots \forall$ μεταφορά s του σημείου RP_1 στην ευθεία RP_1RP_2 {
12	$\vdots \vdots \vdots \forall$ μεταφορά s του σημείου RP_2 στην ευθεία RP_1RP_2 {
13	$\vdots \vdots \vdots \vdots$ Τα νέα σημεία που προκύπτουν είναι SP_i , $i = 1, 2$
14	$\vdots \vdots \vdots \vdots$ Υπολογισμός συσχέτισης cor πολλαπλών μπλοκ με κέντρα $\vdots \vdots \vdots \vdots$ τα σημεία SP_i , $i = 1, 2$
15	$\vdots \vdots \vdots \vdots$ Αν $cor \geq maxCor$ {
16	$\vdots \vdots \vdots \vdots \vdots$ Αν ικανοποιούνται οι γεωμετρικοί περιορισμοί {
17	$\vdots \vdots \vdots \vdots \vdots \vdots$ $maxCor = cor$
18	$\vdots \vdots \vdots \vdots \vdots \vdots$ Κράτα τα σημεία SP_i , $i = 1, 2$ ως σωστά
	$\vdots \vdots \vdots \vdots \vdots \vdots$ }
	$\vdots \vdots \vdots \vdots \vdots$ }
	$\vdots \vdots \vdots \vdots$ }
	$\vdots \vdots \vdots$ }
	$\vdots \vdots$ }
	\vdots }
	}

Πίνακας 5.3: Ο αλγόριθμος αναζήτησης μύτης, στόματος. Στο βήμα 3 υπολογίζεται η κλίση του ευθύγραμμου τμήματος των ματιών, ώστε να ισχύσει η παραλληλία. Στα βήματα 5 και 6 γίνεται η μεταφορά του ευθύγραμμου τμήματος. Στο βήμα 9 το νέο ευθύγραμμο τμήμα περιστρέφεται γύρω από το κέντρο του σύμφωνα με την κλίση α του βήματος 3. Τα βήματα 11 – 12 προσομοιώνουν την κλιμάκωση. Στο βήμα 14 υπολογίζεται η συσχέτιση για $k = 2$ μπλοκ. Τέλος, στο βήμα 16 ελέγχονται οι γεωμετρικοί περιορισμοί.

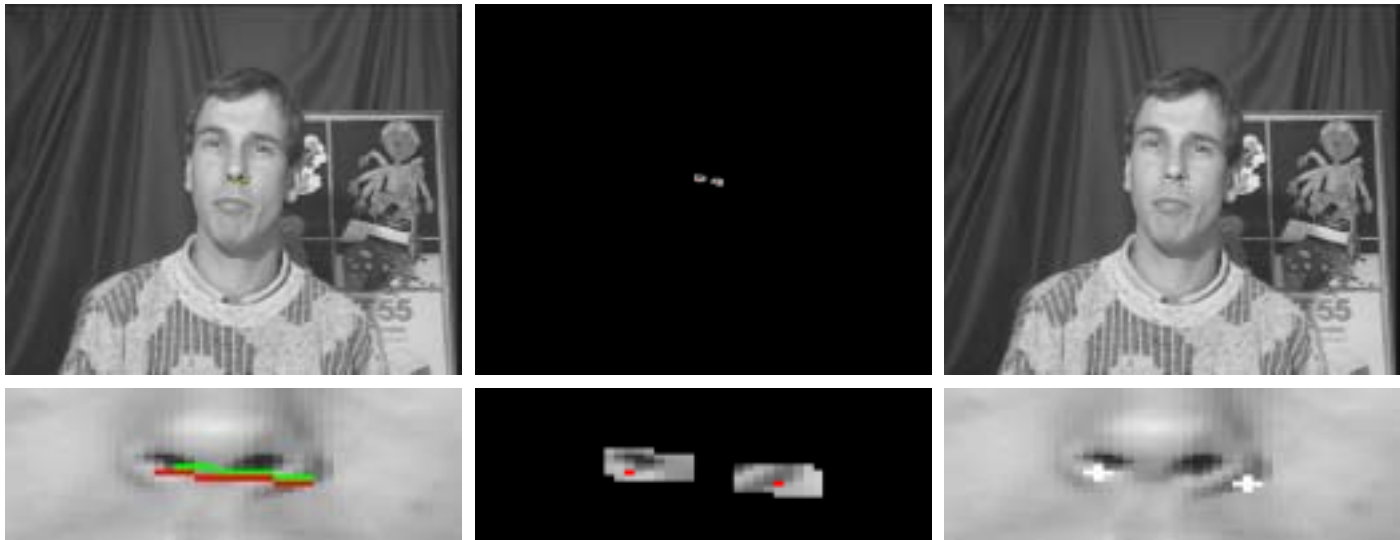
Η μόνη διαφορά στον αλγόριθμο είναι οι γεωμετρικοί περιορισμοί που ισχύουν για τη μύτη και για το στόμα. Ορίζουμε την απόσταση d_n ως την ευκλείδεια απόσταση των δύο σημείων της μύτης, την απόσταση d_m ως την ευκλείδεια απόσταση των δύο σημείων του στόματος, ενώ κρατάμε τις αποστάσεις d_3 και d_4 , όπως ορίστηκαν προηγουμένως για τα μάτια. Η απόσταση της μύτης είναι λίγο μεγαλύτερη από το τμήμα μεταξύ των δύο ματιών (d_4). Η απόσταση d_m είναι μικρότερη της d_3 και κατά πολύ μεγαλύτερη της d_4 . Το στόμα παρουσιάζει μία ιδιομορφία, αφού κινείται κατά την ομιλία και σύμφωνα με διάφορες εκφράσεις (π.χ. χαμόγελο). Σε σχέση, λοιπόν, με όλες τις υπόλοιπες αποστάσεις, η απόσταση των σημείων του στόματος μεταβάλλεται περισσότερο. Οι γεωμετρικοί περιορισμοί της μύτης φαίνονται στον πίνακα 5.4, ενώ αυτοί του στόματος παρουσιάζονται στον πίνακα 5.5. Στα σχήματα 5.7 και 5.8 φαίνεται αντίστοιχα για τη μύτη και για το στόμα, πώς από την εκτίμηση φτάνουμε στην περιοχή αναζήτησης και στο τελικό αποτέλεσμα.

$\left \frac{d_3}{d_n} - \frac{d_3^{prev}}{d_n^{prev}} \right \leq 0.5$
$ d_n - d_n^{prev} \leq 1.0$
$\frac{d_n}{d_4} \leq 1.3 \text{ και } \frac{d_4}{d_n} \leq 1.3$
$d_n \leq 0.4d_3$

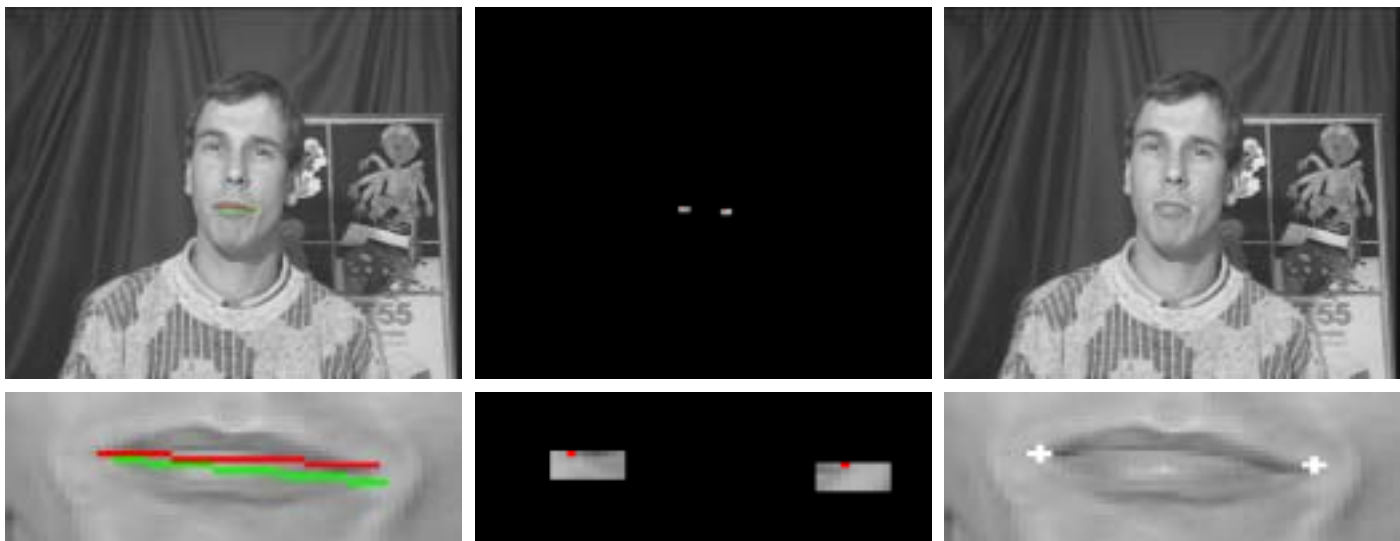
Πίνακας 5.4: Οι γεωμετρικοί περιορισμοί για τη μύτη, όπως προκύπτουν από ανθρωπομορφικούς παράγοντες, υποθέσεις κίνησης και πειράματα. Οι αποστάσεις d_3 και d_4 ορίστηκαν για τα μάτια.

$\left \frac{d_3}{d_m} - \frac{d_3^{prev}}{d_m^{prev}} \right \leq 0.5$
$ d_m - d_m^{prev} \leq 1.5$
$d_4 < 0.6d_m \text{ και } d_m < 0.8d_3$
$d_m \leq 0.75d_3$

Πίνακας 5.5: Οι γεωμετρικοί περιορισμοί για το στόμα, όπως προκύπτουν από ανθρωπομορφικούς παράγοντες, υποθέσεις κίνησης και πειράματα. Οι αποστάσεις d_3 και d_4 ορίστηκαν για τα μάτια.



Σχήμα 5.7: Εικόνα πάνω αριστερά : Η εκτίμηση (πράσινη γραμμή) και η τελική θέση (κόκκινη γραμμή) του ευθύγραμμου τμήματος, που προκύπτει από τον αλγόριθμο. Εικόνα πάνω κέντρο : Η περιοχή αναζήτησης που προκύπτει (με κόκκινο τα τελικά σημεία). Εικόνα πάνω δεξιά : Τα τελικά σημεία, όπως προέκυψαν από τη μέθοδο. Οι κάτω εικόνες αποτελούν μία μεγέθυνση των αντίστοιχων επάνω εικόνων για καλύτερη οπτική απεικόνιση.



Σχήμα 5.8: Εικόνα πάνω αριστερά : Η εκτίμηση (πράσινη γραμμή) και η τελική θέση (κόκκινη γραμμή) του ευθύγραμμου τμήματος, που προκύπτει από τον αλγόριθμο. Εικόνα πάνω κέντρο : Η περιοχή αναζήτησης που προκύπτει (με κόκκινο τα τελικά σημεία). Εικόνα πάνω δεξιά : Τα τελικά σημεία, όπως προέκυψαν από τη μέθοδο. Οι κάτω εικόνες αποτελούν μία μεγέθυνση των αντίστοιχων επάνω εικόνων για καλύτερη οπτική απεικόνιση.

5.5 Παρατηρήσεις

Στην ενότητα αυτή εξηγήσαμε τη σημασία των οκτώ σημείων που χρησιμοποιούμε για την παρακολούθηση. Γίνεται κατανοητό γιατί επελέγησαν τα σημεία αυτά και παρουσιάσαμε πως με τη βοήθεια της πρόβλεψης και χρησιμοποιώντας τέσσερις παραμέτρους περιορίζουμε την περιοχή αναζήτησης. Σημειώνουμε ότι οι παράμετροι αυτοί ορίζονται προς το παρόν σαν σταθερές και ενδέχεται να αλλάζουν από ακολουθία σε ακολουθία, ανάλογα με την κίνηση. Όμως θα μπορούσαν να προσαρμοστούν με αυτόματο τρόπο στην κίνηση και να μεταβάλλονται μέσα από τη μέθοδο. Παρουσιάσαμε αλγορίθμους αναζήτησης των σημείων αυτών και κατ' επέκταση των χαρακτηριστικών (μάτια, μύτη, στόμα) του προσώπου, αφού ορίσαμε την αντιστοίχιση πολλαπλών μπλοκ. Είδαμε ότι τα σημεία ξεχωριστά είναι αρκετά αδύναμα, αλλά με την ομαδοποίησή τους αναπτύξαμε μία μέθοδος που επιβεβαιώνεται από τα αποτελέσματα. Η μέθοδος υποβοηθάται από γεωμετρικούς περιορισμούς, οι οποίοι όμως είναι αρκετά απλοί και δεν έχουν μεγάλο κόστος. Θα σταθούμε σε δύο σημεία : (α) στο ευθύγραμμο τμήμα των σημείων του ματιού, και (β) στην αναζήτηση του στόματος.

Θεωρούμε απαραίτητο τα σημεία που αφορούν τα μάτια να βρίσκονται πάνω σε ένα ευθύγραμμο τμήμα. Το τμήμα αυτό παίζει σημαντικό ρόλο καθ' όλη τη διαδικασία της αναζήτησης των σημείων των ματιών. Αν τα σημεία σε κάποια στιγμή-εικόνα δεν είναι πάνω στο ευθύγραμμο τμήμα, η μέθοδος διαλέγει τα σημεία του ευθύγραμμου τμήματος πλησιέστερα σ' αυτά. Λόγω της προβολής και των διακριτών συντεταγμένων της εικόνας, αυτό μπορεί να συμβεί, αλλά δεν θα επιφέρει λανθασμένο αποτέλεσμα. Δεν θα φτάσουμε δηλαδή ποτέ στο σημείο να χαρακτηρίσουμε τα ευρεθέντα σημεία ως μη ανήκοντα στις άκρες των ματιών.

Οι γεωμετρικοί περιορισμοί είναι απαραίτητοι για τη μέθοδο καθότι συμβάλλουν στην αποφυγή λαθών, που έχουν την τάση να συσσωρεύονται και να μεγαλώνουν το τελικό λάθος. Τα πειραματικά αποτελέσματα επιβεβαιώνουν τη χρήση των περιορισμών αυτών, ιδιαίτερα όταν αντιμετωπίζουμε προβλήματα ποιότητας στην ακολουθία εικόνων. Επισημαίνουμε ότι, ακόμα και σε ελεγχόμενα περιβάλλοντα, η κάμερα δεν δίνει τις ίδιες τιμές από εικόνα σε εικόνα. Σε πολλές περιπτώσεις όμως η αναζήτηση των σημείων του στόματος παρουσιάζει από ικανά έως πολύ καλά αποτελέσματα και χωρίς τους γεωμετρικούς περιορισμούς που αφορούν το στόμα. Αυτό υποθέτουμε ότι οφείλεται στις ιδιαίτερα ισχυρές ακμές που παρουσιάζονται στις άκρες του στόματος, με αποτέλεσμα τα σημεία αυτά να εμπεριέχουν ήδη αρκετή πληροφορία για να δουλέψει η αντιστοίχιση. Η παρατήρηση αυτή είναι ιδιαίτερα θετική καθώς απαλλάσσει τη μέθοδο από μία σειρά υπολογισμών και ελέγχων.

Θα πρέπει να επισημάνουμε όμως και τα αρνητικά αυτής της μεθοδολογίας. Παρατηρούμε, από την επεξήγηση της μεθόδου ότι το λάθος είναι συσσωρευτικό, με αποτέλεσμα να οδηγούμαστε σε σφάλματα που δεν θα επιτρέψουν ανάρρωση. Αυτό γιατί η αντιστοίχιση γίνεται πάντα μεταξύ δύο εικόνων, της εικόνας προς επεξεργασία και

της αμέσως προηγούμενης. Έτσι, αν παρουσιαστεί κάποια στιγμή ένα λάθος, πράγμα που συμβαίνει συχνά, αυτό μπορεί να μην παίζει σοβαρό ρόλο για τις αμέσως επόμενες εικόνες, αλλά εφόσον η αντιστοίχιση εκεί θα γίνεται πλέον με λανθασμένο πρότυπο, είναι πολύ πιθανόν να μας δώσει και λανθασμένο αποτέλεσμα, το λάθος του οποίου θα είναι μεγαλύτερο σε σχέση με το αρχικό, κοκ. Οι επιπτώσεις αυτού του λάθους θα φανούν ύστερα από αρκετές εικόνες, όπου τα ευρεθέντα σημεία θα έχουν ξεφύγει κατά πολύ από τα σωστά. Είμαστε, λοιπόν, αναγκασμένοι να αντιμετωπίσουμε τέτοιες καταστάσεις, πριν να είναι πολύ αργά. Ο ευκολότερος τρόπος να χειριστούμε τέτοια λάθη, είναι, όταν διαπιστωθούν, να ξανααρχικοποιηθεί η μέθοδος με σωστά σημεία. Για να διαπιστώσουμε πότε το λάθος θα είναι αρκετά μεγάλο, ώστε οι συνέπειές του να είναι καταστροφικές, κρατάμε ένα μικρό ιστορικό των συσχετίσεων για κάθε μία από τις τρεις ομάδες των σημείων. Η συσχέτιση μπορεί κάποια στιγμή να αποκλίνει προσωρινά από το ιστορικό αυτό, αλλά αυτό να οφείλεται σε κάποιες συμπτώσεις (διαταραχή των φωτεινοτήτων λόγω φωτισμού, ποιότητας λήψης ή και προσωρινή μη ικανοποίηση των γεωμετρικών περιορισμών, λόγω της προβολής ή και της ίδιας της κίνησης). Όταν όμως συνεχίζει να αποκλίνει σημαίνει ότι οδηγούμαστε σε μεγάλο λάθος και πρέπει να επέμβουμε. Ένα ιστορικό των 5 με 9 τελευταίων συσχετίσεων και μία σύγκριση με την ενδιάμεση τιμή είναι ικανή να μας προσδιορίσει τη στιγμή της επέμβασης. Ιδιαίτερα σημαντικό είναι το λάθος όταν συμβαίνει στην περίπτωση των σημείων των ματιών, γιατί με βάση το ευθύγραμμο τμήμα, που σχηματίζεται από αυτά, προσδιορίζονται και τα υπόλοιπα. Έτσι, όταν τα μάτια προσδιορίζονται σωστά, έχουμε λιγότερες πιθανότητες να κάνουμε λάθος στα υπόλοιπα σημεία (εκτός βέβαια αν μεσολαβούν άλλοι παράγοντες, όπως είναι κάποια κίνηση του στόματος που δεν καλύπτεται από τη μέθοδο ή αν η πληροφορία των σημείων της μύτης είναι ιδιαίτερα χαμηλή). Στις περισσότερες περιπτώσεις μπορούμε να βασιστούμε ικανοποιητικά στα μάτια, και άρα να παραλείψουμε την επιπρόσθετη αυτή διαδικασία για τα δύο άλλα χαρακτηριστικά.

Κεφάλαιο 6

Πειραματικά αποτελέσματα

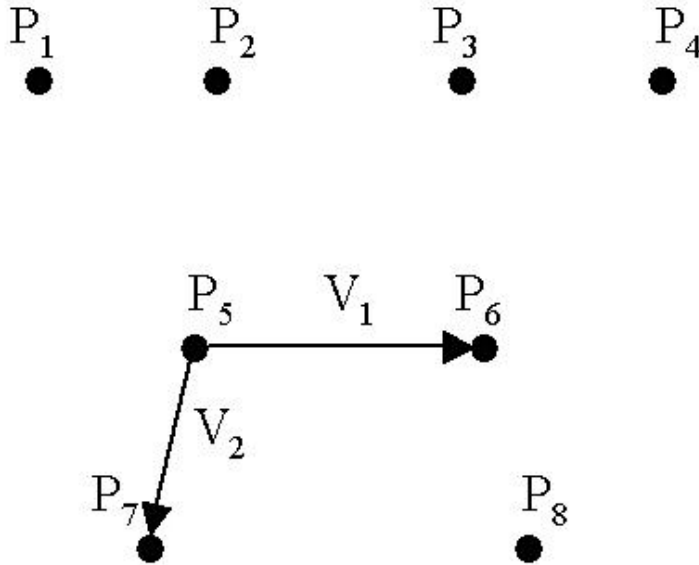
Στο κεφάλαιο αυτό θα εκθέσουμε τα πειραματικά αποτελέσματα που προκύπτουν από την προτεινόμενη μεθοδολογία. Οι δοκιμές έγιναν σε τρία στάδια. Το πρώτο αφορά την εκτίμηση των παραμέτρων κίνησης και γεωμετρίας στον τρισδιάστατο χώρο, ενώ το δεύτερο την παρακολούθηση των σημείων, και κατ' επέκταση του προσώπου σε ψηφιακές ακολουθίες εικόνων στο δισδιάστατο χώρο. Τέλος, επιχειρούμε να συνδυάσουμε τα δύο αυτά στάδια.

6.1 Εκτίμηση παραμέτρων κίνησης και γεωμετρίας στον 3-Δ χώρο

Η εκτίμηση παραμέτρων κίνησης και γεωμετρίας στον 3-Δ χώρο περιγράφηκε στο Κεφάλαιο 4. Για να επαληθεύσουμε τη μεθοδολογία αυτή χρησιμοποιούμε την τεχνική της προσομοίωσης. Ορίζουμε μία επίπεδη επιφάνεια στον 3-Δ χώρο και οκτώ σημεία πάνω σ' αυτήν. Τα σημεία αυτά αντιστοιχούν σε υποτιθέμενα χαρακτηριστικά προσώπου. Γνωρίζοντας τη θέση των σημείων στον 3-Δ χώρο $P_i(x, y, z)$, $i = 1, \dots, 8$, με το $O(0, 0, 0)$ να αντιστοιχεί στη θέση που βρίσκεται ο φακός, είμαστε σε θέση να προβάλλουμε τα σημεία αυτά στο 2-Δ χώρο της εικόνας χρησιμοποιώντας τους τύπους της προοπτικής προβολής (4.1). Για λόγους απλότητας θεωρούμε $f = 1$. Από τις συντεταγμένες του 3-Δ χώρου είμαστε επίσης σε θέση να υπολογίσουμε το κάθετο στο επίπεδο διάνυσμα N , που αποτελεί τον προσανατολισμό του υποτιθέμενου προσώπου. Ορίζουμε δύο διανύσματα πάνω στην επίπεδη επιφάνεια $V_1 = (P_5\vec{P}_6)$ και $V_2 = (P_5\vec{P}_7)$, τα οποία δεν είναι παράλληλα μεταξύ τους, όπως φαίνεται και στο σχήμα 6.1.

Οι συνιστώσες των διανυσμάτων αυτών προκύπτουν από τις :

$$V_1 = \begin{bmatrix} V_1(x) \\ V_1(y) \\ V_1(z) \end{bmatrix} = \begin{bmatrix} P_6(x) - P_5(x) \\ P_6(y) - P_5(y) \\ P_6(z) - P_5(z) \end{bmatrix}$$



Σχήμα 6.1: Τα οκτώ σημεία $P_i(x, y, z)$, $i = 1, \dots, 8$ που βρίσκονται πάνω σε μία επίπεδη επιφάνεια στον τρισδιάστατο χώρο.

$$V_2 = \begin{bmatrix} V_2(x) \\ V_2(y) \\ V_2(z) \end{bmatrix} = \begin{bmatrix} P_7(x) - P_5(x) \\ P_7(y) - P_5(y) \\ P_7(z) - P_5(z) \end{bmatrix}$$

Από την αναλυτική γεωμετρία προκύπτει το διάνυσμα N , που είναι κάθετο στα δύο αυτά διανύσματα, και κατά συνέπεια κάθετο και στην επίπεδη επιφάνεια, σύμφωνα με :

$$N = \begin{vmatrix} V_1(x) & V_1(y) & V_1(z) \\ V_2(x) & V_2(y) & V_2(z) \\ i & j & k \end{vmatrix} = \begin{bmatrix} V_1(y)V_2(z) - V_2(y)V_1(z) \\ V_2(x)V_1(z) - V_1(x)V_2(z) \\ V_1(x)V_2(y) - V_2(x)V_1(y) \end{bmatrix} = \begin{bmatrix} N_x \\ N_y \\ N_z \end{bmatrix}$$

όπου i, j, k τα μοναδιαία διανύσματα πάνω στους άξονες x, y, z αντίστοιχα. Προσαρμόζουμε το N έτσι ώστε να είναι μοναδιαίο και να δείχνει προς το φακό.

Στη συνέχεια τροφοδοτούμε το πρόγραμμα προσομοίωσης με διανύσματα που αφορούν την κίνηση, ένα για τη μεταφορική $T = (T_x, T_y, T_z)$ και ένα για την περιστροφική $\Omega = (\Omega_x, \Omega_y, \Omega_z)$. Χρησιμοποιώντας αυτά τα διανύσματα εφαρμόζουμε την κίνηση V , που ορίζουν, στα αρχικά σημεία και βρίσκουμε τις νέες θέσεις των σημείων $P'_i(x, y, z)$, $i = 1, \dots, 8$ ως εξής :

$$P'_i = P_i + V \Leftrightarrow$$

$$\begin{bmatrix} P'_i(x) \\ P'_i(y) \\ P'_i(z) \end{bmatrix} = \begin{bmatrix} P_i(x) \\ P_i(y) \\ P_i(z) \end{bmatrix} + \begin{bmatrix} V(x) \\ V(y) \\ V(z) \end{bmatrix}$$

όπου :

$$\begin{aligned} V(x) &= -(-Tx - \Omega_y P_i(z) + \Omega_z P_i(y)) \\ V(y) &= -(-Ty - \Omega_z P_i(x) + \Omega_x P_i(z)) \\ V(z) &= -(-Tz - \Omega_x P_i(y) + \Omega_y P_i(x)) \end{aligned}$$

Από τις νέες θέσεις των σημείων $P'_i(x, y, z), i = 1, \dots, 8$ μπορούμε να υπολογίσουμε και πάλι το σωστό κάθετο διάνυσμα στην επίπεδη επιφάνεια N' , όπως περιγράψαμε παραπάνω. Επίσης, χρησιμοποιώντας την προοπτική προβολή προβάλλουμε τα νέα σημεία στον 2-Δ χώρο της εικόνας.

Οι δύο εικόνες, όπως προκύπτουν από τα προηγούμενα, και συγκεκριμένα οι 2-Δ συντεταγμένες των σημείων σε κάθε εικόνα αποτελούν την είσοδο της διαδικασίας, η οποία θα εκτιμήσει τις παραμέτρους κίνησης και γεωμετρίας. Το αποτέλεσμα της εκτίμησης αυτής πρέπει να είναι τα δύο διανύσματα T και Ω , τα οποία αποτελούν την πραγματική κίνηση που εκτελέστηκε για να πάμε από τη μία εικόνα στην άλλη, καθώς και το διάνυσμα N' , που αποτελεί τον προσανατολισμό του υποτιθέμενου προσώπου στη δεύτερη εικόνα.

Έχοντας έξι τιμές, που μπορούμε να χρησιμοποιήσουμε ή όχι, προκύπτουν $2^6 = 64$ συνδυασμοί κίνησης. Δοκιμάστηκαν, λοιπόν, στην πρώτη προσομοίωση όλοι αυτοί οι συνδυασμοί για $T_x = 8, T_y = 5, T_z = 0.5$, όσον αφορά τη μεταφορά και $\Omega_x = 0.052, \Omega_y = 0.044$ και $\Omega_z = 0.026$, όσον αφορά την περιστροφή (οι συνιστώσες του διανύσματος περιστροφής αναγράφονται σε ακτίνια (radiants) και είναι περίπου 3, 2.5, 1.5 μοίρες αντίστοιχα). Έτσι, προκύπτουν όλες οι δυνατές περιπτώσεις κίνησης, απλές και σύνθετες, όπως είναι : μεταφορά σε ένα και μόνο άξονα, μεταφορά σε περισσότερους άξονες, περιστροφή γύρω από έναν και μόνο άξονα, περιστροφή γύρω από περισσότερους άξονες, μεταφορά σε έναν άξονα και συνδυασμοί περιστροφών, κοκ. Επίσης υπολογίζουμε και το γεγονός της μη ύπαρξης κίνησης, δηλαδή τα διανύσματα T και Ω είναι μηδενικά.

Ξέροντας όλα τα σωστά διανύσματα T, Ω και N και βρίσκοντας τα αντίστοιχα διανύσματα που δίνει ο εκτιμητής T', Ω' και N' , τα συγκρίνουμε χρησιμοποιώντας ένα είδος συσχέτισης διανυσμάτων. Η συσχέτιση αυτή υπολογίζεται σύμφωνα με :

$$\frac{V_\sigma \cdot V_\varepsilon}{\|V_\sigma\| \|V_\varepsilon\|}$$

όπου V_σ το σωστό και V_ε το διάνυσμα που βρέθηκε, $V_\sigma \cdot V_\varepsilon$ το εσωτερικό γινόμενο των δύο διανυσμάτων και “ $\| * \|$ ” συμβολίζει το μέτρο του διανύσματος. Η συσχέτιση αυτή φανερώνει κατά πόσο τα δύο διανύσματα ταυτίζονται, όσον αφορά την κατεύθυνση, και

είναι κανονικοποιημένη, δίνοντας τιμές από -1 μέχρι $+1$, όπου η τιμή -1 σημαίνει ότι τα διανύσματα είναι εντελώς αντίθετα, η 0 σημαίνει ότι είναι εντελώς άσχετα, δηλαδή κάθετα μεταξύ τους, και η $+1$ ότι ταυτίζονται.

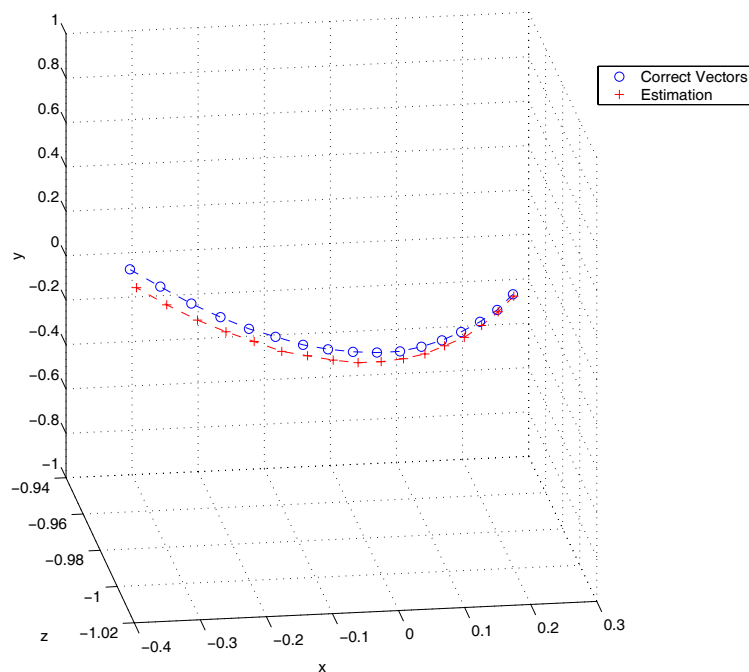
Επισημάνουμε στο σχετικό Κεφάλαιο (4) ότι η μέθοδος δεν μπορεί να βρει το κάθετο διάνυσμα N' όταν η κίνηση αποτελείται μόνο από περιστροφή. Σημειώσαμε όμως ότι αν ξέρουμε το κάθετο διάνυσμα στην προηγούμενη εικόνα (N) είναι δυνατόν να βρεθεί το N' περιστρέφοντας απλά το N με τον πίνακα R , δηλαδή $N' = RN$. Αρκεί, λοιπόν, στις περιπτώσεις αυτές το διάνυσμα της περιστροφής Ω' , που βρέθηκε, να είναι σωστό. Από το Ω προκύπτει απλά ο πίνακας R σύμφωνα με τον τύπο (4.4), όπου ϕ_1, ϕ_2, ϕ_3 είναι οι αντίστοιχες συνιστώσες $\Omega_x, \Omega_y, \Omega_z$ του Ω (σε ακτίνια, radiants).

Τα αποτελέσματα της προσομοίωσης είναι ακριβείας. Τα ευρεθέντα διανύσματα N' , και Ω' είναι, αν όχι ακριβώς, πολύ κοντά στα σωστά και ιδιαίτερα όσον αφορά το διάνυσμα του προσανατολισμού N' που μας ενδιαφέρει. Στις περιπτώσεις, όπου δεν μπορεί να εκτιμηθεί το N' , αρκεί απλώς να εφαρμόσουμε $N' = RN$, όπως περιγράψαμε προηγουμένως. Εφόσον όμως παρατηρούμε ότι το διάνυσμα Ω' είναι σωστό τότε και η περιστροφή RN θα δώσει ένα ευρεθέν διάνυσμα που θα είναι κι αυτό σωστό. Όσον αφορά το ευρεθέν διάνυσμα T' , έχουμε επισημάνει και στο αντίστοιχο Κεφάλαιο ότι δεν είναι δυνατόν να ανακτηθούν κανονικά το κάθετο διάνυσμα και οι παράμετροι της μεταφοράς, διότι μεσολαβεί ένας άγνωστος παράγοντας κλίμακας που δεν μπορεί να υπολογιστεί. Το κάθετο διάνυσμα δεν αποτελεί πρόβλημα, διότι μας ενδιαφέρει ουσιαστικά μόνο η κατεύθυνσή του. Επίσης, θεωρούμε πάντοτε το N μοναδιαίο το οποίο μας αρκεί για να προσδιορίσουμε τον προσανατολισμό του προσώπου. Οι παράμετροι της μεταφοράς όμως παραμένουν άγνωστοι με την έννοια ότι δεν μπορούμε να έχουμε το σωστό μέτρο του διανύσματος T' , λόγω του παράγοντα κλίμακας. Βέβαια, αν κοιτάξουμε την κατεύθυνση, καθώς και τους λόγους των συνιστωσών του διανύσματος T' σε σχέση με το σωστό T θα δούμε ότι δεν αποκλίνει σημαντικά από το σωστό και άρα μπορεί να θεωρηθεί σωστό, πράγμα που γίνεται φανερό με τη συσχέτιση που χρησιμοποιούμε. Η συσχέτιση των διανυσμάτων που συγκρίνονται ήταν επί το πλείστον 1, που σημαίνει ότι το ευρεθέν διάνυσμα ταυτίζεται με το σωστό.

Βέβαια έχουμε και σφάλματα. Τα σφάλματα είναι κατά κύριο λόγο ελάχιστα στο πλήθος αλλά και πάρα πολύ μικρά στην τιμή. Είναι φυσιολογικό να υπάρχουν σφάλματα, εφόσον μιλάμε για ανάκτηση των τρισδιάστατων παραμέτρων από το δισδιάστατο διακριτό χώρο χωρίς καμία άλλη βοήθεια. Όμως, το σφάλμα αυτό είναι πολύ μικρό που μόνο σε μεγάλες αποστάσεις μπορεί να παίξει σημαντικό ρόλο. Εφόσον όμως το πρόσωπο δεν απέχει πολύ από την κάμερα, τα ευρεθέντα διανύσματα καλύπτουν άριστα τις ανάγκες μας, πάντα όσον αφορά την προσομοίωση. Ενδεικτικά αναφέρουμε ότι το μέγιστο σφάλμα δεν ξεπερνάει το 0.002. Σημειώνουμε, επίσης, ότι κατά τη διάρκεια της προσομοίωσης, αλλά και όλων των προσομοιώσεων που ακολουθούν, δεν υπήρχε θόρυβος.

Μία διαφορετική προσομοίωση, που δοκιμάστηκε, είναι να επιβληθεί στα σημεία

μία συνεχόμενη κίνηση και να ανακτηθούν οι παράμετροι κίνησης και γεωμετρίας. Παραθέτουμε δύο περιπτώσεις. Στην πρώτη, η επίπεδη επιφάνεια περιστρέφεται από αριστερά προς τα δεξιά, όπως θα τη βλέπαμε σαν παρατηρητές από το σημείο του φακού. Εφόσον η κίνηση είναι περιστροφική, η μέθοδος αδυνατεί να εκτιμήσει το διάνυσμα του προσανατολισμού N' . Δώσαμε όμως σαν είσοδο στην προσομοίωση το διάνυσμα N του επιπέδου πριν ξεκινήσει την περιστροφή και εφαρμόσαμε το μετασχηματισμό RN για να βρούμε τον προσανατολισμό, ώστε να τον συγκρίνουμε με το σωστό. Η περιστροφή γίνεται γύρω από τον άξονα y και προσομοιώνει ένα πρόσωπο που κοιτάζει λίγο αριστερά και γυρίζει σιγά σιγά (βήμα = 2°) προς τα δεξιά, όπως θα το βλέπαμε εμείς σαν παρατηρητές. Η συσχέτιση έδειξε και πάλι ότι τα ευρεθέντα διανύσματα ταυτίζονται με τα σωστά, δηλαδή συσχέτιση = 1 ή ≈ 1 . Στο σχήμα 6.2 εμφανίζονται με μπλε κύκλους τα σωστά διανύσματα N και με κόκκινους σταυρούς οι εκτιμήσεις N' . Επειδή η κίνηση είναι συνεχής μπορούμε να σχηματίσουμε την καμπύλη της μεταβολής του σωστού διανύσματος N (μπλε διακεκομμένη γραμμή) και την αντίστοιχη καμπύλη της εκτίμησης N' (κόκκινη διακεκομμένη γραμμή) στον τρισδιάστατο χώρο. Παρατηρούμε ότι οι δύο καμπύλες σχεδόν ταυτίζονται, πράγμα που επιβεβαιώνει την ορθότητα των εκτιμήσεων.

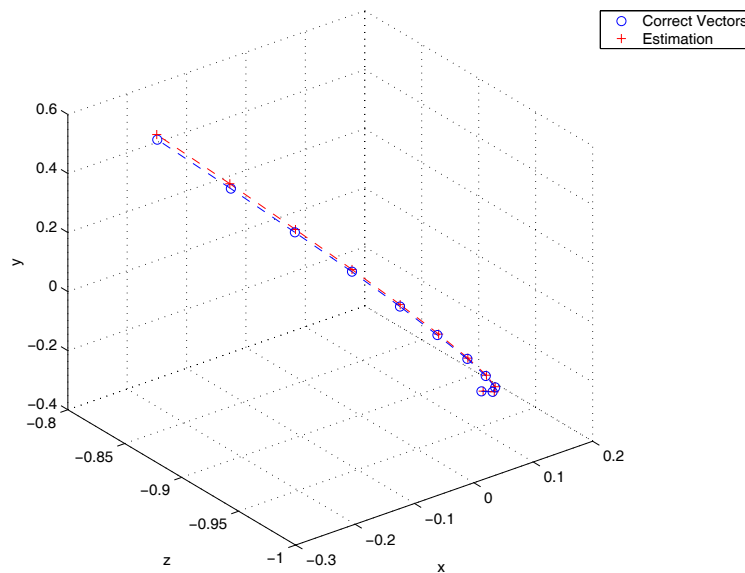


Σχήμα 6.2: Οι καμπύλες του σωστού (μπλε διακεκομμένη γραμμή) και του ευρεθέντος (κόκκινη διακεκομμένη γραμμή) διανύσματος N . Παρατηρούμε ότι σχεδόν ταυτίζονται, πράγμα που επιβεβαιώνει την ορθότητα των εκτιμήσεων. Με μπλε κύκλους φαίνονται τα σωστά και με κόκκινους σταυρούς τα ευρεθέντα διανύσματα του προσανατολισμού. Η προσομοίωση περιλαμβάνει περιστροφή γύρω από τον άξονα y .

Στη δεύτερη περίπτωση, η προσομοίωση είναι πιο πολύπλοκη. Προσομοιώνεται ένα πρόσωπο που κοιτάζει κάτω αριστερά και σιγά σιγά γυρίζει να κοιτάξει πάνω δεξιά.

Γεώργιος Σημαντήρης

Συμπεριλαμβάνονται και μεταφορές. Στον x άξονα έχουμε περιστροφή από -25° ως $+25^\circ$ και μεταφορά από -25 ως $+25$ εκατοστά, στον y περιστροφή από -10° ως $+10^\circ$ και μεταφορά από -10 ως $+10$ εκατοστά και στον z άξονα μόνο μεταφορά από 0.5 ως -0.5 εκατοστά. Η σύγκριση των εκτιμήσεων με τα σωστά διανύσματα έδωσε και πάλι επί το πλείστον συσχέτιση $= 1$ ή ≈ 1 , με ελάχιστα σφάλματα. Αυτό μπορούμε να το διαπιστώσουμε κι από το σχήμα 6.3, που παρουσιάζονται οι καμπύλες μεταβολής των σωστών διανυσμάτων N (μπλε διακεκομμένη γραμμή) και των εκτιμήσεων N' (κόκκινη διακεκομμένη γραμμή).



Σχήμα 6.3: Οι καμπύλες του σωστού (μπλε διακεκομμένη γραμμή) και του ευρεθέντος (κόκκινη διακεκομμένη γραμμή) διανύσματος N . Παρατηρούμε ότι σχεδόν ταυτίζονται, πράγμα που επιβεβαιώνει την ορθότητα των εκτιμήσεων. Με μπλε κύκλους φαίνονται τα σωστά και με κόκκινους σταυρούς τα ευρεθέντα διανύσματα του προσανατολισμού. Η προσομοίωση περιλαμβάνει μεταφορά στους άξονες x , y , z και περιστροφή στους άξονες x , y .

Και στις δύο περιπτώσεις τα αποτελέσματα είναι παραπάνω από ικανοποιητικά. Δεν παραλείπουμε να αναφέρουμε ότι παρουσιάζονται σφάλματα, αλλά και πάλι είναι πάρα πολύ μικρά για να έχουν σοβαρές συνέπειες. Σε πρώτη φάση διαπιστώνουμε, λοιπόν, ότι η μέθοδος εκτίμησης των τρισδιάστατων παραμέτρων κίνησης και γεωμετρίας δουλεύει αρκετά ικανοποιητικά, βασιζόμενοι πάντα στα αποτελέσματα των προσομοιώσεων. Μην ξεχνάμε όμως ότι η προσομοίωση γίνεται με δεδομένα ακριβείας, χωρίς παρεμβολή θορύβου.

6.2 Παρακολούθηση προσώπων με βάση τα σημεία

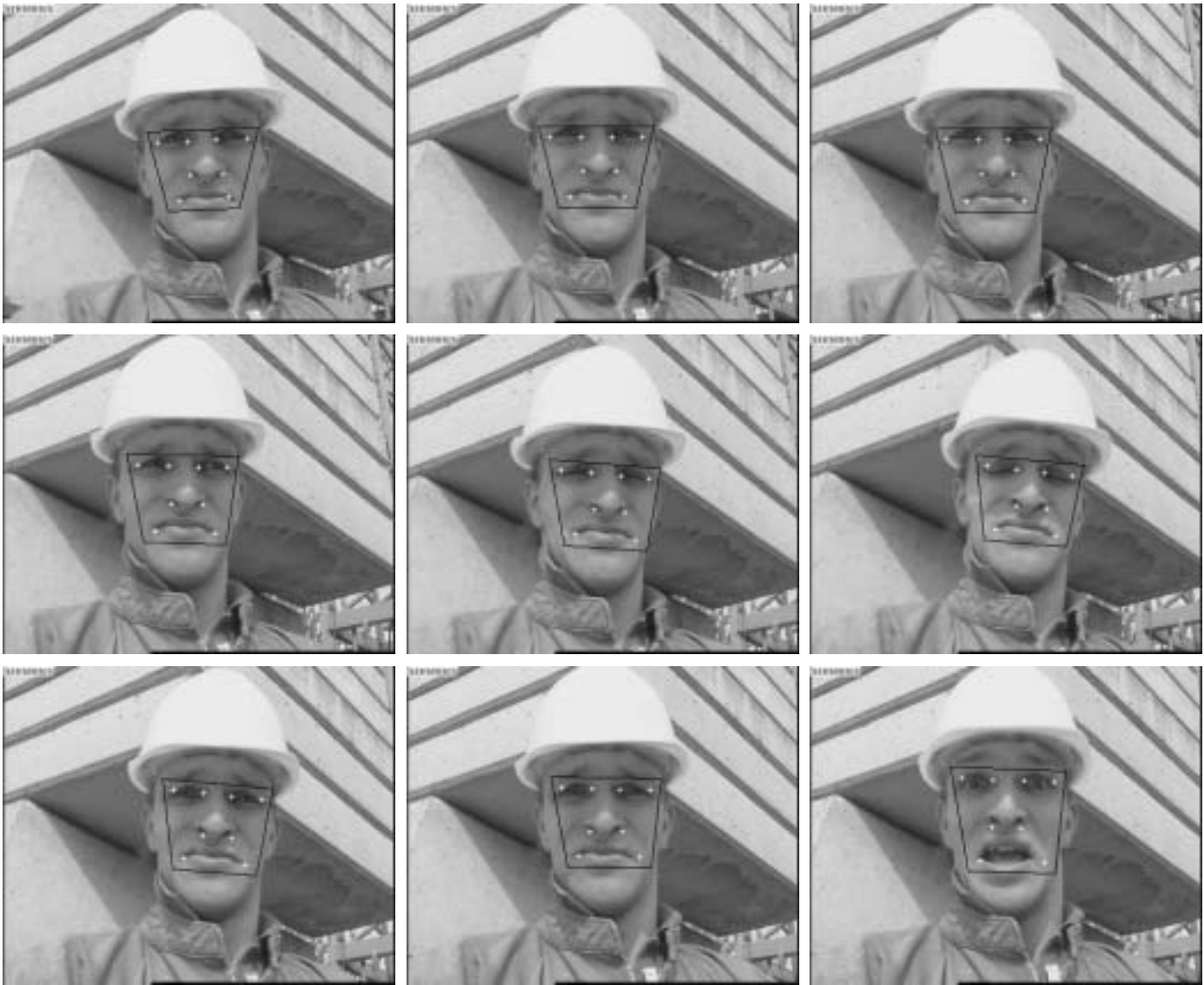
Η μέθοδος της αντιστοίχισης πολλαπλών μπλοκ υποβοηθούμενη από γεωμετρικούς περιορισμούς δοκιμάστηκε σε ψηφιακές ακολουθίες εικόνων που περιέχουν ένα πρόσωπο. Αρχικοποιήθηκαν τα σημεία στις δύο πρώτες εικόνες της ακολουθίας και η εκτίμηση των παραμέτρων κίνησης και γεωμετρίας χρησιμοποιείται μόνο για την πρόβλεψη, όπως περιγράφηκε στο Κεφάλαιο 5. Οι γνωστές για τις δοκιμές ακολουθίες “erik” και “foreman” επιστρατεύτηκαν για το σκοπό αυτό.

Η μέθοδος δοκιμάστηκε με την ακολουθία “erik”, όπου η κίνηση είναι αρκετά μικρή, αλλά περιλαμβάνει δύσκολες καταστάσεις, όπως είναι μία περιστροφή του προσώπου ως προς τον y άξονα, που κορυφώνεται στις εικόνες 25, 26, και όπου το αριστερό άκρο του αριστερού ματιού (όπως βλέπουμε την εικόνα) φτάνει σε σημείο να ταυτίζεται σχεδόν με το περίγραμμα του προσώπου. Δεδομένου ότι στο άκρο αυτό βρίσκεται ένα από τα σημεία που παρακολουθούμε (το P_1) θα περίμενε κανείς να αποτύχει η μέθοδος, αλλά κάτι τέτοιο δεν συμβαίνει. Η μέθοδος λειτουργεί με επιτυχία στη δύσκολη αυτή περίπτωση και καταφέρνει στις επόμενες εικόνες να επαναφέρει σωστά το σημείο κοντά στη θέση που πρέπει να είναι. Επίσης, επισημαίνουμε το γεγονός ότι το πρόσωπο βρίσκεται αρκετά μακριά από το φακό και δύσκολα ξεχωρίζουν τα σημεία που παρακολουθούνται, ακόμα και με το ανθρώπινο μάτι. Αυτό σημαίνει ότι η πληροφορία που εμπεριέχουν τα σημεία είναι λίγη, δεδομένου μάλιστα ότι η ακολουθία εξήχθη από ένα MPEG βίντεο, πράγμα που σημαίνει ότι μεσολαβεί και κάποια συμπίεση. Οι παράμετροι που χρησιμοποιήθηκαν για την ακολουθία αυτή, και που ουσιαστικά με τη βοήθειά τους ορίζεται η περιοχή αναζήτησης των σημείων, είναι $|t_x| = |t_y| = 2$ εικονοστοιχεία (pixels) για τη μεταφορά, τόσο στον x όσο και στον y άξονα, $|r| = 4^\circ$ για την περιστροφή και $|s| = 2$ εικονοστοιχεία για την κλιμάκωση. Αποσπάσματα των αποτελεσμάτων παρουσιάζονται στο σχήμα 6.4.



Σχήμα 6.4: Στις παραπάνω εικόνες βλέπουμε τα αποτελέσματα της μεθόδου παρακολούθησης για αποσπάσματα από την ακολουθία “erik”. Από δεξιά προς τα αριστερά και από πάνω προς τα κάτω τα αποτελέσματα αντιστοιχούν στις εικόνες 2, 7, 12, 17, 22, 26, 31, 36, 41.

Η ακολουθία “foreman” είναι ιδιαίτερα δύσκολη, διότι το πρόσωπο του “επιστάτη” εκτελεί μεγάλη κίνηση, που τις περισσότερες φορές μπορεί να χαρακτηριστεί ανεξέλεγκτη και απρόβλεπτη. Στην εικόνα 62 ο “επιστάτης” κάνει μία εντελώς απρόβλεπτη έκφραση ανοίγοντας διάπλατα τα μάτια και το στόμα με αποτέλεσμα να δυσκολεύει τη μέθοδο ιδιαίτερα, και που τελικά θα οδηγηθεί στην αποτυχία, αφού δεν θα μπορέσει να ανακτήσει πλέον τα σωστά σημεία. Αποτυχία σημαίνει ότι η συσχέτιση των μπλοκ είναι αρκετά μικρή και άρα τα σημεία που βρέθηκαν θεωρούνται λάθος, που σημαίνει ότι πρέπει να ξανα-αρχικοποιηθούν. Μέχρι εκείνη τη στιγμή όμως, τα αποτελέσματα είναι αρκετά ικανοποιητικά, όπως παρουσιάζονται και στο σχήμα 6.5. Οι παράμετροι στην περίπτωση αυτή διαμορφώνονται ως εξής : μεταφορά $|t_x| = |t_y| = 3$, περιστροφή $|r| = 4^\circ$, κλιμάκωση $|s| = 4$. Επισημάνουμε ότι λόγω της μεγάλης περιοχής αναζήτησης που προκύπτει με τις παραμέτρους αυτές, η μεταφορά πραγματοποιήθηκε με βήμα 2 σε κάθε άξονα. Σημειώνουμε ότι και η ακολουθία αυτή εξήχθη από MPEG βίντεο.



Σχήμα 6.5: Στις παραπάνω εικόνες βλέπουμε τα αποτελέσματα της μεθόδου παρακολούθησης για αποσπάσματα από την ακολουθία “foreman”. Από δεξιά προς τα αριστερά και από πάνω προς τα κάτω τα αποτελέσματα αντιστοιχούν στις εικόνες 2, 10, 18, 26, 32, 40, 48, 56, 62.

Στα δύο αυτά παραδείγματα παρατηρούμε περιπτώσεις όπου τα σημεία, των ματιών αλλά και ιδιαίτερα της μύτης, δεν έχουν την αναμενόμενη ακρίβεια. Η επίτευξη ακρίβειας για τα σημεία αυτά είναι ιδιαίτερα δύσκολη υπόθεση. Ειδικά στην περίπτωση της μύτης, τα σημεία δεν εμπεριέχουν αρκετή πληροφορία με αποτέλεσμα η αντιστοίχιση να μην μπορεί να αποδώσει το μέγιστο. Για τα μάτια, ο περιορισμός ότι τα σημεία πρέπει να βρίσκονται πάνω σε ένα ευθύγραμμο τμήμα έχει ως συνέπεια πολλές φορές τα σημεία του ενός ματιού να μην πέφτουν ακριβώς στις σωστές θέσεις, λόγω της ακρίβειας εικονοστοιχείων που διαθέτουμε στο διακριτό χώρο. Θυμίζουμε ότι το ευθύγραμμο τμήμα δεν είναι ακριβώς ευθύγραμμο στο διακριτό χώρο και αναπαρίσταται μέσω του αλγόριθμου *Midpoint*. Επίσης, όταν δύο σημεία ενός ματιού εμπεριέχουν υψηλή πληροφορία, σε σχέση με τα δύο σημεία του άλλου ματιού, το ευθύγραμμο τμήμα θα προσαρμοστεί σ' αυτά, με αποτέλεσμα να "υποφέρουν" τα άλλα δύο. Υπενθυμίζουμε και πάλι ότι στην περίπτωση της ακολουθίας "foreman" η μεταφορά του ευθύγραμμου τμήματος για την εύρεση της περιοχής αναζήτησης πραγματοποιήθηκε με βήμα 2 στον κάθε άξονα, οπότε ένα μέρος της ευθύνης για την έλλειψη ακρίβειας οφείλεται και εκεί.

Όμως, παρόλα τα προβλήματα, η προτεινόμενη μέθοδος καταφέρνει και στις δύο ακολουθίες να δώσει ικανοποιητικά αποτελέσματα. Το πρόσωπο, σαν σύνολο, παρακολουθείται άπταιστα. Τα σημεία παρακολουθούνται με επαρκή ακρίβεια και πάντοτε ορίζουν τα χαρακτηριστικά στα οποία αντιστοιχούν. Ιδιαίτερα στα δύο σημεία του στόματος η μέθοδος εκπλήσσει με την αποτελεσματικότητά της, ειδικά στην περίπτωση του "erik" που μιλάει και επομένως το στόμα κινείται. Θυμίζουμε ότι οι γεωμετρικοί περιορισμοί για το στόμα μπορούν και να απαλειφθούν, πράγμα που έγινε στις συγκεκριμένες δοκιμές.

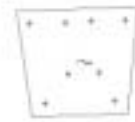
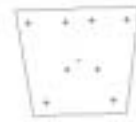
6.3 Συνδυασμός των δύο σταδίων - Η ολοκληρωμένη μεθοδολογία

Για να συνδυάσουμε τα δύο στάδια και να παρουσιάσουμε την ολοκληρωμένη μεθοδολογία, επιστρατεύτηκαν ακολουθίες που ελήφθησαν σε ελεγχόμενο περιβάλλον, για να είμαστε σε θέση να γνωρίζουμε και τον προσανατολισμό του προσώπου στην πρώτη εικόνα και έτσι να αρχικοποιήσουμε πλήρως τη μέθοδο. Το πρόσωπο στις ακολουθίες αυτές κοιτάζει πάντα το φακό στην πρώτη εικόνα της ακολουθίας, οπότε το κάθετο διάνυσμα είναι $N = (0, 0, -1)$. Επίσης, στις ακολουθίες αυτές είναι γεγονός ότι το πρόσωπο βρίσκεται κοντά στην κάμερα και τα σημεία που παρακολουθούμε είναι ιδιαίτερα ευδιάκριτα. Είμαστε, λοιπόν, σε θέση να χρησιμοποιήσουμε πλήρως όλες τις δυνατότητες της προτεινόμενης μεθοδολογίας.

Η ακολουθία "kutsurak" αποτελείται από συνολικά 586 εικόνες. Περιλαμβάνει ένα πρόσωπο, που αρχικά κοιτάει το φακό. Στη συνέχεια το πρόσωπο περιστρέφεται αριστερά-δεξιά και πάνω-κάτω, Τέλος, γίνονται και κάποιες διαγώνιες κινήσεις. Επισημαίνουμε ότι η κίνηση είναι ομαλή, χωρίς να περιλαμβάνει κάτι το απότομο, και απλή, ειδικά στις πρώτες 213 εικόνες. Οι παράμετροι που χρησιμοποιήθηκαν για την

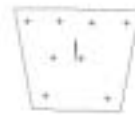
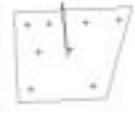
ακολουθία αυτή, όσον αφορά τον προσδιορισμό της περιοχής αναζήτησης των σημείων, είναι $|t_x| = |t_y| = 2$ εικονοστοιχεία (pixels) για τη μεταφορά, τόσο στον x όσο και στον y άξονα, $|r| = 4^\circ$ για την περιστροφή και $|s| = 2$ εικονοστοιχεία για την κλιμάκωση. Παραθέτουμε αποσπασματικά αποτελέσματα, όπου διακρίνονται το πρόσωπο που παρακολουθείται, τα αντίστοιχα σημεία, καθώς και το διάνυσμα της πόζας που εκτιμήθηκε. Το σχήμα 6.6 αφορά τις εικόνες 1 – 158, όπου η κίνηση περιλαμβάνει περιστροφή του προσώπου δεξιά-αριστερά και πάνω-κάτω, ενώ το σχήμα 6.7 τις εικόνες 215 – 297, όπου η κίνηση είναι διαγώνια, συμπεριλαμβάνει δηλαδή μεταφορά και περιστροφή, και άρα πιο πολύπλοκη.







Σχήμα 6.6: Στις παραπάνω εικόνες βλέπουμε τα αποτελέσματα της ολοκληρωμένης μεθόδου, παρακολούθηση σημείων και εκτίμηση πόζας, για αποσπάσματα από την ακολουθία ‘kutsurak’. Από πάνω προς τα κάτω τα αποτελέσματα αντιστοιχούν στις εικόνες 1, 15, 25, 35, 50, 70, 85, 100, 115, 135, 155.





Σχήμα 6.7: Στις παραπάνω εικόνες βλέπουμε τα αποτελέσματα της ολοκληρωμένης μεθόδου, παρακολούθηση σημείων και εκτίμηση πόζας, για αποσπάσματα από την ακολουθία ‘kutsurak’. Από πάνω προς τα κάτω τα αποτελέσματα αντιστοιχούν στις εικόνες 220, 230, 240, 255, 270, 280, 295.

Τα αποτελέσματα της ολοκληρωμένης μεθοδολογίας κρίνονται ικανοποιητικά. Από τη μία μεριά, η παρακολούθηση των σημείων φαίνεται καλή, πράγμα που σημαίνει ότι παρακολουθείται το πρόσωπο και τα επιλεγμένα σημεία προσδιορίζουν τη θέση των πραγματικών χαρακτηριστικών. Από την άλλη μεριά, όμως, η εκτίμηση της πόζας παρουσιάζει προβλήματα και δεν επιβεβαιώνει τα αποτελέσματα των προσομοιώσεων. Μπορούμε να διακρίνουμε ότι ο προσανατολισμός του προσώπου προσδιορίζεται στο περίπου, αλλά όχι επακριβώς. Ο εκτιμητής πόζας, κατά τη διάρκεια των πειραμάτων, έδινε σχετικά αλλά όχι επαρκώς ικανοποιητικά αποτελέσματα, παρόλο που η κίνηση είναι αρκετά απλή. Πράγμα, όμως, αναμενόμενο, που μπορεί να εξηγηθεί. Το πλήθος των σημείων, οι θέσεις των οποίων τροφοδοτούν τον εκτιμητή πόζας, είναι οκτώ. Δεδομένου ότι δεν παρουσιάζουν όλα κάποια συγκεκριμένη κίνηση, που οφείλεται πρώτον στην προβολή και δεύτερον στο διαχωρισμό τους σε ομάδες, συμβαίνει πολύ συχνά να μετακινούνται ορισμένα προς μία κατεύθυνση, ενώ κάποια άλλα προς μία άλλη κατεύθυνση. Αυτό αποτελεί παράγοντα σύγχυσης για τον εκτιμητή, ο οποίος φαίνεται να μπερδεύεται από τέτοια φαινόμενα διάχυσης. Έστω και να σημείο να ξεφεύγει για οποιοδήποτε λόγο από την πορεία που θα έπρεπε να ακολουθήσει είναι ικανό να προκαλέσει σύγχυση στον εκτιμητή. Επίσης, η έλλειψη της ακρίβειας, όσον αφορά τις θέσεις των σημείων, συμβάλλει σ'αυτήν τη σύγχυση. Τέλος, υπάρχουν πολλοί τρόποι κίνησης, σύμφωνα με τους οποίους μπορεί να εξηγηθεί η νέα θέση των σημείων στην εικόνα. Π.χ. μία ελαφρά περιστροφή ως προς τον y άξονα μπορεί να εξηγηθεί και ως απλή μεταφορά, δεδομένου ότι γνωρίζουμε μονάχα τις 2-Δ θέσεις των σημείων και δεν έχουμε καμία άλλη πληροφορία στη διάθεσή μας. Από τις πολλές πιθανές λύσεις, ο εκτιμητής θα δώσει μία, η οποία μπορεί να μην είναι η σωστή, αλλά να ικανοποιεί πλήρως τις σχέσεις, βάσει των οποίων γίνεται η εκτίμηση.

Παρατηρήσαμε κατά τη διάρκεια των πειραμάτων, ότι, όντως, ο εκτιμητής δείχνει περισσότερο “προτίμηση” στη μεταφορά παρά στην περιστροφή. Αυτό έχει αναπόφευκτες συνέπειες στον υπολογισμό του διανύσματος του προσανατολισμού, αφού τα δύο διανύσματα συνδέονται άρρηκτα μεταξύ τους. Έτσι οδηγούμαστε σε λάθη όσον αφορά την εκτίμηση του διανύσματος του προσανατολισμού. Για να αποφύγουμε τέτοιου είδους λάθη αναγκαστήκαμε να επιβάλλουμε επιπρόσθετους περιορισμούς. Θεωρήσαμε, και επιβεβαιωθήκαμε από τα αποτελέσματα, ότι εφόσον διαθέτουμε το ακριβές διάνυσμα του προσανατολισμού από την αρχικοποίηση, είναι καλύτερα να υπολογίζουμε το νέο διάνυσμα προσανατολισμού εφαρμόζοντας απευθείας το μετασχηματισμό RN . Το διάνυσμα περιστροφής που εκτιμάται, κι από το οποίο κατασκευάζεται ο πίνακας περιστροφής R , να μην έχει περίπου τη σωστή κατεύθυνση, αλλά υστερεί σε μέτρο. Με αυτόν τον τρόπο δεν είμαστε σε θέση να καταλογίσουμε ακρίβεια στο ευρεθέν διάνυσμα N , αλλά έχουμε ένα σωστότερο αποτέλεσμα από αυτό του ίδιου του εκτιμητή. Ακόμα, για να αποφύγουμε μερικώς τη σύγχυση, θεωρούμε ότι αν κάποια λίγα σημεία παρουσιάζουν φαινόμενα διάχυσης, σε σχέση με τα υπόλοιπα, επειδή μπορεί τα σημεία αυτά να διορθώνονται από τη μέθοδο, γιατί η προηγούμενή τους εκτίμηση ίσως να ήταν λανθασμένη,

τότε κρατάμε την προηγούμενη εκτίμηση του διανύσματος του προσανατολισμού, μιας και η προσπάθεια εκτίμησης του διανύσματος θα επηρεαστεί και θα είναι πιθανότατα λανθασμένη.

Τέλος, επισημάνουμε ότι η πρώτη υποακολουθία (εικόνες 1 – 58) χρειάστηκε να ξαναρχικοποιηθεί 3 φορές, ενώ η δεύτερη (εικόνες 215 – 297) δύο. Αυτό γιατί, όπως σημειώσαμε και στο αντίστοιχο Κεφάλαιο, το λάθος κατά την αναζήτηση των σημείων είναι συσσωρευτικό, με αποτέλεσμα κάποια στιγμή να γίνεται αρκετά μεγάλο και η συσχέτιση που εφαρμόζεται στα μπλοκ να δίνει ιδιαίτερα χαμηλές τιμές (σε σχέση πάντα με το αντίστοιχο ιστορικό). Όταν φτάσουμε σε τέτοιες τιμές είναι κατανοητό ότι η συνέχιση της μεθόδου μόνο λάθη μπορεί να επιφέρει, οπότε είναι προτιμητέο να σταματήσει για να ξαναρχικοποιηθεί.

Κεφάλαιο 7

Συμπεράσματα - Επίλογος

7.1 Ανασκόπηση

Η παρακολούθηση προσώπων αποτελεί ένα σημαντικό βήμα για πολλές εφαρμογές, όπως είναι συστήματα ασφαλείας και περιβάλλοντα τηλε-συνδιασκέψεων. Συμβάλλει δε ιδιαίτερα στην παραπέρα επεξεργασία, που αφορά τα πρόσωπα, όπως αναγνώριση προσώπων, ανάλυση έκφρασης, παρακολούθηση βλέμματος και αναγνώριση ομιλίας. Μία σειρά από σχετικές εργασίες έχουν δημοσιευθεί, που όμως περιορίζονται στο δισδιάστατο χώρο της εικόνας ή χρησιμοποιούν αρκετά πολύπλοκες μεθόδους για την ανάκτηση των παραμέτρων κίνησης και γεωμετρίας στον τρισδιάστατο χώρο. Παρουσιάσαμε μία μέθοδο αρκετά απλή, που παρακολουθεί χαρακτηριστικά προσώπου, όπως είναι τα μάτια, η μύτη και το στόμα, και κατ' επέκταση παρακολουθεί πρόσωπα σε ψηφιακές ακολουθίες εικόνων. Χρησιμοποιεί οκτώ σημεία, στρατηγικά τοποθετημένα πάνω στα χαρακτηριστικά αυτά, και μοντελοποιεί το πρόσωπο με μία επίπεδη επιφάνεια στον τρισδιάστατο χώρο. Η μοντελοποίηση αυτή είναι από τις απλούστερες στο είδος. Με τη βοήθεια των σημείων η μέθοδος επιχειρεί να ανακτήσει τις τρισδιάστατες παραμέτρους κίνησης, καθώς και τον προσανατολισμό του προσώπου, από δύο συνεχόμενες εικόνες της ακολουθίας. Οι παράμετροι που ανακτήθηκαν χρησιμοποιούνται για να προβλέψουν τις θέσεις των σημείων στην επόμενη εικόνα, όπου με βάση την πρόβλεψη αυτή περιορίζεται σημαντικά η περιοχή αναζήτησης των σημείων. Η περιοχή αναζήτησης ορίζεται από τέσσερις παραμέτρους, που αφορούν τη μεταφορά, την περιστροφή και την κλίμακα, και εντός της περιοχής αυτής γίνεται η διόρθωση των θέσεων των σημείων με βάση την αντιστοίχιση πολλαπλών μπλοκ και γεωμετρικούς περιορισμούς. Παρουσιάστηκαν αποτελέσματα που επαληθεύουν την επιτυχία της μεθόδου εν μέρει για την ανάκτηση του προσανατολισμού του προσώπου και των 3-Δ παραμέτρων κίνησης, όταν η κίνηση είναι απλή και ομαλή, και κυρίως για την παρακολούθηση των σημείων.

7.2 Συμπεράσματα

Η μεθοδολογία που αναπτύχθηκε αποτελεί μία ερευνητική προσπάθεια στο σχετικό τομέα της παρακολούθησης προσώπων. Τα επιμέρους τμήματά της, δηλαδή η ανάκτηση παραμέτρων κίνησης και γεωμετρίας και η παρακολούθηση σημείων, έχουν κατά καιρούς εφαρμοστεί με τον ένα ή τον άλλο τρόπο. Η επίπεδη επιφάνεια που συνδυάστηκε με το πρόσωπο εμφανίζει στοιχεία ότι μπορεί να μοντελοποιήσει σωστά το πρόσωπο και δίνει ικανοποιητικά αποτελέσματα υπό ορισμένες συνθήκες, αλλά και απλοποιεί ουσιαστικά τη μέθοδο σε σχέση με άλλα τρισδιάστατα μοντέλα που εφαρμόστηκαν σε σχετικές εργασίες. Τα σημεία που επιλέχθηκαν για την παρακολούθηση παρουσιάζουν μία σειρά από ιδιότητες που επιτρέπουν στη μέθοδο να κυμαίνεται σε σχετικά πολύ απλό επίπεδο. Εκτός αυτού, ο αριθμός τους είναι μικρός αλλά αρκετός για να ορίσουν το πρόσωπο.

Η ανάκτηση των τρισδιάστατων παραμέτρων κίνησης και γεωμετρίας επιτυγχάνεται μόνο από τις δισδιάστατες θέσεις των οκτώ σημείων πάνω σε δύο συνεχόμενες εικόνες. Αποφεύγονται έτσι χρονοβόρα και δαπανηρά ταιριάσματα με τρισδιάστατα μοντέλα προσώπων καθώς βέβαια και η κατασκευή τέτοιων μοντέλων, που αποτελεί από μόνη της μία χρονοβόρα και ακριβή διαδικασία. Επιπλέον δεν εξασφαλίζεται ότι το κατασκευασμένο μοντέλο θα ταιριάζει σε όλα τα πρόσωπα. Η χρήση της επίπεδης επιφάνειας απλοποιεί την τρισδιάστατη επεξεργασία. Ανακτώνται και οι έξι βαθμοί ελευθερίας που μπορεί να παρουσιάσει ένα πρόσωπο καθώς και ο προσανατολισμός του στον 3-Δ χώρο λύνοντας ένα γραμμικό σύστημα δεκαέξι εξισώσεων και οκτώ αγνώστων. Η λύση του συστήματος επιτυγχάνεται γρήγορα χρησιμοποιώντας τη Singular Value Decomposition για τη λύση γραμμικών συστημάτων. Στη συνέχεια, η Singular Value Decomposition ενός 3×3 πίνακα θα μας δώσει τρεις ιδιοτιμές. Οι ιδιοτιμές θα καθορίσουν πιο από τα τρία θεωρήματα, που παρουσιάσαμε, θα χρησιμοποιηθεί για να ανακτηθούν οι παράμετροι κίνησης και ο προσανατολισμός του προσώπου εφαρμόζοντας απλά τις αντίστοιχες σχέσεις.

Η παρακολούθηση των σημείων γίνεται καθαρά στις δύο διαστάσεις της εικόνας. Τέσσερις μόνο παράμετροι είναι ικανοί να δώσουν μία περιοχή στην οποία θα βρίσκονται σίγουρα οι σωστές θέσεις των σημείων που παρακολουθούνται. Ομαδοποιώντας τα σημεία κατά χαρακτηριστικά, εκμεταλλευόμενοι ιδιότητες της προοπτικής προβολής, της μορφολογίας του προσώπου καθώς και της θέσης του σε σχέση με την κάμερα και χρησιμοποιώντας απλούς γεωμετρικούς περιορισμούς, που προκύπτουν από ανθρωπομορφικές μελέτες και πειράματα, πετυχαίνουμε την παρακολούθηση χωρίς μεγάλο κόστος. Οι μόνοι υπολογισμοί αφορούν την αντιστοίχιση μπλοκ, μία διαδικασία που από σχετικά αδύναμη μετατράπηκε σε αρκετά αποτελεσματική με τη χρήση πολλαπλών μπλοκ. Η αντιστοίχιση κοστίζει, αλλά ο περιορισμός της σε μία μικρή, αλλά ικανή, περιοχή αναζήτησης, κρατάει το υπολογιστικό κόστος της μεθόδου σε ικανοποιητικά επίπεδα, επιτρέποντας μελλοντική παρακολούθηση σε πραγματικό χρόνο (βλ. και ενότητα 7.3).

Τα αποτελέσματα της προτεινόμενης μεθοδολογίας ενθαρρύνουν τον προσανατολισμό της έρευνας προς αυτήν την κατεύθυνση. Η εκτίμηση των παραμέτρων της τρισδιάστατης κίνησης και η ανάκτηση του προσανατολισμού του προσώπου στον τρισδιάστατο χώρο είναι αρκετά ακριβείς όσον αφορά τις προσομοιώσεις, αλλά απαιτείται επιπλέον έρευνα για να δώσουν το ίδιο καλά αποτελέσματα και σε πραγματικές ακολουθίες. Η δε παρακολούθηση των σημείων είναι παραπάνω από ικανοποιητική δεδομένης της απλότητας της μεθοδολογίας.

7.3 Μελλοντικές επεκτάσεις

Η λογική πάνω στην οποία κτίστηκε η προτεινόμενη μεθοδολογία διέπεται από απλότητα. Σίγουρα επιδέχεται βελτίωση, αλλά θα πρέπει αυτή να γίνει αναλογίζοντας και το σχετικό κόστος. Ωστόσο, υπάρχουν ορισμένα σημεία που έχουν άμεση προτεραιότητα για την επίτευξη περισσότερης αυτάρκειας, ταχύτητας και αποτελεσματικότητας.

Πρώτα απ' όλα είδαμε ότι η αρχικοποίηση της μεθόδου εξαρτάται αποκλειστικά από το χρήστη. Ο χρήστης καλείται να προσδιορίσει τις θέσεις συνολικά δεκαέξι σημείων (οκτώ για την πρώτη και άλλα οκτώ για τη δεύτερη εικόνα της ακολουθίας) καθώς και τον 3-Δ προσανατολισμό του προσώπου στην πρώτη εικόνα της ακολουθίας. Αυτό αποτελεί ένα κουραστικό γεγονός για το χρήστη γιατί απαιτείται ακρίβεια. Η αρχικοποίηση μπορεί να γίνει πιο ευχάριστη και πιο φιλική προς το χρήστη με τη χρήση διαφόρων προγραμμάτων διεπαφής, αλλά περισσότερο ικανοποιητικό θα ήταν να αποφευχθεί η ανάμιξή του χρήστη. Αυτό μπορεί να γίνει, όπως περιγράψαμε και στο Κεφάλαιο 3, προσαρμόζοντας ήδη υπάρχοντα προγράμματα ανίχνευσης προσώπων και χαρακτηριστικών προσώπων. Στόχος αυτής της επέκτασης θα είναι η ανεξαρτητοποίηση της μεθόδου από οποιαδήποτε εξωτερική παρέμβαση-επέμβαση.

Η προτεινόμενη μέθοδος παρουσιάζει μη ακριβή αποτελέσματα στην εκτίμηση της κίνησης και του προσανατολισμού, όσο η κίνηση γίνεται πιο σύνθετη, όπως είδαμε στο Κεφάλαιο 6. Μία πιο εκτεταμένη μελέτη της εκτίμησης των 3-Δ παραμέτρων κίνησης και γεωμετρίας μιας επίπεδης επιφάνειας σε συνδυασμό με τις δυνατότητες που παρουσιάζονται στο δισδιάστατο χώρο της εικόνας, καθώς και η αναλυτική μελέτη των επιπτώσεων του θορύβου στο σφάλμα θα μπορούσε να συμβάλει στη βελτίωση αυτών των αποτελεσμάτων.

Η έρευνα και η υλοποίηση της εργασίας αυτής προϋποθέτει τα σημεία που παρακολουθούνται να βρίσκονται πάντα εντός πλαισίου-εικόνας και να είναι ορατά. Πολλές φορές όμως συμβαίνει το πρόσωπο να βγαίνει εκτός πλαισίου ή να επικαλύπτεται μερικώς ή ολόκληρο από κάποιο άλλο αντικείμενο. Ένας ακόμα στόχος είναι, λοιπόν, να συνυπολογίζονται τέτοιου είδους παράγοντες εξαφάνισης και επικάλυψης ώστε η μέθοδος να μην αποτυγχάνει και να πρέπει να ξανα-αρχικοποιείται είτε με παρέμβαση του χρήστη, είτε αυτόματα. Ακόμα, η παρακολούθηση αφορά ένα και μόνο πρόσωπο. Μπορεί, λοιπόν, να επεκταθεί ώστε να περιλαμβάνει περισσότερα πρόσωπα κρατώντας

πληροφορίες για αντίστοιχο αριθμό επιπέδων και σημείων.

Μεγάλο ενδιαφέρον παρουσιάζουν εφαρμογές τηλε-συνδιασκέψεων και βίντεο-τηλεφώνων, που ήδη μπαίνουν στην καθημερινότητά μας. Οι εφαρμογές αυτές, προς το παρόν, καταφέρνουν πιο πολύ να συγχύζουν τους χρήστες-συνομιλητές καθώς υπάρχει η δυσκολία να αντιλαμβανόμαστε ποιος μιλάει σε ποιον και πού έχει κάποιος στραμμένη στην προσοχή του. Η λύση που έχει γίνει αποδεκτή μέχρι στιγμής είναι να αντικαθίστανται τα πραγματικά πρόσωπα με συνθετικά και να επανατοποθετούνται στις οθόνες των συνομιλητών. Εφόσον η μέθοδος αυτή υιοθετεί την ανάκτηση της 3-Δ κίνησης και γεωμετρίας θα μπορούσε να χρησιμοποιηθεί σε τέτοιου είδους περιβάλλοντα. Η μέθοδος με την ήδη υπάρχουσα υποδομή μπορεί να επεκταθεί για να παράγει συνθετικά πρόσωπα, π.χ. τοποθετώντας τη 2-Δ υφή της περιοχής του προσώπου, όπως προκύπτει από την παρακολούθηση των σημείων, πάνω στην επίπεδη επιφάνεια.

Μία πιο προχωρημένη επέκταση θα μπορούσε να είναι η κινητοποίηση της κάμερας από την ίδια τη μέθοδο, ώστε να έχει πάντα στο επίκεντρο το πρόσωπο, που παρακολουθείται. Υποστηρίζοντας ένα περιβάλλον διεπαφής με μικρά μοτέρ, τοποθετημένα στην κάμερα, θα μπορούσε να δίνει οδηγίες προς την κάμερα, έτσι ώστε όχι μόνο να μη χάνεται το πρόσωπο ποτέ από την εικόνα, αλλά να βρίσκεται πάντα στο “κέντρο” της.

Τέλος, η εργασία αυτή αποτελεί ένα εξαιρετικό θεμέλιο για να προχωρήσουμε σε ανάλυση έκφρασης, παρακολούθηση βλέμματος και αναγνώριση ομιλίας. Παρακολουθώντας τα συγκεκριμένα σημεία είμαστε σε θέση να γνωρίζουμε ανά πάσα στιγμή που βρίσκονται τα μάτια, η μύτη και το στόμα, οπότε μία παραπέρα επεξεργασία της πληροφορίας αυτής μπορεί να οδηγήσει σε μία ολοκληρωμένη εφαρμογή σχετικά με το πρόσωπο. Μία εφαρμογή που θα βρει πρόσφορο έδαφος κυρίως στους ίδιους τους υπολογιστές εφοδιάζοντάς τους με κάποιο είδος αίσθησης και αντίληψης.

Παράρτημα Α

Singular Value Decomposition

Η Singular Value Decomposition ενός $M \times N$ πίνακα A , με $M \geq N$, ορίζεται ως η ανάλυση :

$$A = U W V^T$$

όπου ο πίνακας U είναι διαστάσεων $M \times N$ και ορθογώνιος ως προς τις στήλες ($U^T U = I$), ο πίνακας W είναι τετραγωνικός $N \times N$ και διαγώνιος, και ο πίνακας V είναι τετραγωνικός $N \times N$ και ορθογώνιος ($V V^T = V^T V = I$). Τα στοιχεία $w_j, j = 1, \dots, N$ που βρίσκονται στην κύρια διαγώνιο του W ονομάζονται singular values και ισχύει $w_1 \geq w_2 \geq \dots \geq w_N \geq 0$. Αν $M < N$ τα singular values $w_j, j = M + 1, \dots, N$ είναι όλα μηδέν και οι αντίστοιχες στήλες του U είναι επίσης μηδενικές.

Όταν ο A είναι ένας τετραγωνικός $N \times N$ πίνακας, οι πίνακες U, W και V είναι όλοι τετραγωνικοί $N \times N$ και μάλιστα ισχύει $U^{-1} = U^T$ και $V^{-1} = V^T$. Ο πίνακας W είναι διαγώνιος και ο αντίστροφός του W^{-1} περιέχει στοιχεία στη διαγώνιό του που είναι αντίστοιχα αντίστροφα των στοιχείων του W . Δηλαδή :

$$\begin{aligned} A^{-1} &= (U W V^T)^{-1} \Leftrightarrow \\ A^{-1} &= (V^T)^{-1} W^{-1} U^{-1} \Leftrightarrow \\ A^{-1} &= (V^{-1})^{-1} W^{-1} U^T \Leftrightarrow \\ A^{-1} &= V \left[\text{diag} \left(\frac{1}{w_j} \right) \right] U^T \end{aligned}$$

Έστω το γραμμικό σύστημα $Ax = b$. Η Singular Value Decomposition κατασκευάζει ορθοκανονικές βάσεις για το μηδενόχωρο (nullspace) και τον υπόχωρο του b στον οποίο μπορεί να εκταθεί ο A (range(A)), δηλαδή υπάρχουν κάποια x που μπορεί να αντιστοιχιστούν στον υπόχωρο αυτό. Συγκεκριμένα, οι στήλες του U , των οποίων τα αντίστοιχα w_j είναι μη-μηδενικά, αποτελούν ένα ορθοκανονικό σύνολο βασικών διανυσμάτων που καλύπτουν τον υπόχωρο αυτό. Επίσης, οι στήλες του V , των οποίων

τα αντίστοιχα w_j είναι μηδενικά, αποτελούν μία ορθοκανονική βάση του μηδενόχωρου του πίνακα A .

Η Singular Value Decomposition προσφέρεται για να λυθεί το γραμμικό σύστημα $Ax = b$. Αν $b = 0$, η λύση αποτελείται από οποιαδήποτε στήλη του V της οποίας το αντίστοιχο $w_j = 0$. Αν $b \neq 0$ ξεχωρίζουμε δύο περιπτώσεις. (α) Αν το b αποτελεί υπόχωρο του A , τότε υπάρχει λύση x : αντικαθιστούμε $\frac{1}{w_j}$ με 0, αν $w_j = 0$, και υπολογίζουμε από δεξιά προς αριστερά

$$x = V \left[\text{diag} \left(\frac{1}{w_j} \right) \right] (U^T b) \quad (\text{A.1})$$

Αυτό αποτελεί το διάνυσμα λύσης με το μικρότερο $|x|^2$. (β) Αν το b δεν αποτελεί υπόχωρο του A , τότε δεν υπάρχει λύση, αλλά η εξίσωση (A.1) μπορεί να χρησιμοποιηθεί για να κατασκευαστεί μία λύση τέτοια ώστε το σφάλμα $|Ax - b|$ να είναι το ελάχιστο δυνατό.

Παράρτημα Β

Αποδείξεις Θεωρημάτων

Παρακάτω παραθέτονται οι αποδείξεις των θεωρημάτων του Κεφαλαίου 4. Προκύπτουν από τη βιβλιογραφία [85] και βοηθάνε τον αναγνώστη να συνειδητοποιήσει τη λογική του όλου προβλήματος.

Β.1 Θεώρημα 1

Αν η πολλαπλότητα των singular values του πίνακα A είναι δύο, π.χ. $\lambda_1 = \lambda_2 \neq \lambda_3$, τότε η λύση για τις παραμέτρους κίνησης και γεωμετρίας (προσανατολισμός) είναι μοναδική, εκτός ενός κοινού παράγοντα κλίμακας για τις παραμέτρους μεταφοράς και γεωμετρίας, και

$$R = \lambda_1^{-1} A - \left(\frac{\lambda_3}{\lambda_1} - s \right) U_3 V_3^T,$$

$$\begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} = w^{-1} \left(\frac{\lambda_3}{\lambda_1} - s \right) U_3 \text{ και}$$

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = w V_3$$

όπου

$$s = \det(U)\det(V),$$

w ένας παράγοντας κλίμακας,

a, b και c οι παράμετροι της (4.8) (εξίσωση επιπέδου τη χρονική στιγμή t_1),

U_3 και V_3 οι τρίτες στήλες των πινάκων U και V αντίστοιχα.

(B.1)

Επιπλέον, μία αναγκαία και επαρκής συνθήκη, ώστε η πολλαπλότητα των singular values του πίνακα A να είναι δύο, είναι ότι η κίνηση μπορεί να πραγματοποιηθεί περιστρέφοντας το αντικείμενο γύρω από την αρχή των αξόνων και μετά μεταφέροντάς το διαμήκως του κάθετου διανύσματος της επίπεδης επιφάνειας (N).

Απόδειξη : Τα δύο μέλη της (4.36) μπορούν να εξισωθούν με μία συλλογή θετικών τιμών που αντιστοιχούν στα πεδία τιμών των x, y, z και x', y', z' . Έστω d μία τέτοια τιμή. Τότε έχουμε :

$$d = x_n'^2 + y_n'^2 + z_n'^2 \quad (\text{B.2})$$

και

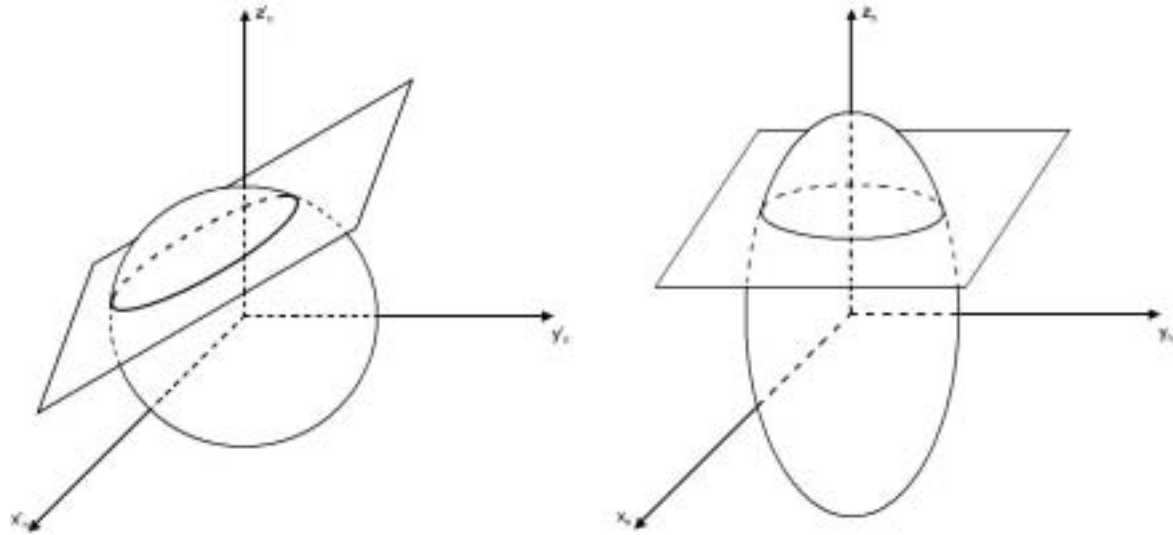
$$d = k^2 (\lambda_1^2 x_n^2 + \lambda_2^2 y_n^2 + \lambda_3^2 z_n^2) \quad (\text{B.3})$$

Η (B.2) ορίζει μία σφαίρα στον 3-Δ χώρο (x_n', y_n', z_n') , ενώ η (B.3) ορίζει ένα ελλειψοειδές στον 3-Δ χώρο (x_n, y_n, z_n) . Εφόσον $\lambda_1 = \lambda_2$, δύο από τους τρεις κύριους άξονες του ελλειψοειδούς είναι ίσοι. Αφού η επιφάνεια του αντικειμένου είναι επίπεδη, η συλλογή των σημείων της επιφάνειας του αντικειμένου, η οποία ικανοποιεί την (B.2), πρέπει να είναι ο κύκλος που προκύπτει από την τομή της σφαίρας με το αντικείμενο (δηλαδή το επίπεδο) τη χρονική στιγμή t_2 (βλ. και σχήμα B.1). Λόγω των (4.36), (B.2) και (B.3), όλα τα σημεία πάνω σ' αυτόν τον κύκλο τη χρονική στιγμή t_2 πρέπει να ικανοποιούν και την (B.3) τη χρονική στιγμή t_1 , δηλαδή αυτά πρέπει να βρίσκονται στην τομή της επιφάνειας του αντικειμένου (επίπεδο) και του ελλειψοειδούς. Λόγω του ότι περιοριστήκαμε σε στερεά αντικείμενα, αυτή η τομή πρέπει να είναι κι αυτή ένας κύκλος. Όμως, η μοναδική δυνατότητα να κόψουμε έναν κύκλο από ένα ελλειψοειδές, που έχει δύο από τους τρεις κύριους άξονες ίσους, είναι το επίπεδο να βρίσκεται κάθετα στον μεγαλύτερο άξονα του ελλειψοειδούς, όπως απεικονίζεται και στο σχήμα B.2. Αυτό, όμως, σημαίνει ότι υπάρχει μόνο ένας δυνατός προσανατολισμός για την επιφάνεια του αντικειμένου πριν εκτελέσει την κίνηση. Αυτό θα μας οδηγήσει παρακάτω στο συμπέρασμα ότι οι παράμετροι κίνησης είναι μοναδικοί.

Σημειώνουμε ότι το λ_1 (και το λ_2 αντίστοιχα) δεν μπορεί ποτέ να είναι μηδέν, γιατί, αν ίσχυε κάτι τέτοιο, το ελλειψοειδές που ορίζεται από την (B.3) θα εκφυλιζόταν σε δύο παράλληλα επίπεδα, και δεν υπάρχει τρόπος η επιφάνεια του αντικειμένου μας να τέμνει δύο παράλληλα επίπεδα ώστε να προκύψει κύκλος.

Όπως παρουσιάζουμε και στο σχήμα B.2, η επιφάνεια του αντικειμένου πρέπει να είναι κάθετη στον άξονα z_n . Εφόσον, όμως, ο άξονας z_n προέρχεται περιστρέφοντας τον άξονα z με τον ορθοκανονικό πίνακα V (όπως απεικονίζεται στο σχήμα 4.2), έχουμε :

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = V \begin{bmatrix} 0 \\ 0 \\ w \end{bmatrix} = wV_3 \quad (\text{B.4})$$



Σχήμα B.1: Επίπεδο που τέμνει τη σφαίρα στον 3-Δ χώρο (x'_n, y'_n, z'_n) τη χρονική στιγμή t_2 (αριστερά) και το ελλειψοειδές στον 3-Δ χώρο (x_n, y_n, z_n) τη χρονική στιγμή t_1 (δεξιά).

όπου a, b και c είναι οι παράμετροι της (4.8), V_3 είναι η τρίτη στήλη του πίνακα V στην (4.26), και w είναι μία αυθαίρετη σταθερά.

Αντικαθιστώντας τις (B.4) και (4.26) στην (4.28) παίρνουμε :

$$R + \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} \begin{bmatrix} 0 & 0 & w \end{bmatrix} V^T = k U \Lambda V^T \quad (\text{B.5})$$

Πολλαπλασιάζοντας την (B.5) από μπροστά με U^T και από πίσω με V , έχουμε :

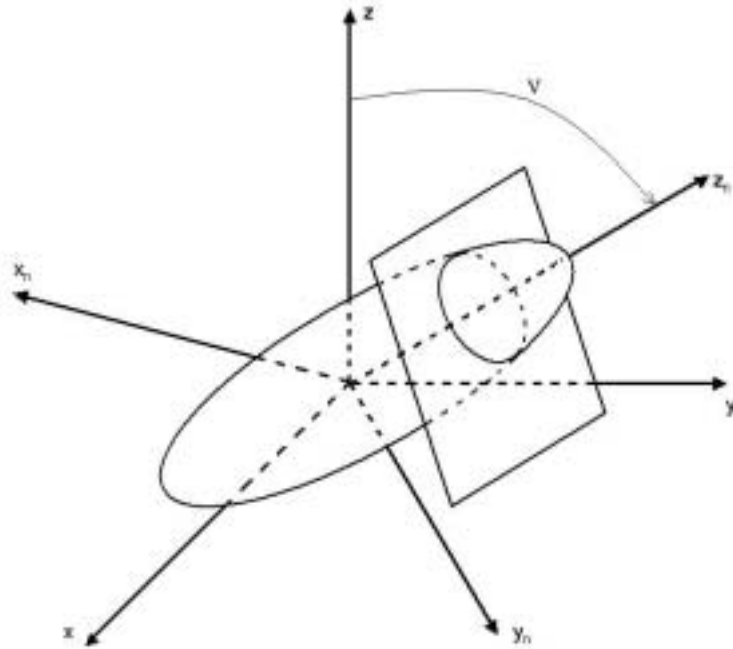
$$U^T R V + U^T \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} \begin{bmatrix} 0 & 0 & w \end{bmatrix} V^T V = k \Lambda \Leftrightarrow \quad (\text{B.6})$$

$$R' + \begin{bmatrix} \Delta x' \\ \Delta y' \\ \Delta z' \end{bmatrix} \begin{bmatrix} 0 & 0 & w \end{bmatrix} = k \begin{bmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \lambda_3 \end{bmatrix} \quad (\text{B.7})$$

όπου :

$$R' = U^T R V \quad (\text{B.8})$$

$$\begin{bmatrix} \Delta x' \\ \Delta y' \\ \Delta z' \end{bmatrix} = U^T \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} \quad (\text{B.9})$$



Σχήμα Β.2: Επίπεδο που τέμνει το ελλειψοειδές στον 3-Δ χώρο (x_n, y_n, z_n) τη χρονική στιγμή t_1 και η σχέση του χώρου (x_n, y_n, z_n) με το χώρο (x, y, z) .

Η εξίσωση (B.7) δίνει :

$$R' = \begin{bmatrix} k\lambda_1 & 0 & -w\Delta x' \\ 0 & k\lambda_1 & -w\Delta y' \\ 0 & 0 & k\lambda_3 - w\Delta z' \end{bmatrix} \quad (\text{B.10})$$

Θα δείξουμε στη συνέχεια ότι τα $\Delta x'$ και $\Delta y'$ στην (B.10) είναι μηδέν, οπότε ο R' είναι διαγώνιος. Εφόσον οι U , V και R στην (B.8) είναι όλοι ορθοκανονικοί πίνακες, τότε και ο R' είναι ορθοκανονικός. Παίρνοντας το εσωτερικό γινόμενο της δεύτερης και τρίτης στήλης του R' και εξισώνοντάς το με μηδέν έχουμε :

$$k\lambda_1 w \Delta y' = 0 \quad (\text{B.11})$$

Τα λ_1 και k δεν μπορεί να είναι μηδέν, γιατί, αν συνέβαινε κάτι τέτοιο, η πρώτη και η δεύτερη στήλη του R' (βλ. (B.10)) θα ήταν μηδενικές, που διαψεύδει το γεγονός ότι ο R' είναι ορθοκανονικός. Προφανώς, ο παράγοντας w δεν μπορεί να είναι κι αυτός μηδέν, γιατί έτσι θα έπρεπε και τα a , b και c να είναι μηδέν, πράγμα που αντιτίθεται στην (4.8). Οπότε, από την (B.11) γίνεται φανερό ότι :

$$\Delta y' = 0 \quad (\text{B.12})$$

Με παρόμοιο τρόπο αποδεικνύεται ότι

$$\Delta x' = 0 \quad (\text{B.13})$$

Οπότε, ο R' είναι διαγώνιος και επειδή είναι και ορθοκανονικός η (B.10), σε συνδυασμό με τις (B.12) και (B.13), δίνει :

$$k\lambda_1 = +1 \text{ ή } -1 \quad (\text{B.14})$$

$$k\lambda_3 - w\Delta z' = +1 \text{ ή } -1 \quad (\text{B.15})$$

Θα δείξουμε τώρα ότι το k είναι θετικό. Από την (4.30) έχουμε :

$$z' = k(a_7x + a_8y + z) \quad (\text{B.16})$$

Επειδή $x = 0$ και $y = 0$ έχουμε $z' = kz$. Επίσης, επειδή το αντικείμενο βρίσκεται μπροστά στην κάμερα, τα z και z' είναι και το δύο θετικά, οπότε και το k είναι θετικό.

Αφού το λ_1 είναι εξ' ορισμού θετικό¹, το δεξί μέλος της (B.14) δεν μπορεί να είναι -1 . Οπότε,

$$k = \frac{1}{\lambda_1} \quad (\text{B.17})$$

και

$$R' = \begin{bmatrix} 1 & & \\ & 1 & \\ & & s \end{bmatrix} \quad (\text{B.18})$$

όπου s είναι είτε $+1$ είτε -1 .

Από την (B.8) ξέρουμε ότι $R' = U^T R V$, άρα $\det(R') = \det(U)\det(R)\det(V)$ και επειδή ο R είναι ορθοκανονικός και συμμετρικός τέτοιος ώστε $\det(R) = 1$ μπορούμε να υπολογίσουμε το s :

$$s = \det(U)\det(V) \quad (\text{B.19})$$

Η (B.15) δίνει

$$\Delta z' = w^{-1} \left(\frac{\lambda_3}{\lambda_1} - s \right) \quad (\text{B.20})$$

Από τις (B.9), (B.20) και λόγω του γεγονότος ότι $\Delta x' = \Delta y' = 0$, έχουμε :

¹Για κάθε singular value λ_i ισχύει $\lambda_i \geq 0$.

$$\begin{aligned}
\begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} &= U \begin{bmatrix} \Delta x' \\ \Delta y' \\ \Delta z' \end{bmatrix} \Leftrightarrow \\
\begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} &= U \begin{bmatrix} 0 \\ 0 \\ w^{-1} \left(\frac{\lambda_3}{\lambda_1} - s \right) U_3 \end{bmatrix} \Leftrightarrow \\
\begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} &= w^{-1} \left(\frac{\lambda_3}{\lambda_1} - s \right) U_3 \tag{B.21}
\end{aligned}$$

Οι εξισώσεις (B.4), (B.5) και (B.21) δίνουν :

$$\begin{aligned}
R &= \lambda_1^{-1} A - \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} \begin{bmatrix} a & b & c \end{bmatrix} \Leftrightarrow \\
R &= \lambda_1^{-1} A - \left(\frac{\lambda_3}{\lambda_1} - s \right) U_3 V_3^T \tag{B.22}
\end{aligned}$$

Στη συνέχεια θα αποδείξουμε ότι η λύση για τον πίνακα περιστροφής R (βλ. και (B.22)) είναι μοναδική και επίσης ότι οι λύσεις, πέραν ενός παράγοντα κλίμακας, των παραμέτρων μεταφοράς Δx , Δy και Δz (βλ. και (B.21)) και των γεωμετρικών παραμέτρων a , b και c (βλ. και (B.4)), που απαρτίζουν το διάνυσμα προσανατολισμού του επιπέδου $N = (a, b, c)$, είναι μοναδικές.

Το πρώτο που θα δείξουμε είναι ότι, εφόσον δίδεται ο A , η στήλη U_3 είναι αμετάβλητη, εκτός του προσήμου. Από την (4.26) έχουμε :

$$AA^T U_3 = \lambda_3^2 U_3 \tag{B.23}$$

Έστω Q κάποιος ορθοκανονικός πίνακας ιδιοδιανυσμάτων του AA^T . Τότε :

$$AA^T = Q^T \begin{bmatrix} \lambda_1^2 & & \\ & \lambda_2^2 & \\ & & \lambda_3^2 \end{bmatrix} Q \tag{B.24}$$

Από τις (B.23) και (B.24) έχουμε :

$$\left\{ Q^T \begin{bmatrix} \lambda_1^2 & & \\ & \lambda_2^2 & \\ & & \lambda_3^2 \end{bmatrix} Q - \lambda_3^2 I \right\} U_3 = 0$$

ή

$$PU_3 = 0$$

όπου

$$P \triangleq AA^T - \lambda_3^2 I = Q^T \begin{bmatrix} \lambda_1^2 - \lambda_3^2 & & \\ & \lambda_1^2 - \lambda_3^2 & \\ & & 0 \end{bmatrix} Q \quad (\text{B.25})$$

Ο πίνακας P είναι τάξεως 2, αφού το $\lambda_1^2 - \lambda_3^2$ στη διαγώνιο του διαγώνιου πίνακα της (B.25) είναι μη μηδενικό. Ακόμα, ο πίνακας P είναι σταθερός, δεδομένου του A , αφού $P = AA^T - \lambda_3^2 I$. Συμπεραίνουμε, λοιπόν, ότι και η στήλη U_3 είναι σταθερή, πέραν του προσήμου βέβαια.

Θα αποδείξουμε τώρα ότι η (B.22) είναι μοναδική, δηλαδή η

$$\left(\frac{\lambda_3}{\lambda_1} - s\right) U_3 V_3^T = \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} \begin{bmatrix} a & b & c \end{bmatrix} \quad (\text{B.26})$$

είναι μοναδική δεδομένου του A . Μπορούμε να διακρίνουμε δύο περιπτώσεις. Η πρώτη είναι όταν $\lambda_3 \neq 0$. Στην περίπτωση αυτή, $s = \text{sgn}(\det(A))$, εφόσον $A = U\Lambda V^T$ οπότε $s = \det(U)\det(V) = (\lambda_1\lambda_2\lambda_3)^{-1} \det(A) = \text{sgn}(\det(A))$. Δεδομένου, λοιπόν, του A , το s είναι σταθερό. Μένει να δειχθεί ότι το $U_3 V_3^T$ στην (B.22) είναι μοναδικό.

Εφόσον, οι στήλες U_3 και V_3 είναι σταθερές, πέραν του προσήμου, όπως αποδείξαμε παραπάνω, αρκεί να δείξουμε ότι αν η V_3 αλλάξει πρόσημο, θα αλλάξει κι η U_3 . Από την (4.26) έχουμε :

$$AV = U\Lambda = \begin{bmatrix} \lambda_1 U_1 & \lambda_2 U_2 & \lambda_3 U_3 \end{bmatrix}$$

συνεπώς

$$AV_3 = \lambda_3 U_3$$

Επειδή τα A και λ_3 είναι σταθερά, δεδομένων δύο προοπτικών όψεων, είναι φανερό ότι εάν η στήλη V_3 αλλάξει πρόσημο, αλλάζει κι η U_3 . Επομένως, το $U_3 V_3^T$ έχει σταθερό πρόσημο. Αποδείξαμε, λοιπόν, ότι το γινόμενο

$$\begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} \begin{bmatrix} a & b & c \end{bmatrix}$$

Γεώργιος Σημαντήρης

είναι μοναδικό, επομένως και ο R στην (B.22) είναι μοναδικός.

Στη δεύτερη περίπτωση, όπου $\lambda_3 = 0$, έχουμε από την (B.22) ότι

$$\begin{aligned} R &= \lambda_1^{-1} A + sU_3V_3^T \Leftrightarrow \\ R &= \lambda_1^{-1} A + \det(U)U_3\det(V)V_3^T \end{aligned} \quad (\text{B.27})$$

Όταν η U_3 αλλάξει πρόσημο, αλλάζει κι η $\det(U)$. Επομένως το πρόσημο της $\det(U)U_3$ στην (B.27) παραμένει σταθερό. Παρομοίως, το πρόσημο της $\det(V)V_3^T$ παραμένει σταθερό, όταν η V_3 αλλάξει πρόσημο. Η μοναδικότητα, λοιπόν, της (B.27) δεν κλονίζεται από την αμφιβολία για τα πρόσημα των U_3 και V_3 .

Το $A^T A$ έχει διπλές ιδιοτιμές, οπότε τα ιδιοδιανύσματα V_1 και V_2 , που αντιστοιχούν στις πολλαπλές ιδιοτιμές $\lambda_1 (= \lambda_2)$, είναι ορθοκανονικά μεταξύ τους, αλλά μπορεί να βρίσκονται οπουδήποτε σε κάποιο σταθερό επίπεδο κάθετο στο V_3 (σημειώνεται ότι εδώ ερμηνεύουμε τα ιδιοδιανύσματα γεωμετρικά, σαν κάποια διανύσματα στον 3-Δ χώρο). Αν ανταλλάγουν οι τάξεις (order) των V_1 και V_2 , που βρίσκονται πάνω στο επίπεδο, κρατώντας όμως τη V_3 σταθερή, θα αλλάξει το πρόσημο της $\det(V)$. Πρέπει να αποδείξουμε, λοιπόν, ότι αν συμβεί αυτό, θα αλλάξει και το πρόσημο της $\det(U)$, που σημαίνει όμως ότι η (B.27) παραμένει σταθερή. Από την (4.26) είναι φανερό ότι

$$\begin{aligned} (\lambda_1^{-1} A) V_1 &= U_1 \\ (\lambda_1^{-1} A) V_2 &= U_2 \end{aligned}$$

Επειδή τα λ_1 και A είναι σταθερά, όταν ανταλλάγουν τα V_1 και V_2 , θα ανταλλάγουν και τα U_1 και U_2 . Άρα, όταν η $\det(V)$ αλλάξει πρόσημο, θα αλλάξει και η $\det(U)$. Και γι' αυτήν την περίπτωση, όταν δηλαδή $\lambda_3 = 0$, το γινόμενο

$$\begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} \begin{bmatrix} a & b & c \end{bmatrix}$$

είναι σταθερό, όπως επίσης και ο R στην (B.22).

θα αποδείξουμε ότι μία αναγκαία συνθήκη για να ισχύει $\lambda_1 = \lambda_2 \neq \lambda_3$ είναι ότι το διάνυσμα μεταφοράς

$$\begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix}$$

είναι παράλληλο στο κάθετο στο επίπεδο διάνυσμα N τη χρονική στιγμή t_2 , δηλαδή παράλληλο στην κατεύθυνση του επιπέδου τη χρονική στιγμή t_2 .

Πριν εκτελεστεί η κίνηση, δηλαδή τη χρονική στιγμή t_1 , το N :

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

είναι κάθετο στο επίπεδο. Για να βρούμε το N τη χρονική στιγμή t_2 αρκεί να περιστρέψουμε το παραπάνω (το N τη χρονική στιγμή t_1) με τον πίνακα R :

$$R \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

Αρκεί απλά να αποδειχθεί ότι υπάρχει q , τέτοιο ώστε

$$\begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} = qR \begin{bmatrix} a \\ b \\ c \end{bmatrix} \quad (\text{B.28})$$

Από τις (B.4), (B.21), (B.22), (B.28) έχουμε

$$\begin{aligned} w^{-1} \left(\frac{\lambda_3}{\lambda_1} - s \right) U_3 &= q \left[\lambda_1^{-1} A - \left(\frac{\lambda_3}{\lambda_1} - s \right) U_3 V_3^T \right] w V_3 \Leftrightarrow \\ w^{-1} \left(\frac{\lambda_3}{\lambda_1} - s \right) U_3 &= qw \lambda_1^{-1} A V_3 - qw \left(\frac{\lambda_3}{\lambda_1} - s \right) U_3 V_3^T V_3 \end{aligned} \quad (\text{B.29})$$

Όμως,

$$\begin{aligned} A &= \lambda_1 U_1 V_1^T + \lambda_2 U_2 V_2^T + \lambda_3 U_3 V_3^T \Leftrightarrow \\ A V_3 &= \lambda_1 U_1 V_1^T V_3 + \lambda_2 U_2 V_2^T V_3 + \lambda_3 U_3 V_3^T V_3 \Leftrightarrow \\ A V_3 &= \mathbf{0} + \mathbf{0} + \lambda_3 U_3 \Leftrightarrow \\ A V_3 &= \lambda_3 U_3 \end{aligned} \quad (\text{B.30})$$

Αντικαθιστώντας την (B.30) στην (B.29) παίρνουμε

$$w^{-1} \left(\frac{\lambda_3}{\lambda_1} - s \right) U_3 = qw \left(\frac{\lambda_3}{\lambda_1} - \frac{\lambda_3}{\lambda_1} + s \right) U_3 = qws U_3 \quad (\text{B.31})$$

Θέτοντας, λοιπόν,

$$q = w^{-2} s^{-1} \left(\frac{\lambda_3}{\lambda_1} - s \right) \quad (\text{B.32})$$

η (B.28) ικανοποιείται.

Αποδείξαμε, λοιπόν, ότι η αναγκαία συνθήκη για να ισχύει $\lambda_1 = \lambda_2 \neq \lambda_3$ είναι ότι η κίνηση μπορεί να πραγματοποιηθεί περιστρέφοντας πρώτα το αντικείμενο γύρω από έναν άξονα, που περνάει από την αρχή των αξόνων, και μετά μεταφέροντάς το διαμηκώς

Γεώργιος Σημαντήρης

του διανύσματος N , που είναι κάθετο στο επίπεδο και φανερώνει την κατεύθυνση προσανατολισμού του επιπέδου, τη χρονική στιγμή t_2 .

Στο επόμενο θεώρημα (βλ. Β.2) θα αποδείξουμε ότι αν η μεταφορά γίνει διαμήκως του διανύσματος N (που είναι κάθετο στο επίπεδο και φανερώνει την κατεύθυνση προσανατολισμού του επιπέδου τη χρονική στιγμή t_2), τότε τα singular values του A δεν μπορούν να είναι όλα διαφορετικά.

Β.2 Θεώρημα 2

Αν τα singular values του A είναι όλα διαφορετικά, π.χ. $\lambda_1 > \lambda_2 > \lambda_3$, τότε υπάρχουν ακριβώς δύο λύσεις για τις παραμέτρους γεωμετρίας και κίνησης, εκτός ενός παράγοντα κλίμακας για τις παραμέτρους μεταφοράς και γεωμετρίας, και

$$R = U \begin{bmatrix} \alpha & 0 & \beta \\ 0 & 1 & 0 \\ -s\beta & 0 & s\alpha \end{bmatrix} V^T$$

$$\begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} = w^{-1} \left[-\beta U_1 + \left(\frac{\lambda_3}{\lambda_2} - s\alpha \right) U_3 \right]$$

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = w [\delta V_1 + V_3]$$

όπου

$$\delta = \pm \left(\frac{\lambda_1^2 - \lambda_2^2}{\lambda_2^2 - \lambda_3^2} \right)^{\frac{1}{2}}$$

$$\alpha = \frac{\lambda_1 + s\lambda_3\delta^2}{\lambda_2(1 + \delta^2)}$$

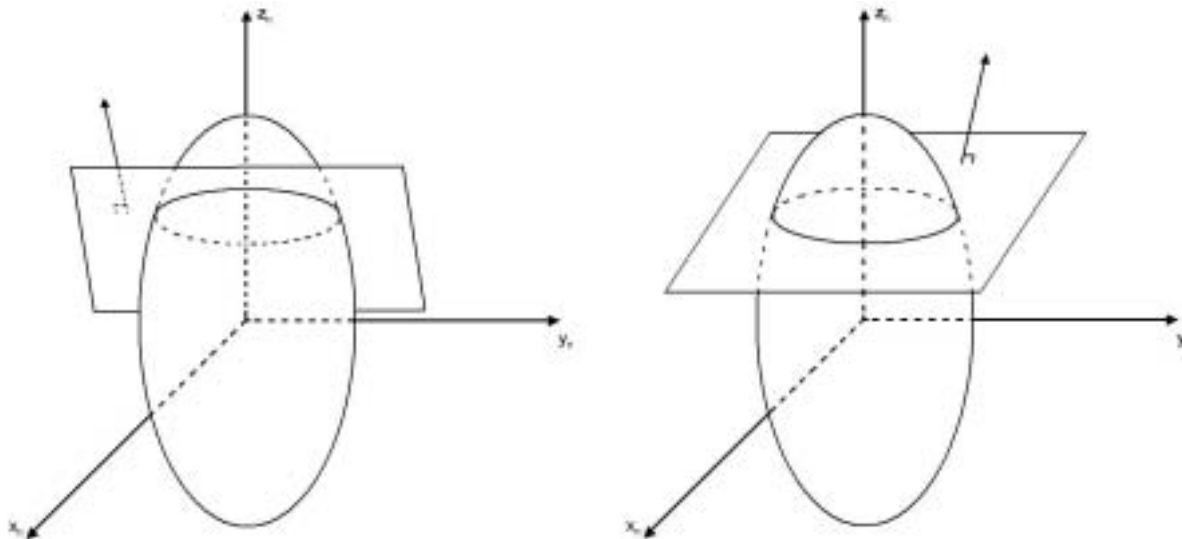
$$\beta = \pm \sqrt{1 - \alpha^2}$$

$$s = \det(U)\det(V)$$

Σε κάθε μία από τις δύο λύσεις, $\text{sgn}(\beta) = -\text{sgn}(\delta)$.

Επιπλέον, μία αναγκαία και επαρκής συνθήκη, ώστε ο πίνακας A να έχει διαφορετικά singular values, είναι ότι η κίνηση μπορεί να πραγματοποιηθεί περιστρέφοντας το αντικείμενο γύρω από την αρχή των αξόνων και μετά μεταφέροντάς το διαμήκως μίας κατεύθυνσης διαφορετικής του κάθετου διανύσματος της επίπεδης επιφάνειας (N) τη χρονική στιγμή t_2 .

Απόδειξη : Εφόσον τα τρία singular values είναι όλα διαφορετικά, οι τρεις κύριοι άξονες του ελλειψοειδούς, που ορίζεται στην (B.3) έχουν διαφορετικά μήκη. Χρησιμοποιώντας το ίδιο επιχείρημα όπως στο θεώρημα 1 (βλ. B.1), η επιφάνεια του αντικειμένου πρέπει να είναι προσανατολισμένη με τέτοιο τρόπο, ώστε η τομή της με το ελλειψοειδές να δίνει κύκλο. Χρησιμοποιώντας βασική αναλυτική γεωμετρία, μπορούμε εύκολα να διαπιστώσουμε ότι υπάρχουν δύο τέτοιοι προσανατολισμοί (βλ. και σχήμα B.3).



Σχήμα B.3: Επίπεδο που τέμνει το ελλειψοειδές στον 3-Δ χώρο (x_n, y_n, z_n) τη χρονική στιγμή t_1 με δύο τρόπους, ώστε να προκύψει κύκλος.

Επειδή $\lambda_1 > \lambda_2 > \lambda_3$, ο μεγαλύτερος κύριος άξονας συμπίπτει με τον άξονα z_n , και το διάνυσμα κάθετο στην επιφάνεια του αντικειμένου είναι

$$w \begin{bmatrix} \delta \\ 0 \\ 1 \end{bmatrix}, \quad \text{όπου} \quad \delta = \pm \left(\frac{\lambda_1^2 - \lambda_2^2}{\lambda_2^2 - \lambda_3^2} \right)^{\frac{1}{2}}$$

στο σύστημα συντεταγμένων (x_n, y_n, z_n) και w είναι κάποια σταθερά. Όπως μπορούμε να δούμε στο σχήμα 4.2, οι (x_n, y_n, z_n) άξονες προκύπτουν από τους (x, y, z) άξονες περιστρέφοντας τους με τον ορθοκανονικό πίνακα V , οπότε

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = wV \begin{bmatrix} \delta \\ 0 \\ 1 \end{bmatrix} \quad (\text{B.33})$$

όπου w κάποια σταθερά. Αντικαθιστώντας την (B.33) στην (4.28) έχουμε :

$$kA = kU\Lambda V^T = R + w \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} \begin{bmatrix} \delta & 0 & 1 \end{bmatrix} V^T \quad (\text{B.34})$$

Πολλαπλασιάζοντας την (B.34) από μπροστά με U^T και από πίσω με V^T έχουμε :

$$k\Lambda = U^T R V + w U^T \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} \begin{bmatrix} \delta & 0 & 1 \end{bmatrix} V^T V$$

Οπότε

$$R' = k\Lambda - \begin{bmatrix} \Delta x' \\ \Delta y' \\ \Delta z' \end{bmatrix} \begin{bmatrix} \delta & 0 & 1 \end{bmatrix} \quad (\text{B.35})$$

όπου

$$R' \triangleq U^T R V \quad (\text{B.36})$$

$$\begin{bmatrix} \Delta x' \\ \Delta y' \\ \Delta z' \end{bmatrix} \triangleq w U^T \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} \quad (\text{B.37})$$

Από την (B.35) γίνεται φανερό ότι

$$R' = \begin{bmatrix} k\lambda_1 - \delta\Delta x' & 0 & -\Delta x' \\ -\delta\Delta y' & k\lambda_2 & -\Delta y' \\ -\delta\Delta z' & 0 & k\lambda_3 - \Delta z' \end{bmatrix} \quad (\text{B.38})$$

Αφού οι U , R και V είναι ορθοκανονικοί πίνακες, από την (B.38) συμπεραίνουμε ότι ο R' είναι επίσης ορθοκανονικός. Παίρνοντας το εσωτερικό γινόμενο των στηλών 2 και 3 και εξισώνοντάς το με μηδέν έχουμε

$$k\lambda_2\Delta y' = 0 \quad (\text{B.39})$$

Εφόσον $\lambda_1 > \lambda_2 > \lambda_3 \geq 0 \Rightarrow \lambda_2 > 0$. Άρα η (B.39) υποδηλώνει ότι $\Delta y' = 0$ και η (B.38) γίνεται :

$$R' = \begin{bmatrix} k\lambda_1 - \delta\Delta x' & 0 & -\Delta x' \\ 0 & k\lambda_2 & 0 \\ -\delta\Delta z' & 0 & k\lambda_3 - \Delta z' \end{bmatrix} \quad (\text{B.40})$$

Λόγω της κανονικότητας της στήλης 2 έχουμε $k\lambda_2 = \pm 1$. Αλλά, επειδή $k > 0$ και $\lambda_2 > 0$ συνεπάγουμε $k\lambda_2 = 1 \Rightarrow k = \frac{1}{\lambda_2}$. Επιπλέον, λόγω του γεγονότος ότι οι στήλες 1 και 3, όπως και οι γραμμές 1 και 3, του R' είναι αμοιβαία ορθογώνιες και τα μέτρα των γραμμών και στηλών του R' είναι μοναδιαία, μπορούμε να δείξουμε ότι

$$R' = \begin{bmatrix} \alpha & 0 & \beta \\ 0 & 1 & 0 \\ -s\beta & 0 & s\alpha \end{bmatrix} \quad (\text{B.41})$$

όπου

$$\alpha = \frac{\lambda_1}{\lambda_2} - \delta \Delta x' = s (k\lambda_3 - \Delta z') \quad (\text{B.42})$$

$$\beta = -\Delta x' = s\delta \Delta z' = \pm \sqrt{1 - \alpha^2} \quad (\text{B.43})$$

$$s = \det(U)\det(V) \quad (\text{B.44})$$

Αφού οι U και V είναι ορθοκανονικοί, από την (B.44) συμπεραίνουμε ότι το s είναι ή $+1$ ή -1 . Θα αποδείξουμε ότι αν και τα $\det(U)$, $\det(V)$ μπορεί να είναι $+1$ ή -1 για κάποιον πίνακα A , το s είναι μοναδικό δεδομένου του πίνακα A .

Θυμίζουμε ότι τα U_1 , U_2 και U_3 είναι τα ιδιοδιανύσματα του AA^T , που αντιστοιχούν στις ιδιοτιμές λ_1^2 , λ_2^2 και λ_3^2 . Επειδή τα $\lambda_1, \lambda_2, \lambda_3$ είναι όλα διαφορετικά, τα U_1, U_2, U_3, V_1, V_2 και V_3 είναι όλα σταθερά, πέραν των προσήμων. Όμως, σύμφωνα με την απόδειξη του θεωρήματος 1 (βλ. B.1), έχουμε :

$$AV_1 = \lambda_1 U_1$$

$$AV_2 = \lambda_2 U_2$$

$$AV_3 = \lambda_3 U_3$$

Φαίνεται, λοιπόν, καθαρά, ότι αν το U_i αλλάζει πρόσημο, τότε και το V_i θα αλλάξει ($i = 1, 2, 3$). Άρα το πρόσημο $\det(U)\det(V)$ παραμένει σταθερό, οπότε το s είναι μοναδικό.

Από τις (B.42) και (B.43) έχουμε :

$$\alpha - \frac{\lambda_1}{\lambda_2} = \beta\delta \quad (\text{B.45})$$

$$\delta \left(\alpha + s \frac{\lambda_3}{\lambda_2} \right) = \beta \quad (\text{B.46})$$

Απαλείφοντας το β στις (B.45) και (B.46) προκύπτει

$$\alpha = \frac{\lambda_1 + s\lambda_3\delta^2}{\lambda_2(1 + \delta^2)}$$

όπου

$$\delta = \pm \left(\frac{\lambda_1^2 - \lambda_2^2}{\lambda_2^2 - \lambda_3^2} \right)^{\frac{1}{2}}$$

Από τις (B.42) και (B.43) συν του γεγονότος ότι $\Delta y' = 0$ έχουμε

Γεώργιος Σημαντήρης

$$\begin{bmatrix} \Delta x' \\ \Delta y' \\ \Delta z' \end{bmatrix} = \begin{bmatrix} -\beta \\ 0 \\ \frac{\lambda_3}{\lambda_2} - s\alpha \end{bmatrix} \quad (\text{B.47})$$

Ενώ από τις (B.37), (B.39) και (B.47) έχουμε :

$$\begin{aligned} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} &= w^{-1}U \begin{bmatrix} \Delta x' \\ \Delta y' \\ \Delta z' \end{bmatrix} \Leftrightarrow \\ \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} &= w^{-1}U \begin{bmatrix} -\beta \\ 0 \\ \frac{\lambda_3}{\lambda_2} - s\alpha \end{bmatrix} \Leftrightarrow \\ \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} &= w^{-1} \left[-\beta U_1 + \left(\frac{\lambda_3}{\lambda_2} - s\alpha \right) U_3 \right] \end{aligned} \quad (\text{B.48})$$

Από τις (B.36) και (B.41) έχουμε :

$$R = U \begin{bmatrix} \alpha & 0 & \beta \\ 0 & 1 & 0 \\ -s\beta & 0 & s\alpha \end{bmatrix} V^T \quad (\text{B.49})$$

Τέλος, από τις (B.33), (B.48), (B.49) και λόγω του γεγονότος ότι το s είναι σταθερό, βλέπουμε ότι υπάρχουν ακριβώς δύο λύσεις, πέραν ενός παράγοντα κλίμακας για τις παραμέτρους μεταφοράς και γεωμετρίας (το διάνυσμα του προσανατολισμού).

Μένει να αποδειχθεί ότι μία αναγκαία και ικανή συνθήκη, ώστε τα singular values να είναι διαφορετικά, είναι ότι το διάνυσμα της μεταφοράς δεν συμπίπτει με το διάνυσμα του προσανατολισμού (το κάθετο διάνυσμα στο επίπεδο) τη χρονική στιγμή t_2 . Η ικανότητα αποδείχθηκε στο θεώρημα 1 (B.1). Η αναγκαιότητα θα αποδειχθεί με αντίφαση.

Στο θεώρημα 1 (B.1) φάνηκε ότι το κάθετο διάνυσμα στο επίπεδο τη χρονική στιγμή t_2 προκύπτει από το κάθετο διάνυσμα (a, b, c) τη χρονική στιγμή t_1 περιστρέφοντάς το κατά R :

$$R \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

Ας υποθέσουμε ότι το διάνυσμα της μεταφοράς

$$\begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix}$$

είναι παράλληλο στο

$$R \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

Τότε :

$$\begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} = hR \begin{bmatrix} a \\ b \\ c \end{bmatrix} \quad (\text{B.50})$$

για κάποια σταθερά h . Με βάση τις (B.33), (B.48), (B.49), και (B.50) έχουμε :

$$\begin{aligned} w^{-1}U \begin{bmatrix} -\beta \\ 0 \\ \frac{\lambda_3}{\lambda_2} - s\alpha \end{bmatrix} &= hU \begin{bmatrix} \alpha & 0 & \beta \\ 0 & 1 & 0 \\ -s\beta & 0 & s\alpha \end{bmatrix} V^T V \begin{bmatrix} \delta \\ 0 \\ 1 \end{bmatrix} \Leftrightarrow \\ \begin{bmatrix} -\beta \\ 0 \\ \frac{\lambda_3}{\lambda_2} - s\alpha \end{bmatrix} &= wh \begin{bmatrix} \alpha & 0 & \beta \\ 0 & 1 & 0 \\ -s\beta & 0 & s\alpha \end{bmatrix} \begin{bmatrix} \delta \\ 0 \\ 1 \end{bmatrix} \end{aligned}$$

το οποίο συνεπάγει :

$$-\beta = wh(\alpha\delta + \beta) \quad (\text{B.51})$$

και

$$\frac{\lambda_3}{\lambda_2} - s\alpha = wh(-s\beta\delta + s\alpha) \quad (\text{B.52})$$

Αντικαθιστώντας τώρα τις (B.42) και (B.43) στις (B.51) και (B.52) έχουμε

$$\Delta x' = \frac{wh\delta}{1 + wh(1 + \delta^2)} \frac{\lambda_1}{\lambda_2} \quad (\text{B.53})$$

Γεώργιος Σημαντήρης

$$\Delta z' = \frac{wh}{1 + wh(1 + \delta^2)} \frac{\lambda_3}{\lambda_2} \quad (\text{B.54})$$

Αλλά, λόγω της (B.43), προκύπτει :

$$-\Delta x' = s\delta\Delta z' \quad (\text{B.55})$$

Αντικαθιστούμε τις (B.53) και (B.54) στην (B.55), οπότε :

$$\frac{wh\delta}{1 + wh(1 + \delta^2)} \frac{\lambda_1}{\lambda_2} = \frac{wh\delta}{1 + wh(1 + \delta^2)} \frac{-s\lambda_3}{\lambda_2}$$

η οποία συνεπάγει $\lambda_1 = -s\lambda_3$. Τα λ_1 και λ_3 είναι εξ' ορισμού μη αρνητικά, άρα θα πρέπει $\lambda_1 = \lambda_3$. Αυτό όμως αντιφάσκει στην αρχική υπόθεση ότι $\lambda_1 \neq \lambda_3$, οπότε είναι άτοπο. Επομένως αποδείχτηκε και η ικανότητα.

B.3 Θεώρημα 3

Η ικανή και αναγκαία συνθήκη, ώστε η πολλαπλότητα των singular values του πίνακα A να είναι τρία, δηλαδή $\lambda_1 = \lambda_2 = \lambda_3$, είναι ότι η κίνηση αποτελείται μόνο από μία περιστροφή γύρω από έναν άξονα, που περνάει απ' την αρχή των αξόνων. Δηλαδή το διάνυσμα της μεταφοράς $(\Delta x, \Delta y, \Delta z)$ είναι μηδέν. Ακόμα, ο πίνακας περιστροφής είναι μοναδικός και $R = \lambda_1^{-1}A$. Η επιφάνεια μπορεί να βρίσκεται οπουδήποτε.

Απόδειξη : Εάν $\lambda_1 = \lambda_2 = \lambda_3$, τότε η (4.36) δίνει

$$x_n'^2 + y_n'^2 + z_n'^2 = k^2\lambda_1^2 (x_n''^2 + y_n''^2 + z_n''^2) \quad (\text{B.56})$$

Εφόσον οποιαδήποτε κίνηση στερεού αντικειμένου μπορεί να αναλυθεί σε περιστροφή ακολουθούμενη από μεταφορά, πρώτα περιστρέφουμε το αντικείμενο έτσι ώστε το (x_n, y_n, z_n) να γίνει (x_n'', y_n'', z_n'') και στη συνέχεια εκτελούμε τη μεταφορά, η οποία αλλάζει το (x_n'', y_n'', z_n'') σε (x_n', y_n', z_n') , δηλαδή :

$$\begin{bmatrix} x_n'' \\ y_n'' \\ z_n'' \end{bmatrix} = R' \begin{bmatrix} x_n \\ y_n \\ z_n \end{bmatrix} \quad (\text{B.57})$$

και

$$\begin{bmatrix} x_n' \\ y_n' \\ z_n' \end{bmatrix} = \begin{bmatrix} x_n'' \\ y_n'' \\ z_n'' \end{bmatrix} + \begin{bmatrix} \Delta x' \\ \Delta y' \\ \Delta z' \end{bmatrix} \quad (\text{B.58})$$

όπου R' , $\Delta x'$, $\Delta y'$ και $\Delta z'$ είναι οι παράμετροι κίνησης στο (x_n, y_n, z_n) χώρο, όπως ορίστηκαν στις (B.8) και (B.9). Από την (B.57) έχουμε :

$$\begin{aligned} x_n''^2 + y_n''^2 + z_n''^2 &= \begin{bmatrix} x_n & y_n & z_n \end{bmatrix} R'^T R' \begin{bmatrix} x_n \\ y_n \\ z_n \end{bmatrix} \Leftrightarrow \\ x_n''^2 + y_n''^2 + z_n''^2 &= x_n^2 + y_n^2 + z_n^2 \end{aligned}$$

Συνδυάζοντας το αποτέλεσμα αυτό με την (B.58) προκύπτει :

$$x_n^2 + y_n^2 + z_n^2 = (x'_n - \Delta x')^2 + (y'_n - \Delta y')^2 + (z'_n - \Delta z')^2 \quad (\text{B.59})$$

Τέλος, από τις (B.56) και (B.59), έχουμε :

$$\begin{aligned} &(k^2 \lambda_1^2 - 1)x_n'^2 + (k^2 \lambda_1^2 - 1)y_n'^2 + (k^2 \lambda_1^2 - 1)z_n'^2 - \\ &- \left[2\Delta x' x'_n + 2\Delta y' y'_n + 2\Delta z' z'_n - \Delta x'^2 - \Delta y'^2 - \Delta z'^2 \right] k^2 \lambda_1^2 = 0 \end{aligned} \quad (\text{B.60})$$

Εφόσον η (B.60) ισχύει για όλα τα x'_n , y'_n και z'_n , θα πρέπει

$$\Delta x' = \Delta y' = \Delta z' = 0$$

και

$$k \lambda_1 = 1 \Rightarrow k = \frac{1}{\lambda_1}$$

Συμπεραίνουμε, λοιπόν, από την (B.9)

$$\Delta x = \Delta y = \Delta z = 0$$

Έτσι όμως, η (4.28) δίνει :

$$R + 0 = kA \Rightarrow R = \lambda_1^{-1} A$$

Αποδείξαμε, λοιπόν, ότι αν $\lambda_1 = \lambda_2 = \lambda_3$, τότε η κίνηση αποτελείται μόνο από μία περιστροφή γύρω από έναν άξονα, που περνάει από την αρχή των αξόνων, και η λύση για τον πίνακα περιστροφής είναι μοναδική. Η επιφάνεια του αντικειμένου μπορεί να βρίσκεται οπουδήποτε. Αυτό αποδεικνύει την αναγκαιότητα.

Θα αποδείξουμε τώρα και την ικανότητα. Αν η κίνηση αποτελείται μόνο από μία περιστροφή γύρω από έναν άξονα, που περνάει από την αρχή των αξόνων, δηλαδή $\Delta x = \Delta y = \Delta z = 0$, τότε από την (4.28) έχουμε :

$$A = k^{-1} R \quad (\text{B.61})$$

Γεώργιος Σημαντήρης

Έστω, $U_A = R$, $V_A = I$ και $\Lambda_A = k^{-1}I$. Τότε η (B.61) γίνεται :

$$\begin{aligned}
 A &= U_A \Lambda_A V_A^T \Leftrightarrow \\
 A &= U_A \begin{bmatrix} k^{-1} & & \\ & k^{-1} & \\ & & k^{-1} \end{bmatrix} V_A^T
 \end{aligned} \tag{B.62}$$

Εφόσον οι U_A και V_A είναι ορθοκανονικοί, η (B.62) δίνει τη Singular Value Decomposition (SVD) του A , με singular values k^{-1} , k^{-1} και k^{-1} . Έτσι, από το γεγονός ότι τα singular values οποιουδήποτε πίνακα είναι μοναδικά, βλέπουμε ότι ο A έχει τρία ίδια singular values. Αυτό αποδεικνύει την ικανότητα.

Βιβλιογραφία

- [1] Γεώργιος Σημαντήρης, Ιωάννης Μαυρικάκης, Ηλίας Γκρίνιας, Christophe Garcia, and Γεώργιος Τζιρίτας. Ανίχνευση Προσώπων και Χαρακτηριστικών Προσώπου. Technical Report CSD-TR-2001-02, University of Crete, Computer Science Department, Heraklion, Greece, May 2001.
- [2] K. Aizawa, C. S. Choi, H. Harashima, and T. S. Huang. Human facial motion analysis and synthesis with application to model-based coding. In M. Sezan and R. Lagendijk, editors, *Motion Analysis and Image Processing*, chapter 11, pages 317-348. Kluwer Academic Publishers.
- [3] J. Atick, P. Griffin, and N. Redlich. Statistical approach to shape from shading : Reconstruction of three-dimensional face surfaces from single two-dimensional images. *Neural Computation*, (8), 1997.
- [4] A. Azarbayejani and A. Pentland. Recursive estimation of motion, structure and focal length. *IEEE Pattern Analysis and Machine Intelligence*, June 1995.
- [5] A. Azarbayejani, T. Starner, B. Horowitz, and A. Pentland. Visually controlled graphics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(15):602-605, June 1993.
- [6] Vera Bakic and George Stockman. Real-time tracking of face features and gaze direction determination. citeseer.nj.nec.com/432670.html.
- [7] Michael J. Black and Yaser Yacoob. Tracking and recognizing facial expressions in image sequences, using local parameterized models of image motion. Technical Report CAR-TR-756, University of Maryland, Center for Automation Research, College Park, January 1995.
- [8] Gary R. Bradski. Computer vision face tracking for use in a perceptual user interface. http://developer.intel.com/technology/itj/q21998/articles/art_2.htm.
- [9] Roberto Brunelli and Tomaso Poggio. Face recognition : Features versus templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(11):1042-1052, 1993.

- [10] E. Cartan. *La Theorie des Groupes Finis et Continus*. Librairie-Imprimerie Gauthier-Villars, Paris, France, 1937.
- [11] Marco La Cascia, Stan Sclaroff, and Vassilis Athitsos. Fast, reliable head tracking under varying illumination : An approach based on registration of texture-mapped 3d models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(4):322-336, April 2000.
- [12] C. S. Choi, K. Aizawa, H. Harashima, and T. Takabe. Analysis and synthesis of facial image sequences in model-based image coding. *IEEE Transactions on Circuits and Systems for Video Technology*, 4(3):257-275, June 1994.
- [13] Antonio Colmenarez, Ricardo Lopez, and Thomas S. Huang. 3d model-based head tracking. *Visual Communication and Image Processing*, 1997.
- [14] Y. Dai and Y. Nakano. Recognition of facial images with low resolution using a hopfield memory model. *Pattern Recognition*, 2(31):159-167, 1998.
- [15] Trevor J. Darrell, Irfan A. Essa, and Alex P. Pentland. Task-specific gesture analysis in real-time using interpolated views. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(12):1236-1242, December 1996.
- [16] Douglas DeCarlo and Dimitris Metaxas. Deformable model-based face shape and motion estimation. *IEEE Proc. of ICFG*, 1996.
- [17] Manolis Delakis and Christophe Garcia. Robust face detection based on convolutional neural networks. In *Proceedings of the 2nd Hellenic Conference on Artificial Intelligence (SETN'02)*, pages 367-378, Thessaloniki, Greece, April 2002.
- [18] L. P. Eisenhart. *Continuous Groups of Transformations*. Dover, New York, 1933.
- [19] A. Eleftheridis and A. Jacquin. Model-assisted coding of video teleconferencing sequences at low bit rates. In *Proceedings of IEEE Int. Symp. Circuits and Systems*, pages 3177-3180, May 1994.
- [20] I. A. Essa and A. Pentland. A vision system for observing and extracting facial action parameters. In *Proceedings Computer Vision and Pattern Recognition, CVPR-94*, pages 76-83, Seattle, WA, June 1994.
- [21] L. G. Farkas. *Anthropometry of the Head and Face*. Raven Press, second edition, 1994.
- [22] M. A. Fischer and R. C. Bolles. Random sample consensus : A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 6(24):381-395, 1981.

- [23] D. J. Fleet and A. D. Jepson. Computation of component image velocity from local phase information. *International Journal of Computer Vision*, 5(1):77, 1990.
- [24] K. Fukunaga. *Introduction to Statistical Pattern Recognition*. Academic Press, Boston, 1990.
- [25] Christophe Garcia and Manolis Delakis. Face detection using trained convolutional filters. To appear in *Proceedings of the XI European Signal Processing Conference (EUSIPCO2002)*, Toulouse, France, September 2002.
- [26] Christophe Garcia and Manolis Delakis. A neural architecture for fast and robust face detection. To appear in *Proceedings of the IEEE-IAPR International Conference on Pattern Recognition (ICPR '02)*, Quebec city, Canada, August 2002.
- [27] Christophe Garcia, George Simandiris, and Giorgos Tziritas. A feature-based face detector using wavelet frames. In *Proceedings of the International Workshop on Very Low Bitrate Video Coding (VLBV 01)*, pages 71-67, Athens, Greece, October 2001.
- [28] Christophe Garcia and Giorgos Tziritas. Face detection using quantized skin color regions merging and wavelet packet analysis. *IEEE Transactions on Multimedia*, 3(1):264-277, September 1999.
- [29] Christophe Garcia, Giorgos Zikos, and Giorgos Tziritas. A wavelet-based framework for face recognition. In *Proceedings of Workshop on Advances in Facial Image Analysis and Recognition Technology*, Freiburg, Germany, June 1998. 5th European Conference on Computer Vision.
- [30] Christophe Garcia, Giorgos Zikos, and Giorgos Tziritas. Face detection in color images using wavelet packet analysis. In *Proceedings of the 6th IEEE International Conference on Multimedia Computing and Systems (ICMCS'99)*, pages 703-708, Florence, Italy, June 1999.
- [31] Christophe Garcia, Giorgos Zikos, and Giorgos Tziritas. Wavelet packet analysis for face recognition. *Image and Vision Computing*, 4(18):289-297, February 2000.
- [32] A. Gee and R. Cipolla. Estimating gaze from a single view of a face. In *ICPR '94*, pages 758-760, 1994.
- [33] Andrew H. Gee and Roberto Cipolla. Fast visual tracking by temporal consensus. Technical Report CUED/F-INFENG/TR 207, Cambridge University, Department of Engineering, Cambridge, England, February 1995.
- [34] Jim Gemmell, Kentaro Toyama, C. Lawrence Zitnick, Thomas Kang, and Steven Seitz. Gaze awareness for video-conferencing : A software approach. *IEEE Multimedia*, 7(4):26-35, October 2000.

- [35] Shaogang Gong, Alexandra Psarrou, Ilias Katsoulis, and Pascalis Palavouzis. Tracking and recognition of face sequences. In *European Workshop on Combined Real and Synthetic Image Processing for Broadcast and Video Production*, Hamburg, Germany, November 1994.
- [36] T. Hogg, D. Rees, and H. Talhami. Three-dimensional pose from two-dimensional images : a novel approach using synergetic networks. In *IEEE International Conference on Neural Networks 2*, pages 1140-1144, 1995.
- [37] Thanarat Horprasert, Yaser Yacoob, and Larry S. Davis. Computing 3-d head orientation from a monocular image sequence. In *Proceedings of the International Conference on Automatic Face and Gesture Recognition*, pages 242-247, Los Alamitos, CA, USA, 1996. IEEE Computer Society Press.
- [38] A. Jonathan Howell and Hilary Buxton. Recognising simple behaviours using time-delay rbf networks. Technical Report CSRP 456, University of Sussex, ISSN 1350-3162, February 1997.
- [39] Fu Jie Huang and Tsuhan Chen. Tracking of multiple faces for human-computer interfaces and visual environments. *IEEE Int. Conf. on Multimedia and Expo.*, July 2000.
- [40] Martin H. Hunke. Locating and tracking of human faces with neural networks. Technical Report CMU-CS-94-155, Carnegie Mellon University, School of Computer Science, Pittsburgh, PA, USA, August 1994.
- [41] Martin H. Hunke and Alex Waibel. Face locating and tracking for human-computer interaction. In *Proceedings of the Twenty-Eight Asilomar Conference on Signal, Systems and Computers*, Monterey, CA, USA, 1994.
- [42] Tony Jebara, Kenneth Russell, and Alex Pentland. Mixtures of eigenfeatures for real-time structure from texture. In *Proceedings of ICCV'98*, Bombay, India, January 1998.
- [43] Tony S. Jebara and Alex Pentland. Parameterized structure from motion for 3d adaptive feedback tracking of faces. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 144-150, San Juan, Puerto Rico, June 1997.
- [44] S.-H. Jeng, H. Y. M. Yao, C. C. Han, M. Y. Chern, and Y. T. Liu. Facial feature detection using geometrical face model : An efficient approach. *Pattern Recognition*, 3(31):273-282, 1998.
- [45] Qiang Ji and Rong Hu. 3d face pose estimation and tracking from a monocular camera. *Image and Vision Computing*, 20(7):499-511, 2002.

- [46] M. Kirby and L. Sirovich. Application of the karhunen-loeve procedure and the characterization of human faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1(12):103-108, 1990.
- [47] E. Levin and R. Pieraccini. Dynamic planar warping for optical character recognition. In *Proceedings of the Int. Conf. Acoust., Speech, Signal Processing*, pages 149-152, 1992.
- [48] S.-H. Lin, S.-Y. Kung, and L.-J. Lin. Face recognition/detection by probabilistic decisionbased neural network. *IEEE Transactions on Neural Networks*, 1(8):114-131, 1997.
- [49] C. Loewner. *Theory of Continuous Groups*. MIT Press, Cambridge, MA, USA, 1971.
- [50] H. C. Longuet-Higgins. The visual ambiguity of a moving plane. In *Proc. Royal Soc. London, B 223*, pages 165-175, 1984.
- [51] H. C. Longuet-Higgins and K. Prazdny. The interpolation of a moving retinal image. In *Proc. Royal Soc. London, B 208*, pages 385-397, 1980.
- [52] Ricardo Lopez and Thomas Huang. 3d head pose computation from 2d images : Templates versus features. In *IEEE International Conference in Image Processing*, pages 220-224, Washington D.C., USA, October 1995. IEEE, IEEE Press.
- [53] Y. Matsushima. *Differentiable Manifolds*. Marcel Dekker, New York, 1972.
- [54] Thomas Maurer and Christoph von der Malsburg. Tracking and learning graphs and pose on image sequences of faces. *IEEE Proc. of ICFG*, pages 176-181, 1995.
- [55] Baback Moghaddam and Alex Pentland. An automatic system for model-based coding of faces. *IEEE Data Compression Conference*, March 1995.
- [56] T. G. Newman and D. A. Demus. Lie theoretic methods in video tracking. In *Proc. Workshop Auto. Missile Guidance*, Redstone, Arsenal, AL, USA, November 1979.
- [57] A. O'Toole, H. Abdi, K. Deffenbacher, and D. Valentin. Low-dimensional representation of faces in higher dimensions of the face space. *Journal of the Optical Society of America*, 10(3):405-410, 1993.
- [58] R. Pappu and P. A. Beardsley. A qualitative approach to classifying gaze direction. In *Proceedings of the Third IEEE International Conference on Automatic Face and Gesture Recognition*, Nara, Japan, 1998.
- [59] A. Pentland, R. W. Picard, and S. Sclaroff. Photobook : Content-based manipulation of image databases. In *Proceedings of the SPIE Storage and Retrieval and Video Databases II*, San Jose, CA, 1994.

- [60] J. L. Perry and J. M. Carney. Human face recognition using a multilayer perceptron. In *IJCNN*, pages 413-416, Washington, D.C., 1990.
- [61] P. J. Phillips, R. McCabe, and R. Chellappa. Biometric image processing and recognition. In *Proceedings of the IX European Signal Processing Conference*, Rhodes, Greece, 1998.
- [62] L. S. Pontryagin. *Topological Groups*. Gordon and Breach, New York, 1966.
- [63] R. Rae and H. J. Ritter. Recognition of human head orientation based on artificial neural networks. *IEEE Transactions on Neural Networks*, 9(2):257-265, 1998.
- [64] Marcia G. Ramos and Sheila S. Hemami. Eigenfeatures coding of videoconferencing sequences. In *SPIE's Symposium on Visual Communications & Image Processing*, 1996.
- [65] Ram R. Rao and Russell M. Mersereau. On merging hidden markov models with deformable templates. In *Proceedings of the Int'l Conf. on Image Processing*, Washington D.C., 1995.
- [66] Mark Rosenblum, Yaser Yacoob, and Larry S. Davis. Human emotion recognition from motion using a radial basis function network architecture. *IEEE Transactions on Neural Networks*, 5(7):1121-1138, 1996.
- [67] H. A. Rowley, S. Baluja, and T. Kanade. Neural network-based face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (20):23-28, 1998.
- [68] Jocelyn Scheirer, Raul Fernandez, and Rosalind W. Picard. Expression glasses : A wearable device for facial expression recognition. In *CHI 99 Conference Proceedings*, Pittsburg, Pennsylvania, USA, May 1999. ACM SIGCHI Conference on Human Factors in Computing Systems.
- [69] P. Sinha. Object recognition via image invariants : A case study. *Investigative Ophthalmology and Visual Science*, (35):1735-1740, 1994.
- [70] L. Sirovich and M. Kirby. Low-dimensional procedure for the characterization of human faces. *Journal of the Optical Society of America*, 4(3):519-524, March 1987.
- [71] K. Sobottka and I. Pitas. Segmentation and tracking of faces in color images. In *Proceedings of the Second International Conference On Auto., Face and Gesture Recognition*, pages 236-241, 1996.
- [72] Rainer Stiefelhagen, Jie Yang, and Alex Waibel. Tracking eyes and monitoring eye gaze. In *Proceedings of the Workshop on Perceptual User Interfaces*, pages 98-100, Banff, Canada, October 1997.

- [73] J. Ström, T. Jebara, S. Basu, and Alex Pentland. Real time tracking and modeling of faces : An ekf-based analysis by synthesis approach. In *Proceedings of the IEEE International Workshop on Modelling People*, pages 55-62, Corfu, Greece, September 1999.
- [74] Muralidhara Subbarao and Allen M. Waxman. On the uniqueness of image flow solutions for planar surfaces in motion. Technical Report CAR-TR-114, University of Maryland, Center for Automation Research, College Park, Maryland, MD, USA, April 1985.
- [75] D. L. Swets and J. Weng. Using discriminant eigenfeatures for image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(18):831-836, August 1996.
- [76] D. Terzopoulos and K. Waters. Analysis and synthesis of facial image sequences using physical and anatomical models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(15):569-579, June 1993.
- [77] W. M. Theimer and H. A. Mallot. Phase-based binocular vergence control and depth reconstruction using active vision. *CVGIP : Image Understanding*, 60(3):343, 1994.
- [78] Kentaro Toyama. Look, ma - no hands! hands-free cursor control with real-time 3d face tracking. In *Proc. Workshop on Perceptual User Interfaces (PUI98)*, San Francisco, CA, USA, November 1998.
- [79] Kentaro Toyama. Prolegomena for robust face tracking. Technical Report MSR-TR-98-65, Microsoft Research, Vision Technology Group, Redmond, WA, USA, November 1998. (presented at the Post-ECCV Workshop on Advances in Facial Image Analysis and Recognition Technology, Freiburg, May 1998).
- [80] Kentaro Toyama and Gregory D. Hager. Incremental focus of attention for robust visual tracking. In *Proc. CVPR*, pages 189-195, 1996.
- [81] Kentaro Toyama and Gregory D. Hager. If at first you don't succeed. . . . In *Proc. AAAI*, pages 3-9, Providence, RI, 1997.
- [82] Roger Y. Tsai and Thomas S. Huang. Estimating three-dimensional motion parameters of a rigid planar patch. *IEEE Transactions on Acoustics, Speech and Signal Processing*, ASSP-29(6):1147-1152, December 1981.
- [83] Roger Y. Tsai and Thomas S. Huang. Estimating three-dimensional motion parameters of a rigid planar patch. In *Proc. 1981 Int. Conf. Pattern Recognition, Image Processing*, Dallas, TX, USA, August 1981.

- [84] Roger Y. Tsai and Thomas S. Huang. Estimating three-dimensional motion parameters of a rigid planar patch, iii : Finite point correspondences and the three-view problem. *IEEE Transactions on Acoustics, Speech and Signal Processing*, ASSP-32(2):213-219, April 1984.
- [85] Roger Y. Tsai, Thomas S. Huang, and Wei-Le Zhu. Estimating three-dimensional motion parameters of a rigid planar patch, ii : Singular value decomposition. *IEEE Transactions on Acoustics, Speech and Signal Processing*, ASSP-30(4):525-534, August 1982.
- [86] A. Tsukamoto, C. Lee, and S. Tsuji. Detection and pose estimation of human face with synthesized image models. In *ICPR '94*, pages 754-757, 1994.
- [87] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Science*, 1(3):71-86, 1991.
- [88] Stéphane Valente and Jean-Luc Dugelay. Face tracking and realistic animations for telecommunican clones. *IEEE Multimedia Computing and Systems*, 7(1):34-43, January 2000.
- [89] Stéphane Valente and Jean-Luc Dugelay. A visual analysis/synthesis feedback loop for accurate face tracking. *Signal Processing : Image Communication*, 16(6):585-608, February 2001.
- [90] H. Wang and S. F. Chang. A highly efficient system for automatic face region detection in mpeg video. *IEEE Transactions on Circuits and Systems For Video Technology*, 4(7):615-628, 1997.
- [91] Ru-Shang Wang and Yao Wang. Facial feature extraction and tracking in video sequences. In *Workshop of Multimedia Signal Processing*, pages 233-238, Princeton, New Jersey, USA, June 1997. IEEE Signal Processing Society, USA Electronic Proceedings.
- [92] J. W. Welsh and D. Shah. Facial-feature image coding using principal components. *Electronic Letters*, 28(22), October 1992.
- [93] C. L. Wilson, C. S. Barnes, R. Chellappa, and S. A. Sirohey. Face recognition technology for law enforcement applications. In *NISTIR 5465*, U.S. Department of Commerce, 1994.
- [94] L. Wiskott, J. M. Fellous, N. Kruger, and C. von der Malsburg. Face recognition by elastic bunch graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7(19):775-779, 1997.

- [95] Yaser Yacoob and Larry S. Davis. Computing spatio-temporal representations of human faces. In *Proceedings Computer Vision and Pattern Recognition, CVPR-94*, pages 70-75, Seattle, WA, June 1994.
- [96] Yaser Yacoob and Larry S. Davis. Recognizing human facial expressions. Technical Report CAR-TR-706, University of Maryland, Center for Automation Research, College Park, MD, USA, May 1994.
- [97] G. Yang and T. S. Huang. Human face detection in a complex background. *Pattern Recognition*, 1(27):55-63, 1994.
- [98] Jie Yang and Alex Waibel. A real-time face tracker. In *Proceedings of WACV'96*, pages 142-147, 1996.
- [99] J. Young. Head and face anthropometry of adult u. s. citizens. Government Report DOT/FAA/AM-93/10, July 1993.
- [100] K. C. Yow and C. Cipolla. Feature-based human face detection. *Image and Vision Computing*, (15):713-735, 1997.
- [101] A .L. Yuille, D. S. Cohen, and P. W. Hallinan. Deformable templates. In A. Blake and A. Yuille, editors, *Active Vision*, pages 21-38. MIT Press, Cambridge MA, 1992.