



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

**Βάση οπτικών δεδομένων για τη μοντελοποίηση και
ανίχνευση της προσοχής του χρήστη σε περιβάλλοντα
φυσικής αλληλεπίδρασης**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Δημήτριος Κ. Σουφλερός

Επιβλέπων : Στέφανος Κόλλιας
Καθηγητής Ε.Μ.Π

Αθήνα, Φεβρουάριος 2010



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

**Βάση οπτικών δεδομένων για τη μοντελοποίηση και
ανίχνευση της προσοχής του χρήστη σε περιβάλλοντα
φυσικής αλληλεπίδρασης**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Δημήτριος Κ. Σουφλερός

Επιβλέπων : Στέφανος Κόλλιας
Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 8^η Μαρτίου 2010.

.....
Στέφανος Κόλλιας
Καθηγητής Ε.Μ.Π.

.....
Ανδρέας-Γεώργιος
Σταφυλοπάτης
Καθηγητής Ε.Μ.Π.

.....
Γεώργιος Στάμου
Λέκτορας Ε.Μ.Π.

Αθήνα, Φεβρουάριος 2010

.....

Δημήτριος Κ. Σουφλερός

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

Copyright © Δημήτριος Κ. Σουφλερός, 2010.

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

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

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

Περίληψη

Η ανάλυση κινούμενης εικόνας και η εξαγωγή πληροφοριών για τα αντικείμενα που περιέχει και τις δραστηριότητές τους αποτελεί έναν σημαντικό κομμάτι του τομέα της όρασης υπολογιστών. Ειδικά η ανθρώπινη συμπεριφορά και η κατανόηση της από τον υπολογιστή ερευνώνται διεξοδικά τα τελευταία χρόνια και τα αποτελέσματα που προκύπτουν έχουν άμεση εφαρμογή σε τεχνολογίες αιχμής, διευρύνοντας ακόμη περισσότερο τους ορίζοντες για περαιτέρω μελέτη και εξέλιξη. Η παρούσα διπλωματική εργασία ασχολείται με την ανίχνευση της θέσης του κεφαλιού του ανθρώπου και την εκτίμηση της κίνησης του στο χώρο. Για τον λόγο αυτό δημιουργήθηκε μία βάση προσώπων σε ακολουθίες καρέ με είκοσι βίντεο καταγεγραμμένα σε τυπικές συνθήκες ενός εργασιακού χώρου, με πολύπλοκο περιβάλλοντα χώρο με έντονη ανθρώπινη δραστηριότητα και τεχνητό φωτισμό. Τα βίντεο προέρχονται από δέκα εθελοντές που κλήθηκαν να πραγματοποιήσουν κινήσεις σε δύο διαφορετικές συνεδρίες. Στην μεν πρώτη κινούνται ελεύθερα με αυτόβουλες περιστροφικές ή μεταφορικές κινήσεις του κεφαλιού σε οριζόντιο και κάθετο άξονα, στην δε δεύτερη ακολουθούν ένα προκαθορισμένο μοτίβο κινήσεων που συνδυάζει αλλαγές στην πόζα του κεφαλιού και της κατεύθυνσης του βλέμματος. Για την ανίχνευση της στάσης του κεφαλιού κατασκευάστηκε και χρησιμοποιήθηκε μία πρότυπη διάταξη τριών φωτοδιόδων τοποθετημένη στο κεφάλι. Με την βοήθεια της, ο αλγόριθμος που υλοποιήθηκε σε Matlab εντοπίζει την θέση των φωτοδιόδων, εκτιμά την στάση του κεφαλιού σε κάθε καρέ του βίντεο και εξάγει πληροφορίες σχετικά με την γωνία περιστροφής του κεφαλιού και την θέση του. Είναι σε θέση να ιχνηλατήσει τις φωτοδιόδους και απομονώσει τα σημεία ενδιαφέροντος σε σύνθετο φόντο με κινούμενα αντικείμενα, έντονο φωτισμό και αντανakλάσεις. Το ground truth που προκύπτει από την ανάλυση της βάσης από τον αλγόριθμο μπορεί να χρησιμεύσουν ως δεδομένα αναφοράς και σύγκρισης για άλλους ερευνητικούς αλγόριθμους που θα αξιοποιήσουν την συγκεκριμένη βάση. Η ιδιαιτερότητα της βάσης που δημιουργήσαμε έγκειται στον συνδυασμό της στάσης του κεφαλιού και της κίνησης των ματιών σε ακολουθίες καρέ, κάτι που κατά της διάρκεια της συγγραφής της παρούσας εργασίας δεν υπήρχε στην διεθνή βιβλιογραφία.

Λέξεις κλειδιά

Βάση προσώπων, πόζα κεφαλιού, κατεύθυνση βλέμματος, εκτίμηση στάσης κεφαλιού, ιχνηλάτηση φωτοδιόδων, γωνία περιστροφής, δεδομένα αναφοράς.

Abstract

Analysis of moving image and extraction of information about the elements included in it and their behaviours is an important part of the field of computer vision. Human behaviour, in particular, and its understanding by the computer have been subject to rigorous research in the last few years, while the outcomes of such research are readily applicable to cutting-edge technologies, expanding the potential for further research and progress even further. The present diploma thesis deals with tracing the position of the human head and with estimating its motion in space. For this purpose, a head pose and eye gaze dataset was created containing twenty videos recorded in typical workplace conditions, within complex background with intense human activity and artificial lighting. The videos depict ten volunteers who were asked to perform head movements, in two different sessions. In the first session, they performed free, self-induced rotational and translational movements, along both the horizontal and vertical axis, while in the second session they followed a predetermined pattern of movements combining changes in head pose and gaze direction. For the detection of head pose, a prototypical device with three LEDs was built, which the subjects wore on their heads. With the use of this device, the relevant program, implemented in Matlab, can detect the position of the LEDs, estimate the head pose in each frame of the video and export information about the rotation angle of the head and its position. It can trace the LEDs and isolate the points of interest even when the background is complex, with moving objects, intense lighting and reflections. The ground truth derived from analysis of the database by the program can serve as a point of reference and comparison for subsequent research projects utilizing this database. What makes our database special is the inclusion of head pose and eye gaze in a common framework, which appear for the first time in video sequences in a public available dataset.

Keywords

Head pose dataset, eye gaze estimation, head pose estimation, LEDs tracing, rotation angle, ground truth.

Πρόλογος

Αυτή η διπλωματική εργασία εκπονήθηκε κατά το ακαδημαϊκό έτος 2009 – 2010 στο Εργαστήριο Ψηφιακής. Επεξεργασίας Εικόνας, Βίντεο και Πολυμέσων (IVML) του Εθνικού Μετσόβιου Πολυτεχνείου. Θα ήθελα να ευχαριστήσω τον καθηγητή κ. Στέφανο Κόλλια για την εμπιστοσύνη που μου έδειξε αναθέτοντάς μου την εργασία αυτή. Ακόμα, θα ήθελα να ευχαριστήσω ιδιαίτερα τον Δρ. Κώστα Καρπούζη για την πολύτιμη και ουσιαστική βοήθεια και καθοδήγηση που μου παρείχε, καθώς και τον υποψήφιο Δρ. Στυλιανό Αστεριάδη, η διδακτορική διατριβή του οποίου αποτέλεσε έναυσμα για τη δική μου διπλωματική εργασία. Η ολοκληρωμένη και εμπειριστατωμένη δουλειά του, στην οποία στηρίχτηκε η έρευνά μου, και το ενδιαφέρον του για νέες ιδέες και προτάσεις, με έκαναν να αγαπήσω ακόμα περισσότερο το αντικείμενο. Καθ' όλη τη διάρκεια της εκπόνησης αυτής της εργασίας. Η συνεργασία μας υπήρξε άψογη καθ' όλη τη διάρκεια της εκπόνησης της εργασίας, και η βοήθειά του υπήρξε πολύτιμη τόσο σε επιστημονικό όσο και σε προσωπικό επίπεδο.

Δημήτριος Σουφλερός
Αθήνα, Φεβρουάριος 2010

Πίνακας περιεχομένων

Κεφάλαιο 1. Εισαγωγή	- 13 -
1.1 Κίνητρα	- 16 -
1.2 Μέθοδοι για την εκτίμηση της στάσης του κεφαλιού.....	- 17 -
1.3 Απαιτήσεις υλοποίησης σε επίπεδο εφαρμογών.....	- 18 -
1.4 Μεθοδολογίες καταγραφής των δεδομένων του ground truth.....	- 19 -
1.5 Συνήθεις υποθέσεις	- 21 -
Κεφάλαιο 2. Βάσεις με εικόνες προσώπων	- 24 -
2.1 Εισαγωγή.....	- 24 -
2.2 Μια εκτενής βάση προσώπων.....	- 24 -
2.3 XM2VTS βάση προσώπων	- 26 -
2.4 Βάση προσώπων “Yale”, έκδοση B	- 26 -
2.5 Βάση προσώπων AR.....	- 27 -
2.6 CVL βάση προσώπων	- 28 -
2.7 INDIAN βάση προσώπων.....	- 28 -
2.8 Βάση Pointing’04.....	- 29 -
2.9 Συγκεντρωτικός πίνακας	- 30 -
2.10 Περιγραφή της βάσης μας (βάση HPEG)	- 31 -
2.11 Περιεχόμενα Οπτικού Δίσκου	- 33 -
Κεφάλαιο 3. Διάταξη LEDs & αλγόριθμος.....	- 35 -
3.1 Εισαγωγή.....	- 35 -
3.2 Περιγραφή της διάταξης των φωτοδιόδων	- 35 -
3.3 Ανάλυση του αλγορίθμου.....	- 38 -
3.3.1 Δομή	- 38 -
3.3.2 MainScript.m	- 39 -
3.3.3 Tracker.m.....	- 46 -
3.3.4 Estimation.m.....	- 48 -
3.3.5 Angle.m.....	- 50 -
3.3.6 SimpleSign.m	- 52 -
Κεφάλαιο 4. Ανάλυση της βάσης HPEG	- 54 -
4.1 Εισαγωγή.....	- 54 -
4.1.1 Αρχεία εξόδου.....	- 54 -
4.1.2 Προκαθορισμένες ρυθμίσεις (Default Settings).....	- 55 -
4.1.3 Επεξήγηση όρων.....	- 55 -
4.1.4 Παράγοντες που επηρεάζουν τα αποτελέσματα	- 58 -

4.2	Ανάλυση των βίντεο της βάσης	- 59 -
4.2.1	Video1a	- 60 -
4.2.2	Video1b	- 62 -
4.2.3	Video2a	- 64 -
4.2.4	Video2b	- 67 -
4.2.5	Video3a	- 69 -
4.2.6	Video3b	- 71 -
4.2.7	Video4a	- 73 -
4.2.8	Video4b	- 76 -
4.2.9	Video5a	- 78 -
4.2.10	Video5b	- 81 -
4.2.11	Video6a	- 83 -
4.2.12	Video6b	- 86 -
4.2.13	Video7a	- 88 -
4.2.14	Video7b	- 91 -
4.2.15	Video8a	- 93 -
4.2.16	Video8b	- 96 -
4.2.17	Video9a	- 98 -
4.2.18	Video9b	- 101 -
4.2.19	Video10a	- 103 -
4.2.20	Video10b	- 106 -
Κεφάλαιο 5.	Συμπεράσματα & μελλοντικές επεκτάσεις.....	- 109 -
5.1	Συμπεράσματα	- 109 -
5.2	Μελλοντικές επεκτάσεις.....	- 111 -
Κεφάλαιο 6.	Βιβλιογραφία.....	- 113 -

Κατάλογος εικόνων - σχημάτων

Εικ. 1.1 Οι τρεις βαθμοί ελευθερίας κίνησης του ανθρώπινου κεφαλιού	- 14 -
Εικ.1.2 Παρόλο που το κεφάλι στα δύο σκίτσα έχει διαφορετική κλίση, τα μάτια βρίσκονται στην ίδια θέση [6].....	- 15 -
Εικ. 1.3 Τύποι μαγνητικών αισθητήρων.....	- 20 -
Εικ. 1.4 Οπτικά συστήματα λήψης κίνησης.....	- 21 -
Εικ. 2.1 Διαρρύθμιση του χώρου καταγραφής της βάσης	- 24 -
Εικ. 2.2 Στιγμιότυπα από ένα υποκείμενο σε διάφορες θέσεις του κεφαλιού και κατευθύνσεις βλέμματος. Παρατηρούμε το δείκτη λέιζερ στο κεφάλι του, που χρησιμεύει στην ρύθμιση της στάσης του κεφαλιού.....	- 25 -
Εικ. 2.3 Στιγμιότυπα από την XM2VTS βάση προσώπων	- 26 -
Εικ. 2.4 Βάση προσώπων “Yale”, έκδοση B.....	- 26 -
Εικ. 2.5 Βάση προσώπων AR	- 27 -
Εικ. 2.6 Βάση προσώπων CVL.....	- 28 -
Εικ. 2.7 INDIAN βάση προσώπων	- 28 -
Εικ. 2.8 Βάση Pointing '04	- 29 -
Εικ. 2.9 Δείγματα κινήσεων στην 1η συνεδρία.....	- 31 -
Εικ. 2.10 Αλληλουχία των κινήσεων στην 2η συνεδρία	- 32 -
Εικ. 3.1 Διάγραμμα κυκλώματος.....	- 35 -
Εικ. 3.2 Αριστερά: Κανονικό χρώμα της φωτοδιόδου σε λειτουργία. Δεξιά: Υπερεκτεθειμένη εικόνα, στην οποία το χρώμα που απεικονίζεται από την κάμερα δεν φαίνεται πράσινο αλλά σχεδόν λευκό.	- 36 -
Εικ. 3.3 Αλλαγή της φωτεινότητας της φωτοδιόδου εντός και εκτός της γωνίας δέσμης	- 36 -
Εικ. 3.4 Επάνω: Η κάθετη τοποθέτηση της φωτοδιόδου και η διαφορά στην φωτεινότητα με την χρήση του ημιδιαφανούς περιβλήματος. Κάτω: Μετά την εφαρμογή της ρύθμισης χρωματικών επιπέδων	- 37 -
Εικ. 3.6 Άποψη της διάταξης μετωπικά και προφίλ. Η φωτεινότητα παραμένει στα ίδια επίπεδα και στις δύο απόψεις.....	- 38 -
Εικ. 3.5 Σωστή τοποθέτηση της διάταξης στο κεφάλι του υποκειμένου	- 38 -
Εικ. 3.7 Σχηματικό διάγραμμα των βημάτων του αλγορίθμου	- 42 -
Εικ. 3.8 Αριστερά: Δυναδική μάσκα από τα στρώματα a^* και b^* . Δεξιά: Τελική δυναδική μάσκα μετά από το λογικό ΚΑΙ της χρωματικής μάσκας και της μάσκας φωτεινότητας από το στρώμα L.	- 44 -
Εικ. 3.9 Τελική δυναδική εικόνα του πρώτου καρέ	- 45 -
Εικ. 3.10 Το καρέ εξόδου της συνάρτησης Tracker.m.....	- 47 -

Εικ. 3.11 Τα κανάλια και οι μάσκες που δημιουργούνται από την συνάρτηση Tracker.m....	- 48 -
Εικ. 3.12 Υπολογισμός της γωνίας περιστροφής του κεφαλιού σε μοίρες στον οριζόντιο άξονα από ένα δοκιμαστικό βίντεο των 20 καρτέ. Υπενθυμίζουμε ότι η αρνητική γωνία περιστροφής στον οριζόντιο άξονα σημαίνει ότι το κεφάλι βρίσκεται δεξιά από την μετωπική θέση.....	- 50 -
Εικ. 3.13 Σχηματική απεικόνιση του υπολογισμού της γωνίας περιστροφής του κεφαλιού σε μοίρες.	- 51 -
Εικ. 3.14 Περιστροφή του κεφαλιού αριστερά της αρχικής μετωπικής στάσης. Φαίνονται τα σημεία ενδιαφέροντος από τα δύο καρτέ, το βελάκι της κατεύθυνσης μετατόπισης, η μετατόπιση σε εικονοστοιχεία, η γωνία περιστροφής και η αλλαγή της συντεταγμένης X.	- 52 -
Εικ. 4.1 Επάνω: Απενεργοποιημένη ρύθμιση χρωματικών επιπέδων. Κάτω: Ενεργοποιημένη ρύθμιση χρωματικών επιπέδων	- 57 -
Εικ. 4.2 (α) Αριστερά, Κάτω (+,+) (β) Δεξιά, Κάτω (-,+) (γ) Αριστερά, Πάνω (+,-) (δ) Δεξιά, Πάνω (-,-)	- 58 -
Εικ. 4.3 Ακούσια περιστροφή του κεφαλιού	- 59 -

Κατάλογος πινάκων

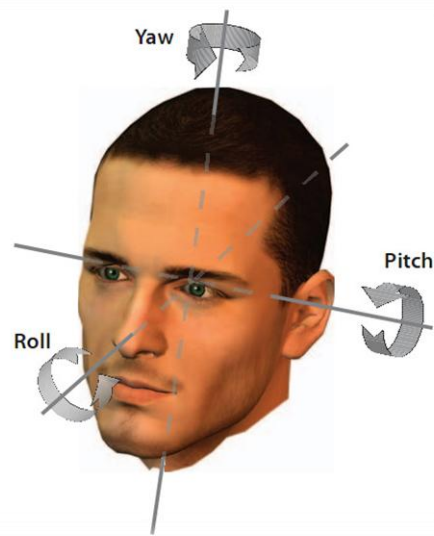
Πίνακας 2.1 Καρέ έναρξης/τερματισμού της κάθε κίνησης για κάθε βίντεο της 2 ^{ης} συνεδρίας	- 33 -
Πίνακας 2.2 Διάρκεια σε καρτέ της κάθε κίνησης για κάθε βίντεο.....	- 33 -
Πίνακας 3.1 Χρωματικοί δείκτες (color_marker).....	- 43 -
Πίνακας 3.2 Επεξήγηση προσήμανσης των τιμών του πίνακα Movement.....	- 49 -
Πίνακας 4.1 Προκαθορισμένες τιμές των παραμέτρων του αλγόριθμου	- 55 -

Κεφάλαιο 1. Εισαγωγή

Από την αρχαιότητα ο άνθρωπος έχει επιδείξει την ικανότητα να ερμηνεύει γρήγορα και αποτελεσματικά τον προσανατολισμό και κίνηση του ανθρώπινου κεφαλιού, δίνοντας του την ικανότητα να εξαγάγει συμπεράσματα για τις προθέσεις των άλλων που βρίσκονται γύρω του και να κατανοεί μία σημαντική μη προφορική μορφή επικοινωνίας. Η ευκολία με την οποία ένα άτομο πραγματοποιεί αυτό το έργο αντιπαράτίθεται στην δυσκολία ενός προβλήματος που προκαλεί τα υπολογίσιμα συστήματα για δεκαετίες τώρα. Στον τομέα της όρασης υπολογιστών, η εκτίμηση της πόζας του κεφαλιού είναι η διαδικασία εξαγωγής συμπερασμάτων για τον προσανατολισμό του ανθρώπινου κεφαλιού από ψηφιακές εικόνες. Απαιτεί μία σειρά από βήματα επεξεργασίας για τον μετασχηματισμό μιας απεικόνισης του κεφαλιού που βασίζεται σε εικονοστοιχεία (pixels) σε μία υψηλού επιπέδου αντίληψη της κατεύθυνσης. Όπως και στις υπόλοιπες υπολογιστικές διαδικασίες για την θέαση του προσώπου, ένας ιδανικός εκτιμητής της πόζας του κεφαλιού πρέπει να παρουσιάζει σταθερότητα σε μία ποικιλία παραγόντων που επηρεάζουν και αλλάζουν την εικόνα. Αυτοί οι παράγοντες συμπεριλαμβάνουν φυσικά φαινόμενα όπως η παραμόρφωση της κάμερας, η προοπτική γεωμετρία, ο δυναμικός φωτισμός από πολλές πηγές, καθώς επίσης διαφορετικά χαρακτηριστικά του προσώπου κάθε ανθρώπου, πιθανές εκφράσεις, και παρουσία εξαρτημάτων όπως γυαλιά και καπέλα.

Παρόλο που μπορεί να μοιάζουν με ρητές προδιαγραφές μίας εργασίας όρασης, η εκτίμηση της κατεύθυνσης του κεφαλιού έχει μια ποικιλία εκδοχών. Σε ένα χονδροειδές επίπεδο, η εκτίμηση της κατεύθυνσης του κεφαλιού αναφέρεται σε αλγορίθμους που αναγνωρίζουν το κεφάλι σε μία ή μερικές διακριτές θέσεις, π.χ. ανάμεσα στην μετωπική και στην αριστερή/δεξιά στάση. Σε ένα πιο λεπτομερές επίπεδο, η εκτίμηση της πόζας μπορεί να είναι μια συνεχής μέτρηση της γωνίας σε πολλούς ταυτόχρονα βαθμούς ελευθερίας (DOF – Degrees Of Freedom). Ένα σύστημα μπορεί να κάνει εκτιμήσεις μόνο σε ένα βαθμό ελευθερίας, όπως για παράδειγμα μία κίνηση του κεφαλιού από αριστερά προς τα δεξιά, αλλά υπάρχουν και πιο πολύπλοκες προσεγγίσεις που εκτιμούν μία πλήρη τρισδιάστατη πόζα του κεφαλιού στον χώρο, ενσωματώνοντας επιπλέον βαθμούς ελευθερίας που συμπεριλαμβάνουν κινήσεις των μυών του προσώπου και του σαγονιού.

Στο πλαίσιο της όρασης υπολογιστών, η εκτίμηση της πόζας του κεφαλιού ερμηνεύεται συχνά ως η ικανότητα εξαγωγής συμπερασμάτων για τον προσανατολισμό του κεφαλιού ενός ατόμου σε σχέση με την οπτική γωνία μίας κάμερας. Πιο αυστηρά, η εκτίμηση της πόζας του κεφαλιού είναι η ικανότητα εξαγωγής συμπερασμάτων σχετικά με τον προσανατολισμό ενός λείου σε συνάρτηση με ένα γενικό τμήμα συντεταγμένων, αλλά αυτή η λεπτή διαφορά απαιτεί γνώση των χαρακτηριστικών και παραμέτρων της



Εικ. 1.1 Οι τρεις βαθμοί ελευθερίας κίνησης του ανθρώπινου κεφαλιού

κάμερας ώστε να αφαιρεθεί η αντιληπτική κλίση λόγω της προοπτικής παραμόρφωσης. Το εύρος κίνησης ενός ενήλικα άντρα περιλαμβάνει μία οβελιαία κάμψη και έκταση (π.χ. κίνηση εμπρός – πίσω του λαιμού) από -60.4° έως 69.6° , μια μετωπική πλευρική κάμψη (π.χ. λύγισμα του λαιμού από τα δεξιά προς τα αριστερά) από -40.9° έως 36.3° , και μία οριζόντια αξονική περιστροφή (π.χ. περιστροφή του κεφαλιού από τα αριστερά προς τα δεξιά) από -79.8° έως 75.3° [1]. Η σχέση μεταξύ της περιστροφικής κίνησης των μυών του προσώπου και της κατεύθυνσης του κεφαλιού συνήθως δεν λαμβάνεται υπόψη (π.χ. η πλευρική άποψη του κεφαλιού δεν είναι ακριβώς η ίδια όταν η κάμερα κοιτάζει από το πλάι και όταν η κάμερα κοιτάζει μετωπικά το σώμα αλλά το κεφάλι είναι στραμμένο πλευρικά). Αν παραλειφθεί αυτό το πρόβλημα, θεωρείται συχνά ότι το ανθρώπινο κεφάλι μπορεί να μοντελοποιηθεί ως αποσπώμενο από το σώμα, στέρεο αντικείμενο. Σύμφωνα με αυτή την λογική, το ανθρώπινο κεφάλι περιορίζεται σε κίνηση σε τρεις βαθμούς ελευθερίας όσον αφορά την πόζα, που μπορούν να χαρακτηριστούν ως pitch, roll και yaw γωνίες όπως φαίνεται και στην εικ. 1.1.

Η εκτίμηση της πόζας του κεφαλιού είναι συνδεδεμένη με την εκτίμηση του βλέμματος ενός ανθρώπου, όπως για παράδειγμα η ικανότητα χαρακτηρισμού της διεύθυνσης και του σημείου εστίασης των ματιών ενός ατόμου. Από μόνη της η εκτίμηση της θέσης του κεφαλιού παρέχει μία χονδροειδή ένδειξη του βλέμματος που μπορεί να εκτιμηθεί έτσι σε περιπτώσεις που τα μάτια του ατόμου δεν είναι ορατά από την κάμερα (όπως π.χ. σε εικόνες με χαμηλή ανάλυση ή σε περιπτώσεις που τα μάτια κρύβονται από αδιαφανή υλικά όπως τα γυαλιά ηλίου). Όταν τα μάτια είναι ορατά, η πόζα του κεφαλιού είναι απαραίτητη για την ακριβή πρόβλεψη της κατεύθυνσης των ματιών. Μελέτες στον τομέα της φυσιολογίας έχουν αποδείξει ότι

η πρόβλεψη του βλέμματος ενός ανθρώπου προέρχεται από έναν συνδυασμό της πόζας του κεφαλιού και της κατεύθυνσης των ματιών [2]. Με την χρήση ψηφιακών συνθέσεων εικόνων που απεικονίζουν καθορισμένες κατευθύνσεις των ματιών σε διαφορετικές κατευθύνσεις του κεφαλιού, οι ερευνητές διαπίστωσαν ότι η ερμηνεία του βλέμματος από έναν παρατηρητή επηρεάζεται από την διεύθυνση του κεφαλιού του στόχου.

Ένα γραφικό παράδειγμα [3] αυτού του φαινομένου παρουσιάστηκε τον 19^ο αιώνα σε ένα σχέδιο όπως φαίνεται στην εικ. 1.2. Στο σχέδιο φαίνονται δύο απόψεις του κεφαλιού που παρουσιάζουν διαφορετικές κατευθύνσεις, αλλά τα μάτια έχουν σχεδιαστεί σε ταυτόσημες διατάξεις. Ρίχνοντας μία ματιά στην εικόνα, είναι ολοφάνερο ότι το βλέμμα που αντιλαμβάνεται ο παρατηρητής επηρεάζεται σημαντικά από την στάση του κεφαλιού. Αν αφαιρεθεί ολοκληρωτικά το κεφάλι και παραμείνουν μόνο τα μάτια, η κατεύθυνση που γίνεται αντιληπτή είναι ακριβώς η ίδια με αυτήν στην περίπτωση που το κεφάλι είναι σε μετωπική διάταξη.



Εικ.1.2 Παρόλο που το κεφάλι στα δύο σκίτσα έχει διαφορετική κλίση, τα μάτια βρίσκονται στην ίδια θέση [6].

Βασισμένοι σε αυτήν την παρατήρηση και στην άποψη ότι η ικανότητα εκτίμησης του ανθρώπινου βλέμματος επεξεργάζεται κατάλληλα την οπτική πληροφορία, γίνεται η υπόθεση ότι ο παθητικός αισθητήρας μίας κάμερας χωρίς καμία προγενέστερη γνώση των συνθηκών φωτισμού, έχει ανεπαρκείς πληροφορίες για να εκτιμήσει με ακρίβεια τον προσανατολισμό ενός ματιού χωρίς παράλληλη γνώση της στάσης του κεφαλιού. Για την υποστήριξη αυτής της λογικής, ας θεωρήσουμε τις αναλογίες του χιτώνα του ματιού (την λευκή περιοχή) που περιβάλλει το μάτι. Η υψηλή αντίθεση μεταξύ του χιτώνα (λευκό) και της ίριδας του ματιού είναι ευδιάκριτη από μακριά, και μπορεί να εξελιχθεί ώστε να διευκολύνει την αντίληψη του

βλέμματος [4]. Ένα μοντέλο για την εκτίμηση της κατεύθυνσης του ματιού που χρησιμοποιεί αυτή την αντίθεση θα χρειαζόταν μια εκτίμηση της στάσης του κεφαλιού έτσι ώστε να ερμηνεύσει την κατεύθυνση του βλέμματος, αφού κάθε κίνηση του κεφαλιού εισάγει μία μετατόπιση του βλέμματος, η οποία όμως δεν επηρεάζει καθόλου την αντίθεση μεταξύ χιτώννα και ίριδας. Συνεπώς, για την υπολογιστική εκτίμηση του ανθρώπινου βλέμματος σε οποιαδήποτε διάταξη, ένας ιχνηλάτης του ματιού πρέπει να συμπληρώνεται με ένα σύστημα εκτίμησης της στάσης του κεφαλιού.

1.1 Κίνητρα

Οι άνθρωποι χρησιμοποιούν τον προσανατολισμό των κεφαλιών τους για να διαβιβάζουν πλούσιες διαπροσωπικές πληροφορίες. Για παράδειγμα, ένα άτομο θα κάνει κίνηση και θα δείξει με την κατεύθυνση του κεφαλιού του, για να υποδείξει ποιος είναι ο επιθυμητός συνομιλητής του. Παρομοίως, σε έναν διάλογο η κατεύθυνση του κεφαλιού είναι ένας μη λεκτικός τρόπος επικοινωνίας μέσω του οποίου μπορεί κάποιος να προτρέψει τον ακροατή του να πάρει εκείνος τον ρόλο του ομιλητή. Έχει μεγάλη σημασία η κίνηση του κεφαλιού σε μία συζήτηση σαν μία μορφή χειρονομίας. Οι άνθρωποι γνέφουν για να δηλώσουν ότι κατανοούν τα όσα λέγονται και χρησιμοποιούν επιπλέον κινήσεις για να δηλώσουν διαφωνία, σύγχυση, ευγένεια και συγκατάθεση. Οι υπερβολικές κινήσεις του κεφαλιού είναι σαν να δείχνει κάποιος με το δάχτυλο και είναι ένας πολύ βολικός τρόπος να πει στον συνομιλητή του να κοιτάξει σε ένα συγκεκριμένο σημείο.

Εκτός των όσων μπορεί να συμπεραίνει κάποιος από την εκούσια κίνηση του κεφαλιού, υπάρχουν και πολλά άλλα που μπορεί να συνάγει παρατηρώντας το. Για παράδειγμα, οι γρήγορες κινήσεις του μπορεί να είναι ένα σημάδι έκπληξης ή πανικού. Αυτό συνήθως πυροδοτεί αντανakλαστικές αντιδράσεις σε έναν παρατηρητή, τις οποίες είναι δύσκολο να τις αγνοήσει ακόμη και αν κάτι αντίθετο έχει λεχθεί. Άλλες σημαντικές παρατηρήσεις μπορούν να γίνουν από την στάση του κεφαλιού στην περίπτωση που κάποιος έχει συγκεντρώσει το βλέμμα του. Αν δύο άνθρωποι εστιάσουν το βλέμμα τους ο ένας στον άλλον είναι συχνά σημάδι ότι πρόκειται να αρχίσουν κάποια συζήτηση. Τέτοιου είδους αμοιβαία βλέμματα μπορεί πολλές φορές να χρησιμοποιούνται ως σημάδια επίγνωσης, όπως για παράδειγμα στην περίπτωση που ένας πεζός θέλει να διασχίσει τον δρόμο, θα περιμένει πρώτα να δει ότι ο οδηγός του οχήματος που έχει σταματήσει μπροστά του τον έχει κοιτάξει και γνωρίζει την παρουσία του. Η παρατήρηση του κεφαλιού ενός ατόμου μπορεί να δώσει πληροφορίες σχετικά και με το περιβάλλον του. Εάν κάποιος μετακινήσει το κεφάλι του προς μία κατεύθυνση πιθανώς υποδεικνύει ότι υπάρχει ένα αντικείμενο ενδιαφέροντος προς την κατεύθυνση αυτή.

Όπως η αναγνώριση φωνής, που έχει ήδη ενσωματωθεί σε πολλές γνωστές τεχνολογίες, έτσι και η εκτίμηση της στάσης του κεφαλιού είναι πολύ πιθανό να γίνει ένα άμεσα διαθέσιμο εργαλείο για την γεφύρωση του χάσματος μεταξύ ανθρώπων και υπολογιστών [5].

1.2 Μέθοδοι για την εκτίμηση της στάσης του κεφαλιού

Στην ενότητα αυτό γίνεται μία προσπάθεια να οργανωθούν και να παρουσιαστούν με συνοπτικό αλλά περιεκτικό τρόπο οι ποικίλες μεθοδολογίες που χρησιμοποιούνται για την εκτίμηση της στάσης του κεφαλιού. Η ταξινόμηση των μεθόδων γίνεται βάση της θεμελιώδους προσέγγισης που υπόκειται της κάθε υλοποίησης. Με αυτόν τον τρόπο αποφεύγονται αμφιβολίες που συνήθως δημιουργούνται όταν γίνεται προσπάθεια επέκτασης της κάθε προσέγγισης εκτός των πρωτότυπων λειτουργικών ορίων της. Συνεπώς, προκύπτουν οι οκτώ παρακάτω διαφορετικές κατηγορίες, που ταξινομούν την κάθε υλοποίηση βάση της θεμελιώδους προσέγγισης που έχει χρησιμοποιηθεί για την εκτίμηση της στάσης του κεφαλιού:

- **Μέθοδοι προτύπων εμφάνισης**, συγκρίνουν μία νέα εικόνα του κεφαλιού με ένα σύνολο προτύπων (το καθένα χαρακτηρίζεται από μία διακριτή πόζα του κεφαλιού) με σκοπό να βρεθεί η εικόνα που ταιριάζει περισσότερο.
- **Μέθοδοι συστοιχίας ανιχνευτών**, που εκπαιδεύουν μία σειρά από ανιχνευτές κεφαλιού ώστε ο καθένας να είναι κλειδωμένος σε μία συγκεκριμένη πόζα και τροφοδοτούν μια διακριτή πόζα στον ανιχνευτή που την υποστηρίζει καλύτερα.
- **Μοντέλα μη- γραμμικής παλινδρόμησης**, χρησιμοποιούν εργαλεία μη-γραμμικής παλινδρόμησης για να αναπτύξουν μία λειτουργική αντιστοίχιση από την εικόνα ή τα χαρακτηριστικά γνωρίσματα σε μία εκτίμηση της στάσης του κεφαλιού.
- **Μέθοδοι ενσωμάτωσης πολυσύνθετων συνόλων**, αναζητούν πολυσύνθετα σύνολα μικρής διάστασης που μοντελοποιούν την διαφοροποίηση στην στάση του κεφαλιού. Οι νέες εικόνες μπορούν να ενσωματωθούν σε αυτά τα σύνολα και στην συνέχεια να χρησιμοποιηθούν σε ενσωματωμένο ταίριασμα ή προτύπων ή μεθόδους παλινδρόμησης.
- **Γεωμετρικοί μέθοδοι**, χρησιμοποιούν την θέση χαρακτηριστικών του προσώπου όπως τα μάτια, το στόμα και η άκρη της μύτης για να καθορίσουν την στάση από την σχετική τους θέση.
- **Μέθοδοι ιχνηλάτησης**, που ανακτούν την γενική αλλαγή στην πόζα του κεφαλιού από την παρατηρούμενη κίνηση μεταξύ των καρέ του βίντεο.
- **Υβριδικές μέθοδοι**, που συνδυάζουν μία ή περισσότερες από τις προαναφερθείσες μεθόδους ώστε να ξεπεράσουν τους περιορισμούς που έχει η κάθε μέθοδος ξεχωριστά.

1.3 Απαιτήσεις υλοποίησης σε επίπεδο εφαρμογών

Παρόλο που η εκτίμηση της στάσης του κεφαλιού θα συνεχίσει να είναι ένας συναρπαστικός χώρος για έρευνα και εξέλιξη, ο κόσμος επιθυμεί ετοιμοπαράδοτα προγράμματα γενικής χρήσης που μπορούν να χρησιμοποιηθούν σε κάθε νέα εφαρμογή. Παρακάτω παρουσιάζονται μερικά από τα βασικά σχεδιαστικά κριτήρια που πρέπει να ακολουθεί ένα τέτοιο πρόγραμμα για να ικανοποιεί τις απαιτήσεις της πλειονότητας των εφαρμογών.

- **Ακριβές:** Το σύστημα πρέπει να παρέχει μία λογική εκτίμηση της πόζας με μέσο απόλυτο σφάλμα μικρότερο από 5°.
- **Μονοκάμερο:** Πρέπει να έχει την δυνατότητα εκτίμησης της πόζας έχοντας εικόνα από μία μόνο κάμερα. Παρόλο που η ακρίβεια μπορεί να βελτιωθεί με την χρήση δύο ή περισσότερων καμερών, δεν πρέπει αυτό να είναι απαραίτητο για την λειτουργία του συστήματος.
- **Αυτόνομο:** Δεν πρέπει να απαιτείται καμία χειροκίνητη αρχικοποίηση, ανίχνευση ή εντοπισμός. Στους περιορισμούς δεν συμπεριλαμβάνεται η χρήση καθαρά ανιχνευτικών προσεγγίσεων που μετρούν την σχετική θέση του κεφαλιού σε σχέση με κάποια αρχική διαμόρφωση, και σχηματικές/γεωμετρικές προσεγγίσεις που υποθέτουν την εξαρχής γνώση κάποιων χαρακτηριστικών του προσώπου.
- **Πολυμελές:** Το σύστημα πρέπει να εκτιμά την στάση του κεφαλιού περισσότερων του ενός ατόμων σε μία εικόνα.
- **Αμετάβλητο στην αλλαγή προσώπων και φωτισμού:** Πρέπει να δουλεύει ανεπηρέαστο από τις δυναμικές μεταβολές των συνθηκών φωτισμού, του περιβάλλοντος λειτουργίας και των διαφορετικών ατόμων που το χρησιμοποιούν.
- **Ανεξάρτητο της ανάλυσης της εικόνας:** Το σύστημα πρέπει να λειτουργεί ομαλά ανεξαρτήτως της απόστασης του αντικειμένου από την κάμερα και της ανάλυσης της εικόνας ή του βίντεο. Πρέπει να είναι ανεκτικό σε κάποια ποσότητα θορύβου καθώς και σε τεχνουργήματα λόγω συμπίεσης της εικόνας.
- **Πλήρης περιοχή κίνησης του κεφαλιού:** Ο αλγόριθμος πρέπει να παρέχει ομαλή και συνεχόμενη εκτίμηση των κινήσεων του κεφαλιού σε όλους τους άξονες και για πολλούς βαθμούς ελευθερίας (DOF), ακόμα και όταν το πρόσωπο δεν κοιτάζει την κάμερα.
- **Πραγματικού χρόνου:** Τέλος το σύστημα πρέπει να πραγματοποιεί τις εκτιμήσεις σε πραγματικό χρόνο, δηλαδή να μπορεί να παρέχει αποτελέσματα σε υψηλό ρυθμό, πάνω από 30fps(καρέ ανά δευτερόλεπτο), ώστε η κίνηση που εκτιμά ο αλγόριθμος να είναι σε συγχρονισμό με αυτή που πραγματοποιεί το αντικείμενο.

Αν και κανένα σύστημα από μόνο του δεν έχει ανταποκριθεί πλήρως στα παραπάνω κριτήρια, φαίνεται ότι η λύση δεν απέχει πολύ. Είναι πολύ πιθανό με την χρήση των σημερινών μεθόδων σε μία υβριδική προσέγγιση που θα συνδυάζει δύο ή και περισσότερες προσεγγίσεις, να μπορούν να ικανοποιηθούν όλες οι παραπάνω απαιτήσεις των σύγχρονων εφαρμογών [7].

1.4 Μεθοδολογίες καταγραφής των δεδομένων του ground truth

Για την αξιολόγηση και σύγκριση των διάφορων συστημάτων για την εκτίμηση της στάσης του κεφαλιού, θεωρείται απαραίτητη μία ακριβής μέθοδος για την μέτρηση του ground truth (δεδομένων αναφοράς) των εκτιμήσεών τους. Τυπικά, το ground truth είναι απαραίτητο για την εκπαίδευση ενός αλγορίθμου για την εκτίμηση της στάσης του κεφαλιού. Στην παρακάτω λίστα περιγράφονται μερικές από τις πιο συνηθισμένες μεθόδους που χρησιμοποιούνται, ξεκινώντας από την λιγότερο και τελειώνοντας με την περισσότερο ακριβή μέθοδο.

- **Κατευθυντικές υποδείξεις:** Ένα σύνολο δεικτών τοποθετούνται σε διακριτά σημεία σε ένα δωμάτιο και κάθε άτομο - υποκείμενο πρέπει να κατευθύνει το κεφάλι του προς κάθε μια από αυτές τις θέσεις καθόσον η κάμερα λαμβάνει εικόνες [8]. Αυτή η μέθοδος είναι μια φτωχή πηγή του ground truth. Αρχικά γίνεται η υπόθεση ότι το κεφάλι του κάθε ατόμου είναι στην ακριβώς ίδια θέση στον τρισδιάστατο χώρο, τέτοια ώστε κάθε κατεύθυνση του κεφαλιού να αντιστοιχεί και στην ίδια γωνία της στάσης του κεφαλιού. Δεύτερο και κυριότερο, γίνεται η υπόθεση ότι το άτομο έχει την ικανότητα να κατευθύνει με ακρίβεια το κεφάλι του προς ένα σημείο. Δυστυχώς αυτή η κίνηση είναι ιδιαίτερα υποκειμενική και ο κόσμος συνήθως την πραγματοποιεί με μικρή ακρίβεια. Για παράδειγμα υποκειμενικά λάθη παρατηρούνται στην ευρέως χρησιμοποιούμενη βάση προσώπων *Pointing '04* που περιγράφεται σε παρακάτω κεφάλαιο [9].
- **Κατευθυντικές υποδείξεις χρήση δείκτη λείζερ:** Αυτή η μέθοδος είναι πανομοιότυπη με την προηγούμενη με την διαφορά ότι χρησιμοποιείται ένα δείκτης λείζερ τοποθετημένος στο κεφάλι του υποκειμένου [10]. Με αυτόν τον τρόπο μπορεί να κατευθύνει με μεγαλύτερη ακρίβεια το κεφάλι του στα διακριτά σημεία που έχουν χαραχθεί στον τοίχο. Παρά την αυξημένη ακρίβεια, στη μέθοδο συνεχίζεται να γίνεται η υπόθεση ότι τα κεφάλια των ατόμων - υποκειμένων βρίσκονται στην ίδια θέση στον τρισδιάστατο χώρο, έτσι ώστε η ίδια κατεύθυνση του κεφαλιού να αντιστοιχεί και στις ίδιες γωνίες της στάσης του. Στην πράξη, αυτό είναι σχετικά δύσκολο να επιτευχθεί, μιας και οι

άνθρωποι έχουν την τάση να μετακινούν ελαφρά το κεφάλι τους όταν το περιστρέφουν προς μία κατεύθυνση (shift).

- **Χειροκίνητος υπομνηματισμός:** Οι εικόνες των προσώπων παρατηρούνται από ένα άτομο που χαρακτηρίζει τις στάσεις του κεφαλιού σύμφωνα με την δική του αντίληψη. Για ένα χονδροειδές σύνολο στάσεων του κεφαλιού και για έναν μόνο βαθμό ελευθερίας αυτό μπορεί να είναι αρκετό, αλλά σε καμία άλλη περίπτωση.
- **Διατάξεις πολλαπλών καμερών:** Σε αυτήν την μέθοδο, πολλαπλές κάμερες σε γνωστές θέσεις λαμβάνουν ταυτόχρονα την στάση του κεφαλιού από διαφορετικές γωνίες. Αν έχει προσεχθεί η θέση του κεφαλιού να είναι η ίδια για κάθε γωνία περιστροφής κατά την διάρκεια των λήψεων, η μέθοδος μπορεί να εξάγει δεδομένα για το ground truth υψηλής ακρίβειας. Το μειονέκτημα είναι ότι μπορεί να εφαρμοσθεί μόνο σε κοντινές εικόνες και δεν μπορούν να εφαρμοσθούν σε ακριβείς πόζες ή βίντεο πραγματικού χρόνου.
- **Μαγνητικοί αισθητήρες:** Οι μαγνητικοί αισθητήρες όπως το *Polhemus FastTrak* και το *Flock of Birds*, λειτουργούν εκπέμποντας και μετρώντας ένας μαγνητικό πεδίο. Ο αισθητήρας μπορεί να τοποθετηθεί στο κεφάλι του υποκειμένου και να χρησιμοποιηθεί για να καθορίσει την θέση και την γωνία προσανατολισμού του [11]. Ε-



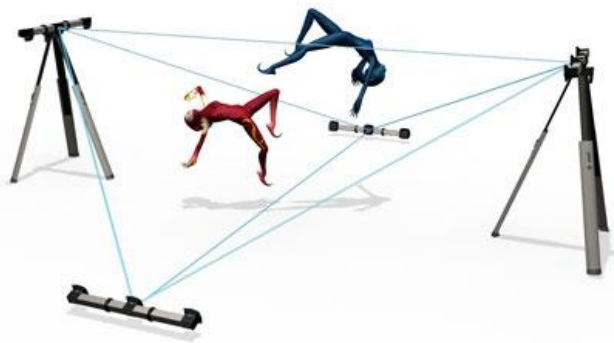
Εικ. 1.3 Τύποι μαγνητικών αισθητήρων

φόσον αυτές οι εκτιμήσεις της πόζας είναι αντικειμενικές, αποτελεί την ευρύτερα χρησιμοποιούμενη μέθοδο εξαγωγής δεδομένων για το ground truth. Τα αποτελέσματά της προσφέρουν θεωρητική ακρίβεια της τάξης της 1° , αλλά στην πράξη έχει παρατηρηθεί ότι είναι πολύ ευαίσθητα στον θόρυβο και στην παρουσία ακόμα και πολύ μικρών κομματιών μετάλλου. Συνεπώς το περιβάλλον που μπορεί να χρησιμοποιηθεί η μέθοδος είναι αυστηρά περιορισμένο και συγκεκριμένες εφαρμογές όπως η εκτίμηση της πόζας του κεφαλιού στο αυτοκίνητο είναι αδύνατον να εκμεταλλευτούν την ακρίβεια των μετρήσεων των μαγνητικών αισθητήρων.

- **Αδρανειακοί αισθητήρες:** Χρησιμοποιούν εταχυσιόμετρα, γυροσκόπια ή άλλες ές ανίχνευσης κίνησης συνήθως μαζί με ένα φίλτρο Kalman ώστε να περιορίζεται ο βος. Οι πιο φτηνοί αισθητήρες όπως ο *Mindflex InertiaCube*, δεν μετρούν την θέση, αλλά μόνο τις γωνίες προσανατολισμού σε 3 βαθμούς ελευθερίας (DOF). Το πλεονέκτημα σε σχέση με τους μαγνητικούς αισθητήρες είναι ότι δεν δέχονται παρεμβολές από μεταλλικά αντικείμενα ενώ έχουν την ίδια ωρητική ακρίβεια με τους μαγνητικούς. Για την εκτίμηση της στάσης του κεφαλιού, ο αισθητήρας μπορεί να τοποθετηθεί στο κεφάλι του υποκειμένου κατά την διάρκεια της λήψης των δεδομένων [12].



- **Οπτικά συστήματα σύλληψης κίνησης:** Τα συστήματα αυτά είναι ιδιαίτερα ικανά αλλά και ακριβά, γι' αυτό χρησιμοποιούνται συνήθως σε επαγγελματικές κινηματογραφικές παραγωγές που απαιτείται σύλληψη της κίνησης όλων των αρθρώσεων του ανθρώπινου σώματος. Συνήθως, μία συστοιχία από βαθμονομημένες κά-



Εικ. 1.4 Οπτικά συστήματα λήψης κίνησης

μερες στο κοντινό υπέρυθρο χρησιμοποιούν στερεοσκοπικούς αλγόριθμους, που δέχονται και επεξεργάζονται δεδομένα από πολλές κάμερες ταυτόχρονα, ακολουθώντας τις ανακλαστικές επιφάνειες που έχουν τοποθετηθεί πάνω σε όλο το σώμα του υποκειμένου. Για την εκτίμηση της στάσης του κεφαλιού αυτές οι επιφάνειες μπορούν να τοποθετηθούν στη πίσω πλευρά του κεφαλιού του υποκειμένου και να χρησιμοποιηθούν για την ανίχνευση της απόλυτης θέσης και κατεύθυνσης του [13]. Μερικά χαρακτηριστικά συστήματα που λειτουργούν με αυτή την τεχνική είναι τα *MX* της *Vicon* και *Visualeyez* της *Phoenix Technologies*.

1.5 Συνήθεις υποθέσεις

Ένα σύστημα εκτίμησης της στάσης του κεφαλιού για να είναι γενικής χρήσης, θα πρέπει να είναι αμετάβλητο στην ταυτότητα του υποκειμένου, να έχει επαρκές εύρος ελευθερίας κινήσεων, να μην χρειάζεται χειροκίνητες επεμβάσεις για να λειτουργήσει σωστά και θα

πρέπει εγκαθίσταται εύκολα σε συνηθισμένες συσκευές. Αν και μερικά συστήματα καλύπτουν όλες αυτές τις απαιτήσεις σε έναν βαθμό, συνήθως υποθέτουν μία ή περισσότερες συνθήκες που απλοποιούν το πρόβλημα της εκτίμησης της πόζας σε βάρος της γενικής χρησιμότητας του συστήματος. Παρακάτω αναφέρονται μερικές από τις πιο συνηθισμένες υποθέσεις που πραγματοποιούνται στην διεθνή ερευνητική βιβλιογραφία.

- **Υπόθεση συνεχούς ροής βίντεο.** Υπάρχει συνεχόμενη ροή βίντεο με μόνο μικρές αλλαγές στην πόζα μεταξύ των καρέ, έτσι ώστε να μπορεί να γίνει η εκτίμηση με αυξανόμενες σχετικές μετακινήσεις μεταξύ των καρέ.
- **Υπόθεση αρχικοποίησης.** Η στάση του κεφαλιού είναι γνωστή όταν ξεκινάει η διαδικασία εκτίμησης. Στην πράξη, το υποκείμενο πρέπει να διατηρήσει μία μετωπική πόζα μέχρι να ξεκινήσει το σύστημα εκτίμησης, ή το σύστημα περιμένει μέχρι ένας ανιχνευτής μετωπικής στάσης που ενσωματώνει, ανακαλύψει μία μετωπική στάση.
- **Αντιολισθητική υπόθεση.** Η πόζα του κεφαλιού υπολογίζεται μόνο για μία μικρή περίοδο κατά την οποία δεν πρέπει να υπάρχουν μεγάλες ανωμαλίες στην οπτική πληροφορία. Αν η υπόθεση παραβιαστεί, το σύστημα θα είναι υποκείμενο σε ολίσθηση, και θα συνεχίσει την εκτίμηση της στάσης έχοντας σημαντικά σφάλματα στις εξόδους του.
- **Υπόθεση στερεοσκοπικής θέασης.** Το υποκείμενο είναι ορατό από δύο ή περισσότερες κάμερες σε αρκετά κοντινή απόσταση ώστε να διακρίνει τις πληροφορίες βάθους κατά μήκος του προσώπου. Εναλλακτικά, οι πληροφορίες βάθους μπορούν να αποκτηθούν με άλλους εξειδικευμένους τρόπους, όπως με αισθητήρες time-of-flight [14].
- **Υπόθεση μετωπικής θέασης.** Οι διαφοροποιήσεις της στάσης περιορίζονται μόνο σε αυτές που όλα ή συγκεκριμένα κρίσιμα χαρακτηριστικά του προσώπου είναι ορατά από μετωπική οπτική γωνία.
- **Υπόθεση ενός βαθμού ελευθερίας.** Το κεφάλι επιτρέπεται να περιστρέφεται μόνο γύρω από έναν άξονα.
- **Υπόθεση γνωστής θέσης χαρακτηριστικών.** Παρέχονται στο σύστημα πληροφορίες για την θέση κάποιων χαρακτηριστικών του κεφαλιού. Αυτό τυπικά γίνεται με την χειροκίνητη επισήμανση των χαρακτηριστικών στα δεδομένα του τεστ.
- **Υπόθεση γνωστών ταυτοτήτων.** Το σύστημα λειτουργεί μόνο για το άτομο ή το σύνολο των ατόμων πάνω στα οποία έχει εκπαιδευτεί προηγουμένως.
- **Υπόθεση τεχνητών δεδομένων.** Το σύστημα λειτουργεί μόνο με τεχνητές εικόνες που δεν περιέχουν τις διαφοροποιήσεις που βρίσκονται σε ένα σύνολο εικόνων από τον πραγματικό κόσμο.

Αυτές οι υποθέσεις περιορίζουν την χρηστικότητα των συστημάτων, ακόμα και αν δείχνουν να λειτουργούν επιτυχώς σε περιορισμένο περιβάλλον. Χωρίς να λαμβάνεται υπόψη η ακρίβεια της εκτίμησης, είναι σημαντικό να αναγνωρισθούν τα συστήματα που μπορούν να εφαρμοσθούν σε πραγματικές συνθήκες όπως στην αυτοκίνηση και στην νοημοσύνη των υπολογιστών. Αυτά τα συστήματα πρέπει να παρέχουν εκτιμήσεις ανεξαρτήτως προσώπου σε τουλάχιστον δύο βαθμούς ελευθερίας χωρίς καμία χειροκίνητη παρέμβαση. Για συστήματα που εκτιμούν την στάση με χρήση διακριτών εικόνων, ο αριθμός των πάγιων στάσεων πρέπει να είναι αρκετά μεγάλος ώστε να καλύπτει όλο το εύρος των περιστροφών [7].

Κεφάλαιο 2. Βάσεις με εικόνες προσώπων

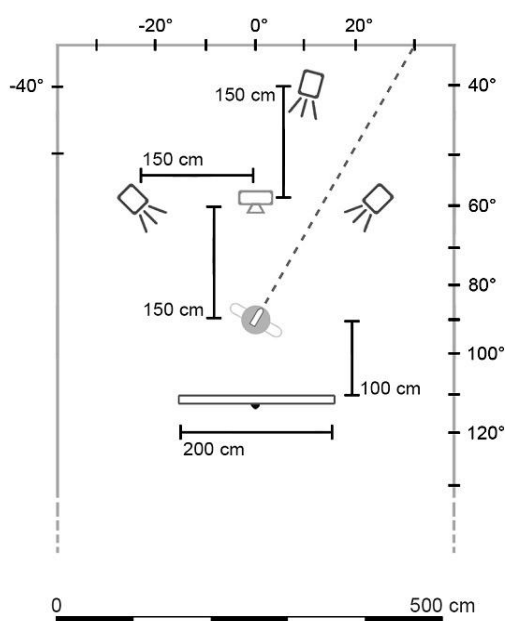
2.1 Εισαγωγή

Στο κεφάλαιο 2 παρουσιάζουμε διάφορες βάσεις εικόνων και βίντεο προσώπων που έχουν δημιουργηθεί τα τελευταία χρόνια και βοηθούν τους ερευνητές στην δοκιμή και αξιολόγηση των αλγορίθμων εκτίμησης της στάσης του κεφαλιού και του της κατεύθυνσης του βλέμματος. Οι βάσεις ποικίλουν σε μέγεθος, ποιότητα, χαρακτηριστικά και δικαιώματα χρήσης. Στις επόμενες σελίδες αναφέρουμε μερικές γνωστές βάσεις προσώπων και τις ιδιαιτερότητές τους. Στο τέλος περιγράφουμε την βάση προσώπων που δημιουργήσαμε για την αξιολόγηση του δικού μας προγράμματος και παραθέτουμε τα περιεχόμενα του συνοδευτικού οπτικού δίσκου.

2.2 Μια εκτενής βάση προσώπων

Στην βάση αυτή [15] καταγράφεται από τους ερευνητές μια εκτενής συλλογή από 20 διαφορετικά αντικείμενα που συμπεριλαμβάνουν πρόσωπα σε διαφορετικούς συνδυασμούς στάσης κεφαλιού και βλέμματος. Το αποτέλεσμα είναι ένα σύνολο από 2220 έγχρωμες εικόνες (111 ανά άτομο). Οι εικόνες έχουν δημιουργηθεί κάτω από ελεγχόμενες συνθήκες, π.χ. έχει χρησιμοποιηθεί μια τεχνική για να επιβεβαιώνεται ότι τα αντικείμενα έχουν προσαρμόσει κατάλληλα τα μάτια και το κεφάλι τους στις σωστές κατευθύνσεις. Επίσης όλες οι εικόνες έχουν χαρακτηριστεί χειροκίνητα με σημάδια που δηλώνουν σημαντικά χαρακτηριστικά του προσώπου.

Ο εξοπλισμός που έχει χρησιμοποιηθεί για την βάση είναι μια ψηφιακή φωτογραφική μηχανή (Casio Exilim 850) σε έναν συνηθισμένο χώρο εργασίας, τρεις σημειακοί προβολείς (800W ο καθένας), ένα λευκό πανί ως φόντο (200x170 cm) και ένα τρίποδο όπως φαίνεται στην εικ. 2.1. Παρέχονται λεπτομερείς πληροφορίες για το ground truth. Έχει χρησιμοποιηθεί μία τεχνική με έναν δείκτη λέιζερ για τον έλεγχο των συγκεκριμένων κατευθύνσεων του κεφαλιού και του βλέμματος. Συγκεκριμένα, ο λέιζερ δείκτης προσαρτήθηκε στο κεφάλι κάθε



Εικ. 2.1 Διαρρύθμιση του χώρου καταγραφής της βάσης

ατόμου έτσι ώστε να μπορεί να παρακολουθείται η κόκκινη κουκίδα του λέιζερ πάνω στον τοίχο και να κανονίζουν τα άτομα την κατεύθυνση του κεφαλιού τους. Επιπλέον, με την χρήση ενός λέιζερ σταθμήμετρου ενσωματωμένου με έναν μετρητή γωνιακών διαστάσεων πάνω σε ένα τρίποδο, σχεδιάστηκαν σημάδια στον τοίχο για την απεικόνιση των γωνιών στον οριζόντιο και κάθετο άξονα. Τα άτομα μπορούσαν πλέον να κινούν το κεφάλι τους μέχρι η ακτίνα του λέιζερ να δείξει σε ένα συγκεκριμένο σημείο στον τοίχο. Όταν έφταναν στην θέση αυτή τα άτομα έπρεπε να γυρίσουν σταδιακά το βλέμμα τους στο σημείο αυτό χωρίς να μετακινήσουν το κεφάλι τους (και συνεπώς την ακτίνα λέιζερ) από το συγκεκριμένο σημείο.

Ακολουθώντας αυτή την τακτική έχουν καταγραφεί για κάθε διαφορετική γωνία του κεφαλιού εννιά διαφορετικές συνθήκες βλέμματος, από -40° έως 40° με βήμα 10° σε σχέση με την επικρατούσα θέση

του κεφαλιού. Οι πόζες του κεφαλιού έχουν καταγραφεί για οριζόντια περιστροφή του κεφαλιού από 0° έως 90° με βήμα 10° . Το ίδιο έχει γίνει και για τις κάθετες πόζες του κεφαλιού για κλίση από -20° έως 20° και για τις θέσεις με οριζόντια γωνία 0° , 30° και 60° . Κάθε πόζα έχει ληφθεί σε συνδυασμό με δια-



Εικ. 2.2 Στιγμιότυπα από ένα υποκείμενο σε διάφορες θέσεις του κεφαλιού και κατευθύνσεις βλέμματος. Παρατηρούμε το δείκτη λέιζερ στο κεφάλι του, που χρησιμεύει στην ρύθμιση της στάσης του κεφαλιού.

φορετικές κάθετες αποκλίσεις του βλέμματος (-20° , 0° , $+20^\circ$) σε σχέση με την θέση του κεφαλιού.

Τέλος για να αποφευχθεί η καταγραφή του βλεφαρίσματος των ματιών έχουν γίνει τρεις συνεχόμενες λήψεις για κάθε συνδυασμό, ενώ στο τέλος διάλεξαν την καλύτερης εικόνας από τις τρεις. Η ανάλυση τους είναι 1600×1200 pixels και αποθηκεύθηκαν σε jpeg κωδικοποίηση (καλή ποιότητα).

Η βάση παρέχεται δωρεάν στο κοινό για καταφόρτωση από την ιστοσελίδα της στο διαδίκτυο.

2.3 XM2VTS βάση προσώπων

Στην βάση XM2VTS [16] έχει γίνει καταγραφή ενός μεγάλου αριθμού υψηλής ποιότητας εικόνων, ήχου και βίντεο και 3Δ μοντέλων που θα βοηθήσουν την κοινότητα των ερευνητών στην δοκιμή των αλγορίθμων για την ανίχνευση του προσώπου που έχουν αναπτύξει. Στην λήψη της βάσης 295 εθελοντές επισκέφτηκαν το στούντιο εγγραφής για τέσσερις φορές με περίπου ένα μήνα διάστημα κάθε φορά. Σε κάθε επίσκεψη γίνονταν δύο διαφορετικές καταγραφές. Στην πρώτη ο εθελοντής έλεγε κάποιες συγκεκριμένες λέξεις μπροστά στην κάμερα, ενώ στην δεύτερη έπρεπε να περιστρέψει το κεφάλι του με συγκεκριμένο τρόπο. Στην τρίτη, γινόνταν κατασκευή ενός τρισδιάστατου μοντέλου υψηλής ακρίβειας του κεφαλιού του εθελοντή με την χρήση ενός στερεοσκοπικού συστήματος.



Εικ. 2.3 Στιγμιότυπα από την XM2VTS βάση προσώπων

Ο εξοπλισμός που χρησιμοποιήθηκε για την βάση απαρτιζόταν από μία ψηφιακή κάμερα Sony VX100E και ένα βιντεοκασετόφωνο DHR1000UX. Οι λήψεις έγιναν σε χρωματική ανάλυση 4:2:0 και ακουστική συχνότητα 32kHz και ρυθμό δειγματοληψίας 16bit. Η κωδικοποίηση των εικόνων ήταν PPM και η ανάλυσή τους 720x576. Δεν γίνεται λόγος για το αν παρέχεται ή όχι ground truth.

Η βάση δεν παρέχεται δωρεάν, καθώς απαιτείται και αναφορά στην βάση και στο ερευνητικό έργο M2VTS. Όλο το υλικό της βάσης μπορεί κάποιος να το παραγγείλει από την ιστοσελίδα της βάσης και να του αποσταλεί σε ένα σετ 20 κασετών DV mini.

2.4 Βάση προσώπων “Yale”, έκδοση B

Η συγκεκριμένη βάση [17] αποτελείται από 5760 εικόνες 10 ατόμων, που το καθένα έχει καταγραφεί κάτω από 576 διαφορετικούς τρόπους θέασης (9 διαφορετικές πόζες x 64 συνθήκες φω-



Εικ. 2.4 Βάση προσώπων “Yale”, έκδοση B

τισμού). Ο συνολικός όγκος δεδομένων της βάσης είναι περίπου 1GB. Για κάθε διαφορετική στάση του κεφαλιού του κάθε ατόμου υπάρχει ένα συμπιεσμένο αρχείο περίπου 11MB μέγεθος που περιέχει τις 64 λήψεις διαφορετικού φωτισμού.

Οι φωτογραφίες είναι στην κλίμακα του γκρι (8-bit) σε ανάλυση 640x480 και έχουν αποθηκευτεί με PGM RAW κωδικοποίηση ώστε να μπορούν να φορτωθούν εύκολα στο Matlab. Οι λήψεις έγιναν με μια φωτογραφική μηχανή Sony XC-75.

Παρέχονται ακριβείς πληροφορίες για το ground truth σε ξεχωριστό αρχείο για κάθε σετ εικόνων με κατάληξη .crop και ίδιο όνομα με το αρχείο των εικόνων. Στα αρχεία αυτά περιέχονται σε μορφή πίνακα οι συντεταγμένες για το δεξί και αριστερό μάτι και για το στόμα του απεικονιζόμενου ατόμου.

Η βάση παρέχεται δωρεάν για καταφόρτωση από ftp διακομιστή για χρήση σε ερευνητικές εργασίες. Αν κάποια πειραματικά αποτελέσματα εξαχθούν με την βοήθεια αυτής της βάσης πρέπει να γίνει αναφορά στην χρήση της καθώς και στην ερευνητική εργασία των δημιουργών της.

2.5 Βάση προσώπων AR

Δημιουργοί της είναι οι Aleix Martinez και Robert Benavente στο κέντρο όρασης υπολογιστών στο U.A.B [18]. Απαρτίζεται από πάνω από 4000 έγχρωμες εικόνες 126 διαφορετικών προσώπων (70 αντρικά και 56 γυναικεία). Οι περισσότερες είναι μετωπικές λήψεις με διαφορετικές εκφράσεις, συνθήκες φωτισμού και αδιαφανή αντικείμενα, όπως γυαλιά ηλίου και



Εικ. 2.5 Βάση προσώπων AR

κασκόλ. Οι λήψεις έγιναν σε εσωτερικό χώρο υπό αυστηρά ελεγχόμενες συνθήκες και το κάθε άτομο έλαβε μέρος σε δύο συνόδους με διαφορά δύο εβδομάδων. Η ίδια φωτογραφία έχει ληφθεί και στις 2. Συνολικά υπάρχουν 26 λήψεις για κάθε άτομο με συγκεκριμένο θέμα στην καθεμία (13 στην πρώτη σύνοδο και άλλες 13 στην δεύτερη με ακριβώς την ίδια σειρά και θέμα όπως στην πρώτη).

Οι εικόνες έχουν αποθηκευτεί σε αρχεία RGB RAW και σε ανάλυση 768x576 με 24bit βάθος χρώματος. Η βάση είναι διαθέσιμη στο κοινό και μπορεί να αποκτηθεί από την ιστοσελίδα του δημιουργού. Είναι εντελώς δωρεάν για ερευνητές που επιθυμούν να δοκιμάσουν τα συστήματά τους. Απαγορεύεται αυστηρά οποιαδήποτε εμπορική έκδοση και χρήση της.

2.6 CVL βάση προσώπων

Δημιουργήθηκε από τον Peter Peer στο University of Ljubljana [19] και έχει τα εξής χαρακτηριστικά:

- Έλαβαν μέρος 114 εθελοντές
- 7 εικόνες για κάθε άτομο
- Ανάλυση των εικόνων : 640x480 pixels
- Διαμόρφωση: jpeg
- Εξοπλισμός καταγραφής: Sony Digital Mavica
- Συνθήκες φωτισμού: ομοιόμορφος , μη μεταβλητός, χωρίς χρήση φλας και χωρίς χρήση κάποιου ειδικού φόντου
- Το 90% των εθελοντών ήταν άντρες και η ηλικία τους κατά μέσο όρο 18 ετών.
- Διαφορετικές συνθήκες των λήψεων(Ground truth πληροφορίες): Προφίλ αριστερό/δεξί, γωνία 45° αριστερά/δεξιά, μετωπική, με διάφορες εκφράσεις (χαμόγελο, γέλιο) σε μετωπική στάση.
- Απαιτείται γραπτή άδεια από τους δημιουργούς της βάσης για την χρήση της για οποιονδήποτε σκοπό. Διανέμεται δωρεάν.



Εικ. 2.6 Βάση προσώπων CVL

2.7 INDIAN βάση προσώπων

Η δημιουργία της βάσης έγινε στο Indian Institute of Technology Kanpur και έλαβαν μέρος 40 εθελοντές [20]. Για κάθε άτομο έχουν καταγραφεί 11 διαφορετικές στάσεις του κεφαλιού, ενώ σε



Εικ. 2.7 INDIAN βάση προσώπων

μερικές περιπτώσεις έγιναν μερικές επιπλέον λήψεις. Όλες οι εικόνες έχουν ανοιχτόχρωμο

ομογενές φόντο και τα αντικείμενα βρίσκονται αρχικά σε μετωπική θέση. Στις επόμενες λήψεις το κάθε άτομο αλλάζει την στάση του κεφαλιού του με συγκεκριμένο τρόπο και σειρά, ενώ στις τελευταίες λήψεις εκφράζει κάποια συναισθήματα (χαμόγελο, γέλιο, λύπη). Δεν γίνεται κάποια αναφορά για το αν παρέχεται πιο λεπτομερές ground truth από τις παραπάνω πληροφορίες για την στάση του κεφαλιού κατά προσέγγιση.

Αναλυτικά οι κατευθύνσεις του κεφαλιού που περιλαμβάνονται στην βάση: Μετωπική, αριστερή, δεξιά, πάνω, κάτω, πάνω-αριστερή, πάνω-δεξιά και οι εκφράσεις χαμόγελο, γέλιο, λύπη/αηδία.

Οι εικόνες είναι αποθηκευμένες σε αρχεία jpeg, ενώ η ανάλυση τους είναι 640x480 pixels σε κλίμακα του γκρι (256 επίπεδα). Η βάση δεν είναι διαθέσιμη για εμπορική χρήση, παρά μόνο διατίθεται δωρεάν ως βοήθημα για ερευνητικούς σκοπούς.

2.8 Βάση Pointing'04

Η βάση [21] αποτελείται από 15 σέτ εικόνων. Το κάθε σέτ περιέχει 2 σειρές των 93 εικόνων του ίδιου ατόμου σε διαφορετικές στάσεις του κεφαλιού. Υπάρχουν συνολικά 15



Εικ. 2.8 Βάση Pointing '04

άτομα στην βάση, που φορούν γυαλιά ή όχι και έχουν διαφορετικούς τόνους δέρματος. Η κατεύθυνση του κεφαλιού καθορίζεται από 2 γωνίες(οριζόντια – κάθετη), που ποικίλει από -90° έως $+90^\circ$. Οι εικόνες είναι πακεταρισμένες και συμπίεσμένες σε αρχεία tar.gz που έχουν μέγεθος περίπου 1MB. Κάθε τέτοιο αρχείο περιέχει μία σειρά των 93 εικόνων. Οι εικόνες είναι αποθηκευμένες σε μορφή jpeg ενώ οι γωνίες του κεφαλιού που είναι καταγεγραμμένες για κάθε άτομο είναι οι εξής:

Κάθετες γωνίες	-90, -60, -30, -15, 0, +15, +30, +60, +90
Οριζόντιες γωνίες	-90, -75, -60, -45, -30, -15, 0, +15, +30, +45, +60, +75, +90

Για την λήψη της βάσης χρησιμοποιήθηκε το σύστημα FAME της ομάδας PRIMA στο INRIA Rhone-Alpes. Παρέχεται ground truth και να καθοριστούν οι διαφορετικές γωνίες στις στάσεις του κεφαλιού έγινε βαθμονόμηση στον τοίχο του δωματίου και τοποθετήθηκαν σημάδια στις γωνίες που αναφέρθηκαν παραπάνω ώστε να γνωρίζει κάθε άτομο που πρέπει να κοιτάζει σε κάθε νέα λήψη. Αρχικά έπρεπε να κοιτάζει ευθεία στις 0 μοίρες κάθετα και οριζόντια, ενώ στις επόμενες λήψεις έπρεπε να γυρίζει διαδοχικά το κεφάλι του σε κάθε ένα από τα 93 συνολικά σημάδια στον τοίχο.

Τέλος η συγκεκριμένη βάση προσώπων διανέμεται δωρεάν για οποιαδήποτε χρήση από τον δημιουργό της, αρκεί να γίνει αναφορά στην ερευνητική εργασία του.

2.9 Συγκεντρωτικός πίνακας

	Εικόνες/βίντεο	Ground Truth	Δωρεάν	Μεταβαλλόμενος Φωτισμός	Έγχρωμη
Εκτενής Βάση	NAI/OXI	NAI	NAI	OXI	NAI
XM2VTS	NAI/NAI	OXI	OXI	OXI	NAI
Yale έκδοση B	NAI/OXI	NAI	NAI	NAI	OXI
AR	NAI/OXI	OXI	NAI	NAI	NAI
CVL	NAI/OXI	NAI	NAI	OXI	NAI
INDIAN	NAI/OXI	OXI	NAI	OXI	OXI
Pointing'04	NAI/OXI	NAI	NAI	OXI	NAI

2.10 Περιγραφή της βάσης μας (βάση HPEG)

Η βάση προσώπων HPEG (Head Pose & Eye Gaze) που δημιουργήσαμε στο εργαστήριο αποτελείται από 20 συνολικά έγχρωμα βίντεο [22]. Συμμετείχαν 10 άτομα, και καταγράψαμε 2 βίντεο για το καθένα (δύο συνεδρίες). Για την καταγραφή των βίντεο χρησιμοποιήσαμε μία usb webcam, Hercules Dualpix Exchange, με δυνατότητα καταγραφής βίντεο μέγιστης ανάλυσης 1280 x 960 pixels, και ρυθμό ανανέωσης 30 fps. Το λογισμικό που χρησιμοποιήθηκε για την καταγραφή ήταν το Hercules Webcam Station Evolution SE. Δεν έγινε καμία αλλαγή στις ρυθμίσεις τις κάμερας όσον αφορά την ισορροπία του λευκού, τον κορεσμό των χρωμάτων, την φωτεινότητα και την αντίθεση λήψης. Τα βίντεο όλα καταγράφηκαν με ανάλυση 640 x 480 pixels, ρυθμό ανανέωσης 30 fps σε ασυμπίεστα αρχεία avi. Στην ονομασία των βίντεο ο αριθμός δηλώνει το άτομο και το γράμμα, a ή b, την πρώτη ή δεύτερη συνεδρία.



Εικ. 2.9 Δείγματα κινήσεων στην 1η συνεδρία

Στην πρώτη συνεδρία, το υποκείμενο πρέπει αρχικά να βρίσκεται σε μετωπική στάση προς την κάμερα και στην συνέχεια να κινήσει ελεύθερα το κεφάλι του προς οποιαδήποτε κατεύθυνση θέλει. Η κίνηση μπορεί να είναι είτε περιστροφική, είτε μεταφορική, είτε συνδυασμός των δύο. Ο μόνος περιορισμός που υπάρχει είναι και οι τρεις φωτοδίοδοι να είναι πάντα ορατές από την κάμερα.

Η κάμερα βρίσκεται σε απόσταση περίπου ενός μέτρου από το κεφάλι του υποκειμένου, στο ύψος των ματιών του και φαίνεται ένα μεγάλο τμήμα του σώματός του. Το φόντο καλύπτει μεγάλο μέρος της εικόνας και είναι ο πραγματικός χώρος εργασίας που έγιναν οι λήψεις με έντονη ανθρώπινη δραστηριότητα σε πολλά από αυτά, χωρίς να έχει γίνει καμία τροποποίηση ώστε να είναι πιο ευδιάκριτα τα LEDs. Συνολικά το βίντεο της πρώτης συνεδρίας διαρκεί 200 καρέ, δηλαδή περίπου 6,5 δευτερόλεπτα.

Στην δεύτερη συνεδρία, το υποκείμενο πρέπει να ακολουθήσει χρονικά την αλληλουχία κινήσεων του κεφαλιού και βλέμματος όπως φαίνεται στην εικ. 2.10. Η κάμερα βρίσκεται περίπου σε απόσταση 30 cm από το υποκείμενο, ώστε να είναι ευδιάκριτα όλα τα χαρακτηριστικά του προσώπου. Και εδώ δεν έχει γίνει καμία αλλαγή στο φόντο. Το κάθε βίντεο διαρκεί 12 δευτερόλεπτα. Οι πληροφορίες για το ground truth που έχουμε για την δεύτερη συνεδρία απεικονίζονται στους πίνακες 2.1 και 2.2. Πρόκειται για την διάρκεια της κάθε κίνησης μετρημένη σε καρέ. Για κάθε βίντεο και για κάθε κίνηση έχουν σημειωθεί το καρέ έναρξης και το καρέ τερματισμού της κίνησης. Όπως φαίνεται και από την εικ. 2.10 έχουμε έξι διαφορετικούς συνδυασμούς της στάσης του κεφαλιού και της κατεύθυνσης του βλέμματος. Έγινε προσπάθεια ώστε η κάθε μία από τις έξι κινήσεις να διαρκεί δύο δευτερόλεπτα. Ωστόσο, η πραγματική διάρκεια είναι μικρότερη γιατί μεταξύ δύο διαδοχικών κινήσεων μεσολαβεί ένα μεταβατικό χρονικό διάστημα από την μία κίνηση στην άλλη (πίνακας 2.1). Επίσης, πρέπει να σημειωθεί ότι ο φωτισμός παραμένει αμετάβλητος σε όλα τα βίντεο, ενώ δεν έχει γίνει μέτρηση της πραγματικής γωνίας περιστροφής του κεφαλιού στην δεύτερη συνεδρία. Φυσικά στην πρώτη συνεδρία δεν είναι δυνατόν να έχουμε κάποια δεδομένα μετρήσεων για το ground truth, αφού το υποκείμενο εκτελεί κινήσεις αυτόβουλα και χωρίς να ακολουθεί κάποια προκαθορισμένα μοτίβα.

0 δευτ : μετωπική στάση και κοιτάζει την κάμερα.

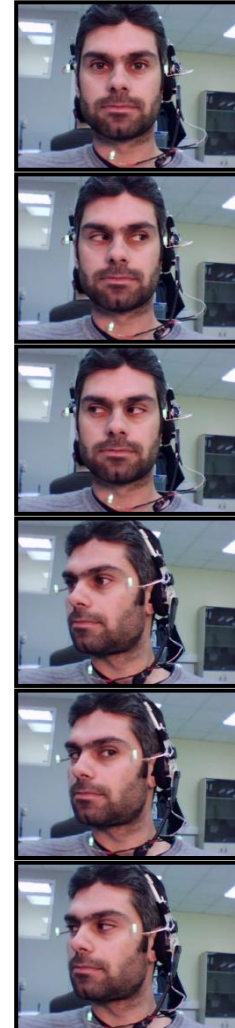
2 δευτ: μετωπική στάση, βλέμμα αριστερά

4 δευτ: μετωπική στάση, βλέμμα δεξιά

6 δευτ: στροφή κεφαλιού δεξιά, βλέμμα στην ευθεία του κεφαλιού

8 δευτ: κεφάλι στα δεξιά ακίνητο όπως πριν, βλέμμα στροφή αριστερά

10 δευτ: κεφάλι στα δεξιά ακίνητο όπως πριν, βλέμμα στροφή δεξιά



Εικ. 2.10 Αλληλουχία των κινήσεων στην 2η συνεδρία

Πίνακας 2.1 Καρέ έναρξης/τερματισμού της κάθε κίνησης για κάθε βίντεο της 2^{ης} συνεδρίας.

Name	EYES C		EYES L		EYES R		HEAD R		EYES RL		EYES RR	
	ST	END	ST	END	ST	END	ST	END	ST	END	ST	END
video1b	1	57	65	125	132	201	219	259	270	329	335	381
video2b	1	55	59	114	120	173	225	263	267	317	326	379
video3b	1	60	67	123	132	187	207	240	244	302	306	364
video4b	1	61	66	106	113	176	195	231	238	289	293	348
video5b	1	65	72	125	129	183	205	255	262	315	323	372
video6b	1	64	71	117	125	188	211	243	251	305	317	360
video7b	1	63	69	123	129	174	202	247	310	348	256	302
video8b	1	61	68	116	145	193	210	265	273	315	324	381
video9b	1	63	69	121	134	188	216	257	264	307	322	363
video10b	1	59	80	142	158	250	282	330	334	378	380	429

Πίνακας 2.2 Διάρκεια σε καρέ της κάθε κίνησης για κάθε βίντεο της 2^{ης} συνεδρίας.

NUMBER OF FRAMES PER POSE						
video1b	56	60	69	40	59	46
video2b	54	55	53	38	50	53
video3b	59	56	55	33	58	58
video4b	60	40	63	36	51	55
video5b	64	53	54	50	53	49
video6b	63	46	63	32	54	43
video7b	62	54	45	45	38	46
video8b	60	48	48	55	42	57
video9b	62	52	54	41	43	41
video10b	58	62	92	48	44	49

EYES C : Κεφάλι μετωπικά, μάτια στο κέντρο

EYES R : Κεφάλι μετωπικά, μάτια στα δεξιά

EYES L : Κεφάλι μετωπικά, μάτια στα αριστερά

HEAD R : Κεφάλι δεξιά, μάτια στο κέντρο

EYES RL : Κεφάλι δεξιά, μάτια στα δεξιά

EYES RR : Κεφάλι δεξιά, μάτια στα αριστερά

Η βάση είναι διαθέσιμη για δωρεάν καταφόρτωση και χρήση της σε ερευνητικούς σκοπούς στην ηλεκτρονική διεύθυνση <http://www.image.ece.ntua.gr/~stiast/HPEG/>.

2.11 Περιεχόμενα Οπτικού Δίσκου

Στον συνοδευτικό οπτικό δίσκο περιέχονται όλα τα απαραίτητα αρχεία για την εκτέλεση του προγράμματος καθώς και το υλικό της βάσης μαζί με τα αποτελέσματα από την ανάλυση της από το πρόγραμμα. Αναλυτικά περιέχονται:

- Ο φάκελος “Data” που περιέχει τα αρχεία με τις μεταβλητές εξόδου του προγράμματος για όλα τα βίντεο της βάσης. Τα αρχεία έχουν κατάληξη .mat και κατάλληλη ονομασία ώστε να είναι ευδιάκριτο σε ποιο βίντεο αντιστοιχούν. Είναι προσπελάσιμα από οποιαδήποτε έκδοση του Matlab και τα περισσότερα δεδομένα που περιέχουν μπορούν να εισαχθούν σε οποιαδήποτε εφαρμογή λογιστικών φύλλων ως πίνακες.
- Ο φάκελος “HPEG Database” που περιέχει δύο υποφακέλους, τους “Session 1” και “Session 2” εντός των οποίων υπάρχουν τα βίντεο από τις δύο συνεδρίες της βάσης μας. Τα βίντεο είναι συμπιεσμένα με κωδικοποίηση XviD ISO MPEG-4 με ρυθμό δεδομένων 1006 kbps. Δεν έχουν συμπεριληφθεί οι ασυμπίεστες εκδόσεις των βίντεο της βάσης εξαιτίας του μεγάλου μεγέθους τους, που συνολικά ξεπερνάει τα 7 GB σε αποθηκευτικό χώρο. Ωστόσο η υλοποίηση του αλγόριθμου σε matlab δέχεται μόνο ασυμπίεστα βίντεο, γι’ αυτό πριν την εισαγωγή τους στο πρόγραμμα για επεξεργασία θα πρέπει να μετατραπούν σε ασυμπίεστα αρχεία avi.
- Ο φάκελος “Output Videos” περιέχει τα βίντεο εξόδου του αλγορίθμου. Περιέχει και αυτός δύο υποφακέλους με ονομασίες “Session 1” και “Session 2” αντίστοιχα για την κάθε συνεδρία. Ομοίως με τον φάκελο “Head Pose Database” τα βίντεο είναι συμπιεσμένα με κωδικοποίηση XviD ISO MPEG-4 με ρυθμό δεδομένων 1006 kbps.
- Τα αρχεία “TimeStamps.pdf” και “TimeStamps.txt” που περιέχουν τις πληροφορίες του ground truth για την δεύτερη συνεδρία.
- Ο φάκελος “Matlab Scripts” που περιέχει όλα τα αρχεία που χρειάζονται για την εκτέλεση του προγράμματος. Όπως περιγράφουμε και στο κεφάλαιο 3, για την έναρξη του προγράμματος, ανοίγουμε το αρχείο “Mainscript.m” από το Matlab και το εκτελούμε.
- Την παρούσα διπλωματική εργασία σε ηλεκτρονική μορφή, αποθηκευμένη σε αρχείο τύπου pdf.

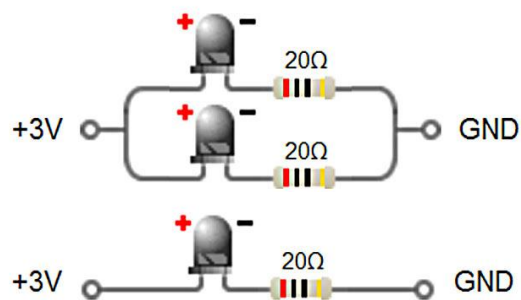
Κεφάλαιο 3. Διάταξη LEDs & αλγόριθμος

3.1 Εισαγωγή

Σε αυτό το κεφάλαιο περιγράφουμε την διάταξη των φωτοδιόδων που κατασκευάσαμε και στην συνέχεια αναλύουμε το πρόγραμμα για την ανίχνευση των LEDs και την εκτίμηση της στάσης του κεφαλιού. Το πρόγραμμα είναι γραμμένο εξ ολοκλήρου σε Matlab (έκδοση MATLAB Version 7.1.0.246 (R14) Service Pack 3), που ενσωματώνει τα Image Acquisition Toolbox Version 1.9(R14SP3), Image Processing Toolbox Version 5.1 (R14SP3) και Video and Image Processing Blockset Version 1.2 (R14SP3). Επειδή έχει γίνει εκτενής χρήση εντολών που ενσωματώνονται κυρίως σε αυτά τα toolboxes, δεν συνίσταται η εκτέλεσης των script σε παλαιότερη έκδοση του MatLab που ενσωματώνει προγενέστερες εκδόσεις των toolboxes. Προσπαθήσαμε να μην χρησιμοποιήσουμε πληθώρα προγραμματιστικών όρων στην ανάλυση, αλλά να παρουσιάσουμε την ροή του προγράμματος, τα βήματα που ακολουθήσαμε και τις δυσκολίες που συναντήσαμε στην υλοποίηση του.

3.2 Περιγραφή της διάταξης των φωτοδιόδων

Για την διάταξη των LEDs χρησιμοποιήσαμε τρεις φωτοδιόδους υψηλής φωτεινότητας (23000 mcd) πράσινου χρώματος (μήκος κύματος 525 nm) με γωνία δέσμης 15°. Τα LEDs έχουν τυπική τιμή τάσης ορθής πόλωσης $V_F = 3,4V$ και τυπική τιμή ρεύματος ορθής πόλωσης $I_F = 25\text{ mA}$. Στην περίπτωση μας δεν θέλουμε να λειτουργούν σε τόσο μεγάλη ένταση, γι' αυτό επιλέγουμε τις παρακάτω τιμές για τα στοιχεία του κυκλώματος.

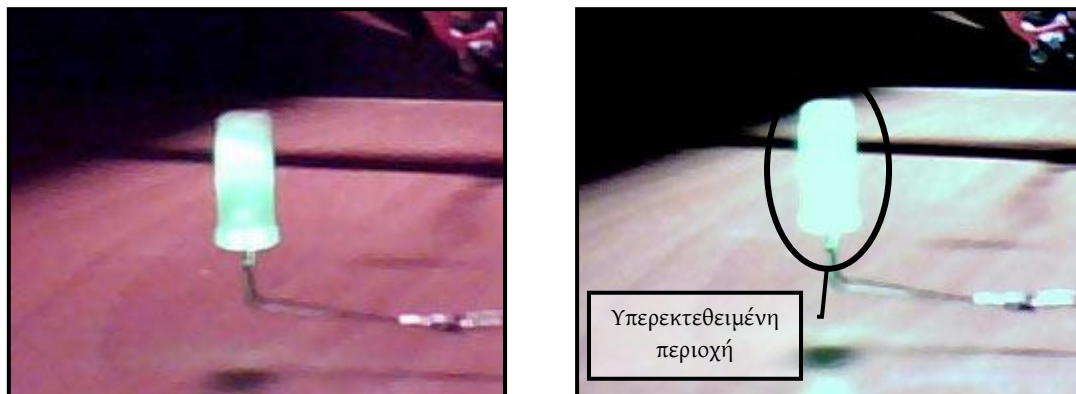


Εικ. 3.1 Διάγραμμα κυκλώματος

Τάση πόλωσης LED:	2,87 V
Ρεύμα πόλωσης LED:	6,5 mA
Τάση πηγής (μπαταρίας):	3 V
Αντίσταση :	20 Ω

Στο ένα κύκλωμα συνδέουμε παράλληλα τις 2 φωτοδιόδους ενώ στο άλλο κύκλωμα την τρίτη μόνη της (εικ. 3.1). Δεν συνδέσαμε και τα τρία LED μαζί για λόγους ευκολίας εγκατάστασης πάνω στα ακουστικά που χρησιμοποιήσαμε ως βάση. Κάθε κύκλωμα αποτελείται

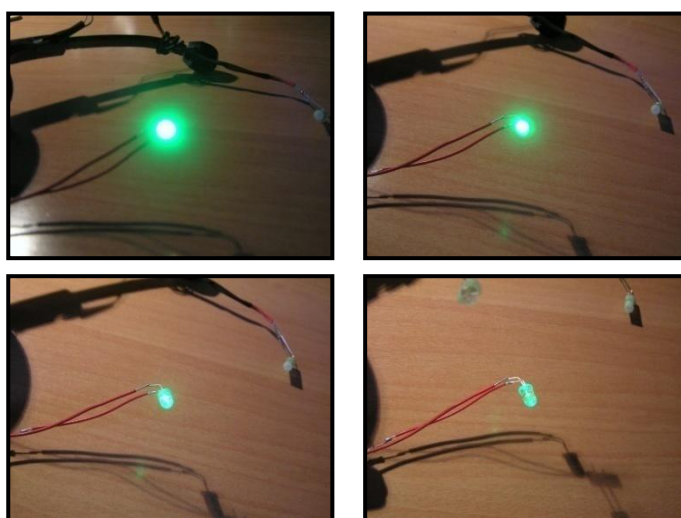
από μία μπαταρία CR2032 των 3V, την θήκη της, ένα διακόπτη δύο θέσεων (ON-OFF), τα LEDs και μία αντίσταση για κάθε LED που χρησιμεύει για την σωστή πόλωσή του στην επιθυμητή τάση.



Εικ. 3.2 **Αριστερά:** Κανονικό χρώμα της φωτοδιόδου σε λειτουργία, **Δεξιά:** Υπερεκτεθειμένη εικόνα, στην οποία το χρώμα που απεικονίζεται από την κάμερα δεν φαίνεται πράσινο αλλά σχεδόν λευκό.

Η επιλογή της αντίστασης γίνεται βάσει του τύπου $R = \frac{V_S - V_F}{I_F}$ όπου V_S είναι η τάση της πηγής, V_F η τάση πόλωσης της φωτοδιόδου και I_F το ρεύμα πόλωσης. Ο προσδιορισμός της επιθυμητής έντασης ακτινοβολίας των φωτοδιόδων πραγματοποιήθηκε με την υλοποίηση του κυκλώματος πάνω σε breadboard, με την διαφορά ότι στην θέση της αντίστασης είχαμε τοποθετήσει μία μεταβλητή αντίσταση, ώστε να μπορούμε να ρυθμίζουμε την ένταση του ρεύματος και συνεπώς την ένταση ακτινοβολίας της φωτοδιόδου. Έχοντας εγκαταστήσει το breadboard με το κύκλωμα μπροστά από μία κοινή webcam και παρακολουθώντας μέσα από την οθόνη του υπολογιστή, μεταβάλλουμε την ένταση ακτινοβολίας της φωτοδιόδου μέχρι να βρούμε μία τιμή στην οποία το LED ήταν ευδιάκριτο σε φυσικό φωτισμό δωματίου και παράλληλα δεν προκαλεί υπερεκτεθειμένες περιοχές στην εικόνα λόγω της μεγάλης έντασης ακτινοβολίας του (εικ. 3.2).

Οι φωτοδιόδοι υψηλής φωτεινότητας έχουν το μειονέκτημα της στενής γωνίας δέσμης. Οι συγκεκριμένες έχουν γωνία δέσμης 15° . Συνεπώς, καθώς πε-



Εικ. 3.3 Αλλαγή της φωτεινότητας της φωτοδιόδου εντός και εκτός της γωνίας δέσμης

ριστρέφουμε το LED η φωτεινότητά τους μειώνεται σημαντικά, και πέρα από τις 15° πέφτει κάτω από το 50% (εικ. 3.3). Εάν τα τοποθετήσουμε στο κεφάλι με τέτοιον τρόπο ώστε στην μετωπική στάση να έχουμε την μέγιστη ακτινοβολία, μετά τις 15° περιστροφής του κεφαλιού προς οποιαδήποτε κατεύθυνση, πρακτικά τα LED δεν φωτοβολούν και δεν μπορούν να ανιχνευτούν από τον ιχνηλάτη του προγράμματος. Οπότε πρέπει να τοποθετηθούν και να τροποποιηθούν κατάλληλα ώστε να μην αλλάζει η φωτεινότητά τους με την περιστροφή του κεφαλιού. Για τον λόγο αυτό, στρέφουμε την πλευρά μέγιστης φωτεινότητας του LED 90° προς τα πάνω, και το τοποθετούμε σε πλαστικό ημιδιαφανές περίβλημα παρόμοιας απόχρωσης (εικ. 3.4). Το ημιδιαφανές περίβλημα διαχέει την δέσμη του φωτός και επιφέρει την απαιτούμενη ομοιογένεια.

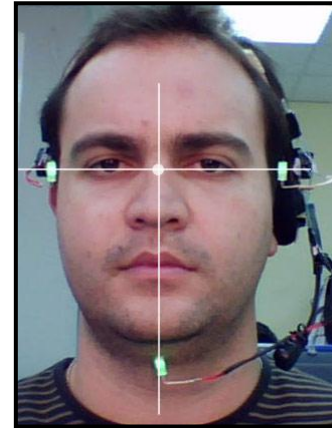


Εικ. 3.4 Επάνω: Η κάθετη τοποθέτηση της φωτοδιόδου και η διαφορά στην φωτεινότητα με την χρήση του ημιδιαφανούς περιβλήματος. **Κάτω:** Μετά την εφαρμογή της ρύθμισης χρωματικών επιπέδων

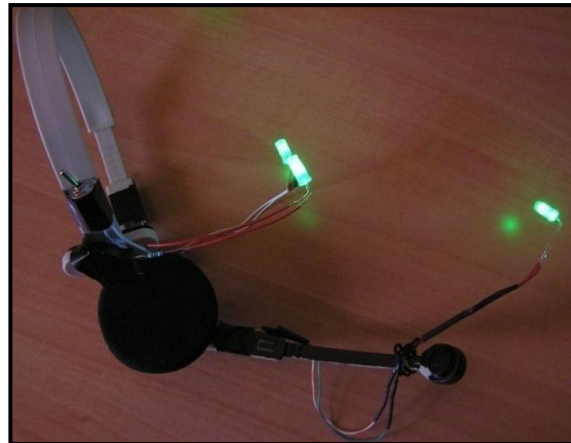


Οι φωτοδιόδοι έχουν τοποθετηθεί πάνω στην βάση με τέτοιο τρόπο ώστε οι δύο επάνω να είναι στο ύψος των ματιών και η νοητή ευθεία που τα ενώνει να είναι παράλληλη με την ευθεία των ματιών. Επίσης, δεν πρέπει να βρίσκονται πολύ μακριά από το πρόσωπο, αλλά ούτε να εφάπτονται σε αυτό, ώστε να μην αλλοιώνουν τα φυσικά χαρακτηριστικά. Το τρίτο LED πρέπει να βρίσκεται κάτω ακριβώς από το πηγούνι, σε ευθεία με την μύτη, ώστε να ισαπέχει από τα δύο πάνω (εικ. 3.5). Το σημείο τομής των δύο ευθειών που φαίνονται στην εικ. 3.5 είναι ένα σημείο ενδιαφέροντος (αναφέρεται ως Cspot στο πρόγραμμα) και χρησιμοποιείται εκτενώς σε υπολογισμούς στην διαδικασία εκτίμησης.

Η απόσταση μεταξύ των δύο πάνω φωτοδιόδων χρησιμεύει για την εκτίμηση της κίνησης στον οριζόντιο άξονα (αριστερά, δεξιά). Η απόσταση μεταξύ του σημείου που βρίσκεται στο μέσο της απόστασης των δύο πάνω (Cspot) και της κάτω φωτοδίοδου χρησιμεύει για την εκτίμηση της κίνησης του κεφαλιού στον κατακόρυφο άξονα (πάνω, κάτω). Το πρόγραμμα που δημιουργήσαμε για την ανίχνευση των φωτοδιόδων και την εκτίμηση της στάσης του κεφαλιού λειτουργεί με οποιοδήποτε μήκος κύματος της φωτεινής δέσμης των LEDs, αρκεί να εκπέμπουν και τα τρία στο ίδιο. Εμείς χρησιμοποιήσαμε πράσινα LEDs, επειδή διαφέρουν από την κόκκινη απόχρωση του δέρματος και γενικά είναι πιο δύσκολο να υπάρχουν, σε μεγάλη έκταση, επιφάνειες σε απόχρωση του πράσινου σε εσωτερικούς χώρους. Επίσης είναι πολύ συχνό κάμερες χαμηλής ποιότητας να δίνουν κόκκινη ή μπλε απόχρωση στα βίντεο που καταγράφουν αλλοιώνοντας τα πραγματικά χρώματα. Οπότε, με την χρήση φωτοδιόδων πράσινου χρώματος επιτυγχάνουμε την βέλτιστη αντίθεση με το φόντο ακόμα και με προβληματικό εξοπλισμό.



Εικ. 3.5 Σωστή τοποθέτηση της διάταξης στο κεφάλι του υποκειμένου



Εικ. 3.6 Άποψη της διάταξης μετωπικά και προφίλ. Η φωτεινότητα παραμένει στα ίδια επίπεδα και στις δύο απόψεις.

3.3 Ανάλυση του αλγορίθμου

3.3.1 Δομή

Ο αλγόριθμος αποτελείται από το βασικό script « MainScript.m» στο οποίο ξεκινάει και τελειώνει η διαδικασία ανίχνευσης των φωτοδιόδων και της εκτίμησης της στάσης του κεφαλιού. Μέσα σε αυτό υπάρχουν άλλα 4 σημαντικά scripts/συναρτήσεις, εκ των οποίων το

σημαντικότερο είναι το «Tracker.m» και στο οποίο βρίσκεται ο αλγόριθμος εκτίμησης. Συγκεκριμένα, το πρόγραμμα απαρτίζεται από τα παρακάτω αρχεία:

- **MainScript.m:** Το βασικό τμήμα του προγράμματος στο οποίο γίνεται η αρχικοποίηση της διαδικασίας ανίχνευσης των φωτοδιόδων, η ανάλυση του πρώτου καρέ και η αρχικοποίηση των πινάκων εξόδου. Από εδώ ξεκινά η εκτέλεση και των υπόλοιπων συναρτήσεων.
- **Tracker.m:** Ο κύριος αλγόριθμος για την ιχνηλάτηση των φωτοδιόδων στα καρέ του βίντεο. Εδώ πραγματοποιούνται και σημαντικές μετρήσεις και αποθηκεύονται για την τελική εκτίμηση της στάσης του κεφαλιού.
- **Estimation.m:** Συνάρτηση που εκτιμά την περιστροφή του κεφαλιού σε εικονοστοιχεία στον οριζόντιο και τον κατακόρυφο άξονα.
- **Angle.m:** Συνάρτηση που εκτιμά την περιστροφή του κεφαλιού σε μοίρες στον οριζόντιο και στον κατακόρυφο άξονα.
- **Simplesign.m:** Βοηθητική συνάρτηση που χρησιμοποιείται για την εκτίμηση της πλευράς περιστροφής και την προσήμανση της τιμής της γωνίας που έχει εκτιμήσει η συνάρτηση angle.m.

3.3.2 MainScript.m

Παρακάτω παρουσιάζονται συγκεντρωτικά σε έναν πίνακα όλες οι σημαντικές μεταβλητές του βασικού κορμού του αλγορίθμου και γίνεται μία σύντομη περιγραφή τους, ενώ στην συνέχεια αναλύονται οι διαδικασίες που περιέχει.

3.3.2.1 Περιγραφή μεταβλητών

>> Μεταβλητές Εξόδου <<		
DistanceRatios :	Διάνυσμα <(frames)x1>	Περιέχει σε κάθε στοιχείο του την τιμή του λόγου της οριζόντιας προς την κατακόρυφη απόσταση των LEDs για κάθε καρέ του βίντεο ($\frac{LEDHDistances}{LEVDistances}$)
LEDCoordinates :	Πίνακας <(frames)x6>	Περιέχει τις συντεταγμένες των τριών LEDs σε κάθε καρέ του βίντεο. Κάθε γραμμή του πίνακα αντιπροσωπεύει και ένα καρέ. Οι στήλες 1, 2 είναι οι συντεταγμένες (x, y) του LED στο δεξί μάτι, οι στήλες 3, 4 του LED κάτω από το πηγούνι και οι στήλες 5, 6 του LED στο αριστερό μάτι.

LEDCspots :	Πίνακας <(frames)x2>	Περιέχει τις συντεταγμένες του μέσου σημείου ενδιαφέροντος μεταξύ των δύο πάνω LEDs. Κάθε γραμμή αντιστοιχεί σε ένα καρέ του βίντεο και η στήλη 1 στην συντεταγμένη X και η στήλη 2 στην συντεταγμένη Y.
LEDHDistances :	Διάνυσμα <(frames)x1>	Εδώ γίνεται η καταγραφή της απόστασης (σε εικονοστοιχεία) των δύο πάνω LEDs για κάθε καρέ. Όπως και προηγουμένως η κάθε γραμμή αντιστοιχεί σε ένα καρέ.
LEDVDistances :	Διάνυσμα <(frames)x1>	Ομοίως με το διάνυσμα LEDHDistances, αλλά εδώ καταγράφεται η απόσταση του κάτω LED με το σημείο ενδιαφέροντος (Cspot).
Movement :	Πίνακας 3D <(frames-1)x1x2>	Στην Z διάσταση του τρισδιάστατου πίνακα έχουμε δύο διανύσματα. Στο πρώτο (Movement[:,1]) αποθηκεύουμε την περιστροφή του κεφαλιού μεταξύ διαδοχικών καρέ μετρημένη σε εικονοστοιχεία στον οριζόντιο X άξονα, ενώ στο δεύτερο (Movement[:,2]) την περιστροφή του κεφαλιού στον άξονα Y.
Rotation :	Πίνακας 3D <(frames-1)x1x2>	Στην z διάσταση του τρισδιάστατου πίνακα έχουμε δύο διανύσματα. Στο πρώτο (Rotation[:,1]) αποθηκεύουμε την συνολική περιστροφή του κεφαλιού από το πρώτο καρέ μετρημένη σε εικονοστοιχεία στον οριζόντιο X άξονα, ενώ στο δεύτερο (Movement[:,2]) την συνολική περιστροφή του κεφαλιού στον άξονα Y.
frames :	Ακέραιος	Ο συνολικός αριθμός των καρέ του βίντεο.
AnglesX :	Διάνυσμα <(frames)x1>	Καταγράφουμε την συνολική περιστροφή του κεφαλιού από την αρχική θέση του, στον οριζόντιο άξονα X, μετρώντας της σε μοίρες.
AnglesY :	Διάνυσμα <(frames)x1>	Καταγράφουμε την συνολική περιστροφή του κεφαλιού από την αρχική θέση του, στον κατακόρυφο άξονα Y, μετρώντας της σε μοίρες.
sgnX :	Διάνυσμα <(frames)x1>	Εδώ αποθηκεύεται η κατεύθυνση της περιστροφικής κίνησης του κεφαλιού στον οριζόντιο άξονα X. Σημειώνουμε 1, όταν το πρόγραμμα ανιχνεύει το κεφάλι να βρίσκεται προς τα αριστερά της αρχικής μετωπικής θέσης. Αντίθετα, όταν το κεφάλι βρίσκεται δεξιά σημειώνουμε με την τιμή -1.
sgnY :	Διάνυσμα <(frames)x1>	Ομοίως με το διάνυσμα sgnX αλλά στον άξονα Y. Όταν η τιμή είναι 1, το κεφάλι έχει περιστραφεί προς τα πάνω, ενώ όταν η τιμή είναι -1 έχει κινηθεί προς τα κάτω.

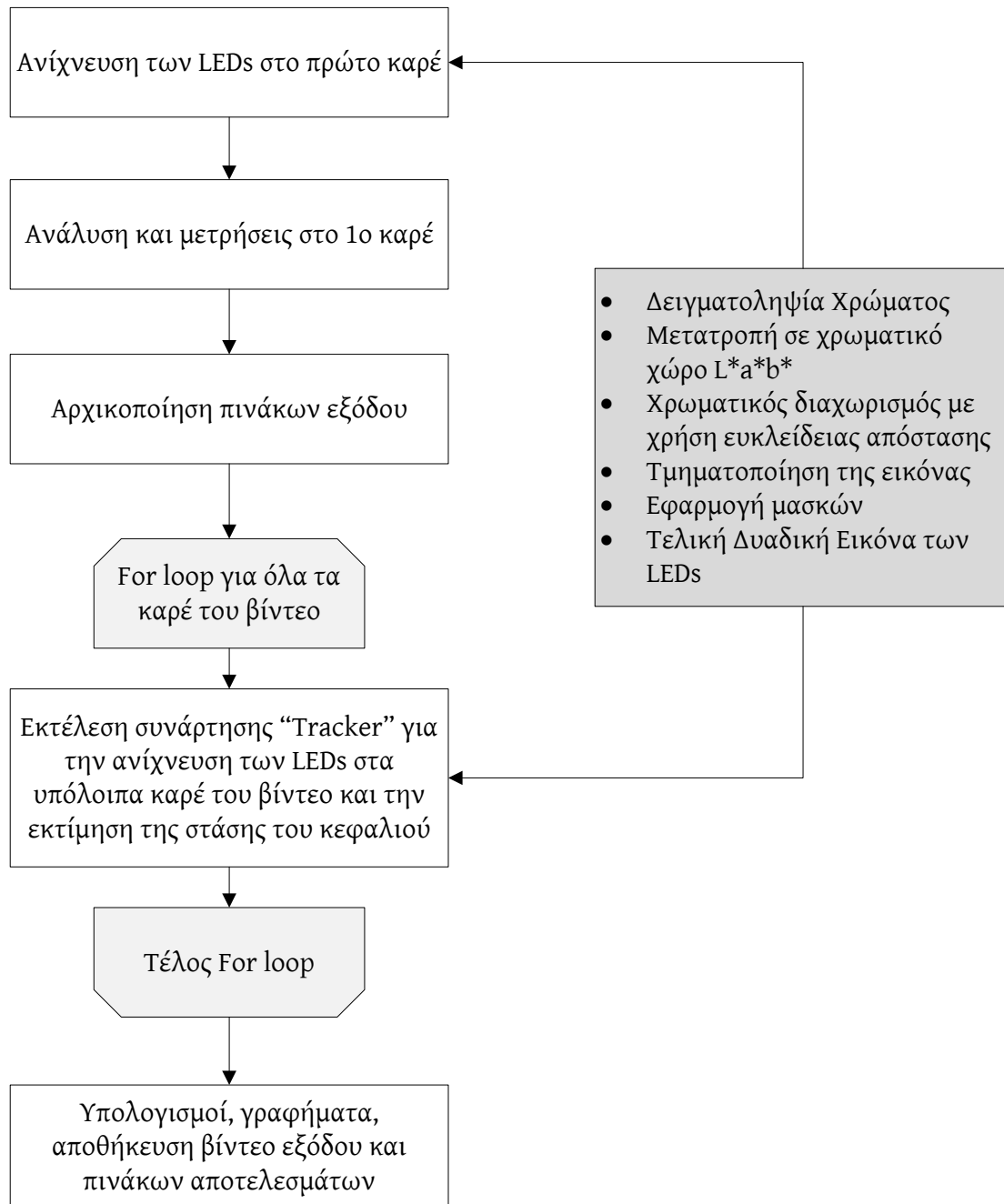
>> Μεταβλητές Εισόδου <<

videofile :	String	Εισάγουμε το όνομα του βίντεο εισόδου που θέλουμε να επεξεργαστούμε.
InitialFrame :	Εικόνα RGB (3D Πίνακας)	Εδώ αποθηκεύεται το πρώτο καρέ του βίντεο στο οποίο οριοθετούμε την περιοχή μίας φωτοδιόδου, για να μπορέσει το πρόγραμμα στην συνέχεια να τις ανιχνεύσει και να τις ιχνηλατήσει στα υπόλοιπα καρέ του βίντεο.
sample_regions :	Πίνακας 3D	Εδώ αποθηκεύονται οι μάσκες που δημιουργούμε με το ποντίκι πάνω στο αρχικό καρέ του βίντεο (InitialFrame). Η πρώτη μάσκα (sample_regions[:,1]) είναι η περιοχή δειγματοληψίας των χαρακτηριστικών αναγνώρισης των φωτοδιόδων. Η δεύτερη μάσκα είναι η περιοχή ενδιαφέροντος. Την οριοθετούμε με το ποντίκι ώστε να συμπεριλάβουμε και τις τρεις φωτοδιόδους και όσον το δυνατόν πλησιέστερα στο κεφάλι γίνεται.
>> Άλλες Μεταβλητές <<		
green :	Διαδικός Πίνακας	Η μάσκα που προκύπτει μετά την εκτέλεση του αλγόριθμου της ευκλείδειας απόστασης στο πρώτο καρέ του βίντεο.
Frame1LEDs :	Πίνακας	Η τελική εικόνα που προκύπτει μετά την ολοκλήρωση της εκτέλεσης του αλγόριθμου της τμηματοποίησης. Η περιοχή του κάθε LED είναι αριθμημένη με διαφορετικό ακέραιο αριθμό.
BadFrames :	Διάνυσμα	Περιέχει τα καρέ του βίντεο που το πρόγραμμα δεν μπόρεσε να ανιχνεύσει με επιτυχία και τις τρεις φωτοδιόδους. Χρήσιμο για την εύρεση των παραγόντων αποτυχίας του προγράμματος.
Mon :	Διάνυσμα με δομές	Εδώ αποθηκεύονται ένα προς ένα τα καρέ του βίντεο που μετά την επεξεργασία τους από το πρόγραμμα. Χρησιμοποιείται στη συνέχεια για την δημιουργία του βίντεο εξόδου.

3.3.2.2 Δειγματοληψία χρώματος – χρωματικός χώρος L*a*b*

Αρχικά εισάγουμε το όνομα του βίντεο που θέλουμε να επεξεργαστούμε και εκτελούμε το πρόγραμμα. Εμφανίζεται σε νέο παράθυρο το πρώτο καρέ του βίντεο και ορίζουμε την περιοχή ενός εκ των τριών LEDs ώστε να προσδιορισθεί ο μέσος όρος του χρώματος που έχουν τα LEDs στο συγκεκριμένο βίντεο (sample_regions[:,1]). Στην συνέχεια οριοθετούμε με το ποντίκι μία δεύτερη μάσκα (sample_regions[:,2]) που είναι η περιοχή ενδιαφέροντος του πρώτου καρέ, και χρησιμεύει στο να απορριφθεί άχρηστη πληροφορία από τον περιβάλλοντα χώρο που μπορεί να δημιουργήσει προβλήματα στον αλγόριθμο. Ο χρωματικός μέσος της

πρώτης μάσκας αποθηκεύεται και στην συνέχεια υπολογίζεται το κέντρο βάρος του χώρου που επιλέξαμε και το εμβαδό του. Στο επόμενο βήμα γίνεται η μετατροπή από τον χρωματικό χώρο που είναι καταγεγραμμένο το βίντεο στον χρωματικό χώρο $L^*a^*b^*$.



Εικ. 3.7 Σχηματικό διάγραμμα των βημάτων του αλγορίθμου

Επιλέγουμε τον συγκεκριμένο χρωματικό χώρο επειδή μας επιτρέπει να ποσοτικοποιήσουμε τις χρωματικές διαφορές που διακρίνει εύκολα το ανθρώπινο μάτι. Αποτελείται από το στρώμα φωτεινότητα L^* , το στρώμα χρωματικότητας a^* που υποδεικνύει που υπάρχει χρώμα κατά μήκος του άξονα κόκκινο – πράσινο και το στρώμα b^* που υποδεικνύει που υπάρχει χρώμα κατά μήκος του άξονα μπλε – κίτρινο. Επίσης εφαρμόζουμε ένα φίλτρο median σε κάθε στρώμα ώστε να απορρίψουμε τυχόν θόρυβο της εικόνας.

3.3.2.3 Ευκλείδεια απόσταση

Στην συνέχεια δημιουργούμε έναν πίνακα με τους χρωματικούς δείκτες. Ο ένας χρωματικός δείκτης είναι ο μέσος όρος των εικονοστοιχείων σε κάθε στρώμα της περιοχής του LED που έχουμε επιλέξει στην αρχή, και ο άλλος, ο μέσος όρος των εικονοστοιχείων της υπόλοιπης εικόνας. Οι χρωματικοί δείκτες μας ενδιαφέρουν στα στρώματα a^* και b^* , γιατί σε αυτά τα στρώματα υπάρχει αποθηκευμένη η πληροφορία του χρώματος της εικόνας. Σχηματικά δημιουργείται ο πίνακας 3.1.

Πίνακας 3.1 Χρωματικοί δείκτες (color_marker)

M.O χρώματος στο a^* της περιοχής του LED.	M.O χρώματος στο b^* της περιοχής του LED.
M.O χρώματος στο a^* της υπόλοιπης εικόνας	M.O χρώματος στο b^* της υπόλοιπης εικόνας

Στην συνέχεια ακολουθεί μία διαδικασία διαλογής των εικονοστοιχείων. Συγκεκριμένα, για κάθε εικονοστοιχείο της εικόνας, συγκρίνουμε την τιμή του στα στρώματα a^* και b^* με την τιμή των χρωματικών δεικτών και αποφασίζουμε αν ανήκει σε κάποια φωτοδίοδο ή όχι. Η απόφαση λαμβάνεται εφαρμόζοντας τον τύπο της ευκλείδειας απόστασης. Πρώτα υπολογίζουμε την ευκλείδεια απόσταση των τιμών του εικονοστοιχείου από τους χρωματικούς δείκτες της περιοχής του LED:

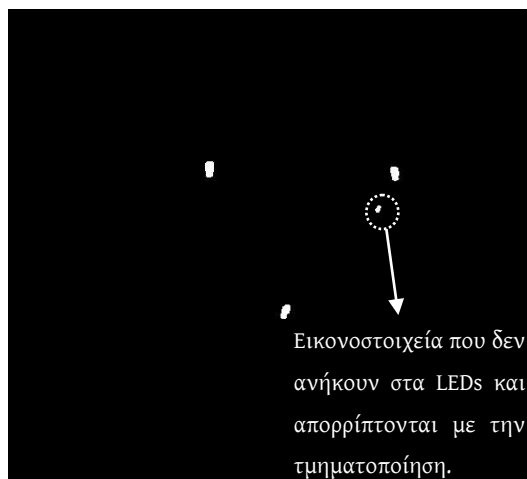
$$\sqrt{(value_{a^*} - color_marker(1,1))^2 + (value_{b^*} - color_marker(1,2))^2}$$

Στην συνέχεια την απόσταση από τους δείκτες της υπόλοιπης περιοχής:

$$\sqrt{(value_{a^*} - color_marker(2,1))^2 + (value_{b^*} - color_marker(2,2))^2}$$

Όπου $value_{a^*}$ η τιμή του εικονοστοιχείου στο a^* στρώμα, $value_{b^*}$ η τιμή του εικονοστοιχείου στο b^* στρώμα, $color_marker(x,y)$ η τιμή του πίνακα 3.1 των χρωματικών δε-

ικτών στην θέση (x, y) . Αν η πρώτη απόσταση είναι μικρότερη θεωρούμε ότι το εικονοστοιχείο ανήκει στο LED και σημειώνουμε σε έναν πίνακα, στην ίδια θέση, την τιμή 1. Αν είναι μικρότερη η δεύτερη ευκλείδεια απόσταση σημειώνουμε την θέση με την τιμή 0. Με τον τρόπο αυτό δημιουργούμε την επιθυμητή δυαδική εικόνα (μάσκα), στην οποία φαίνονται με λευκό τα εικονοστοιχεία που εκτίμησε ο αλγόριθμος ότι χρωματικά πλησιάζουν την απόχρωση των LEDs, ενώ με μαύρο τα υπόλοιπα. Η πρώτη αυτή μάσκα που δημιουργείται βασίζεται στην χρωματική διαφορά των LEDs από την υπόλοιπη εικόνα. Στην συνέχεια δημιουργούμε άλλη μία μάσκα που βασίζεται στην φωτεινότητά τους. Για τον λόγο αυτό χρησιμοποιούμε την πληροφορία που είναι αποθηκευμένη στο L στρώμα. Θέτουμε ως κάτω όριο την μέση τιμή της φωτεινότητας των εικονοστοιχείων της περιοχής του LED που επιλέξαμε στην αρχή και σημειώνουμε τα εικονοστοιχεία με μεγαλύτερη τιμή φωτεινότητας με 1, και με μικρότερη με 0. Το λογικό ΚΑΙ των δύο масκών είναι μια αρχική εικόνα της θέσης των 3 φωτοδιόδων (green). Όπως φαίνεται και στις εικόνες η μάσκα δεν διαχωρίζει τέλεια τις φωτοδιόδους και υπάρχουν και άλλα σημεία, τα οποία όμως εξαλείφονται στην συνέχεια με την διαδικασία της τμηματοποίησης.

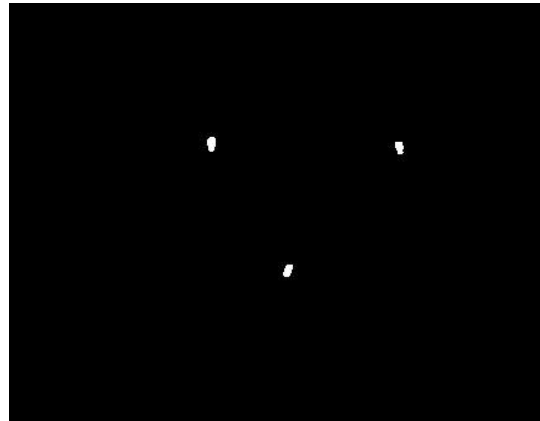


Εικ. 3.8 Αριστερά: Δυαδική μάσκα από τα στρώματα α^* και b^* . **Δεξιά:** Τελική δυαδική μάσκα μετά από το λογικό ΚΑΙ της χρωματικής μάσκας και της μάσκας φωτεινότητας από το στρώμα L.

3.3.2.4 Τμηματοποίηση (Segmentation)

Στην τμηματοποίηση, σημειώνουμε τα εικονοστοιχεία κάθε διακριτής λευκής περιοχής με έναν αύξοντα αριθμό. Για παράδειγμα, στην δεξιά εικόνα της εικ. 3.8 τα εικονοστοιχεία την πρώτης από αριστερά προς δεξιά και από πάνω προς τα κάτω λευκής περιοχής λαμβάνουν την τιμή 1, τα εικονοστοιχεία της επόμενης περιοχής την τιμή 2 κ.ο.κ. Με αυτόν τον τρόπο «βαφτίζουμε» κάθε περιοχή και μπορούμε να λάβουμε πληροφορίες για κάθε μία ξεχωριστά, όπως το εμβαδόν και το κέντρο βάρους της που μας ενδιαφέρουν. Έχοντας αποθηκεύσει τις

πληροφορίες αυτές προχωρούμε στην λυση του πρώτου καρέ, το οποίο είναι τερα κρίσιμο για την επιτυχία της ανίχνευσης των φωτοδιόδων. Θέτουμε ένα κάτω και ένα άνω όριο στην επιτρεπόμενη τιμή που μπορεί να έχει το εμβαδόν μίας λευκής περιοχής ώστε να θεωρηθεί ως LED. Για να ορίσουμε τα όρια αυτά, χρησιμοποιούμε τους συντελεστές minArea και maxArea, οι οποίοι



Εικ. 3.9 Τελική δυαδική εικόνα του πρώτου καρέ

πολλαπλασιάζονται με την τιμή του εμβαδού της περιοχής που επιλέξαμε στην αρχή. Συνεπώς όλες οι διακριτές λευκές περιοχές που είναι μικρότερες ή μεγαλύτερες από τα όρια που θέτουμε απορρίπτονται (εικ. 3.9). Το αποτέλεσμα είναι μία δυαδική εικόνα που αποτελείται μόνο από τις λευκές περιοχές που αντιπροσωπεύουν τις τρεις φωτοδιόδους.

3.3.2.5 Κλήση εξωτερικών συναρτήσεων – υπολογισμοί – γραφήματα

Ακολουθώ με την σειρά που αναφέρονται εκτελούμε τα παρακάτω βήματα:

- Αρχικοποιούμε τους πίνακες εξόδου του προγράμματος με τα αποτελέσματα από το πρώτο καρέ. Εκεί αποθηκεύουμε τις συντεταγμένες του κέντρου βάρους (LEDCoordinates) και το εμβαδό (LEDAreas) του κάθε LED, την απόσταση μεταξύ των δύο πάνω (LEDHDistances), τις συντεταγμένες του σημείου ενδιαφέροντος στο μέσο της απόστασης των δύο πάνω LEDs (LEDCSpots), την απόσταση μεταξύ του σημείου ενδιαφέροντος και του κάτω LED (LEDVDistance) και τέλος τον λόγο των δύο αποστάσεων (DistanceRatios).
- Δημιουργούμε το πρώτο καρέ του βίντεο εξόδου (Mon(1)).
- Αρχικοποιούμε τον πίνακα που θα αποθηκεύσουμε στην συνέχεια όλα τα επεξεργασμένα καρέ για να συνθέσουμε το βίντεο εξόδου.
- Δημιουργούμε τον πίνακα BadFrames, στον οποίον το πρόγραμμα αποθηκεύει τα προβληματικά καρέ κατά την διάρκεια της επεξεργασίας του βίντεο.

Με το πέρας της παραπάνω ακολουθίας έχουμε ολοκληρώσει την καταγραφή των απαραίτητων πληροφοριών του πρώτου καρέ του βίντεο. Προχωρούμε στην επεξεργασία του υπόλοιπου βίντεο, κάνοντας χρήση της συνάρτησης Tracker.m (οι λειτουργίες της περιγράφονται παρακάτω) μέσα σε μία for - loop για τον συνολικό αριθμό των εναπομεινάντων καρέ. Οι έξοδοι της συνάρτησης αποθηκεύονται στους πίνακες εξόδου του προγράμματος που είχα-

με αρχικοποιήσει προηγουμένως και συνεχίζεται η δημιουργία του βίντεο εξόδου. Παράλληλα γίνεται έλεγχος για προβληματικά καρέ, ενώ έχουν ληφθεί μέτρα για την ομαλή λειτουργία του προγράμματος σε περίπτωση που βρεθούν κάποια.

Όταν τελειώσει το for – loop έχουμε επεξεργαστεί όλο το βίντεο και έχουμε συλλέξει και αποθηκεύσει όλα τα δεδομένα που χρειαζόμαστε. Στην συνέχεια, εκτελούμε την συνάρτηση Estimation.m (αναλύεται παρακάτω) με την οποία εκτιμούμε την περιστροφή του κεφαλιού σε εικονοστοιχεία στον οριζόντιο X, και κατακόρυφο Y άξονα. Με την συνάρτηση angle.m (περιγράφεται παρακάτω) υπολογίζουμε την γωνία περιστροφής του κεφαλιού στους δύο άξονες X, Y καθώς και το πρόσημο της με την συνάρτηση simplesign.m, που υποδεικνύει την κατεύθυνση της περιστροφής (αριστερά ή δεξιά, πάνω ή κάτω). Τέλος σχεδιάζουμε κάποια γραφήματα των αποτελεσμάτων της εκτίμησης της στάσης του κεφαλιού, τα σημαντικότερα των οποίων χρησιμοποιούμε στο κεφάλαιο 4 στην ανάλυση της βάσης από το πρόγραμμα.

3.3.3 Tracker.m

3.3.3.1 Περιγραφή μεταβλητών της συνάρτησης

>> Μεταβλητές Εισόδου <<

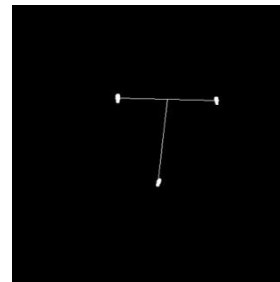
Frame :	Το καρέ του βίντεο εισόδου που θέλουμε να επεξεργαστεί η συνάρτηση.
NumFram :	Ο αριθμός του τρέχοντος καρέ.
ColorMarkers :	Οι χρωματικοί μέσοι για την περιοχή των LEDs για κάθε κανάλι του $L^*a^*b^*$ χώρου όπως έχουν προκύψει από το πρώτο καρέ.
previousCentroid :	Οι συντεταγμένες των φωτοδιόδων στο προηγούμενο καρέ του βίντεο.
previousArea :	Το εμβαδό της περιοχής της κάθε φωτοδιόδου όπως έχει μετρηθεί στο προηγούμενο καρέ.

>> Μεταβλητές Εξόδου <<

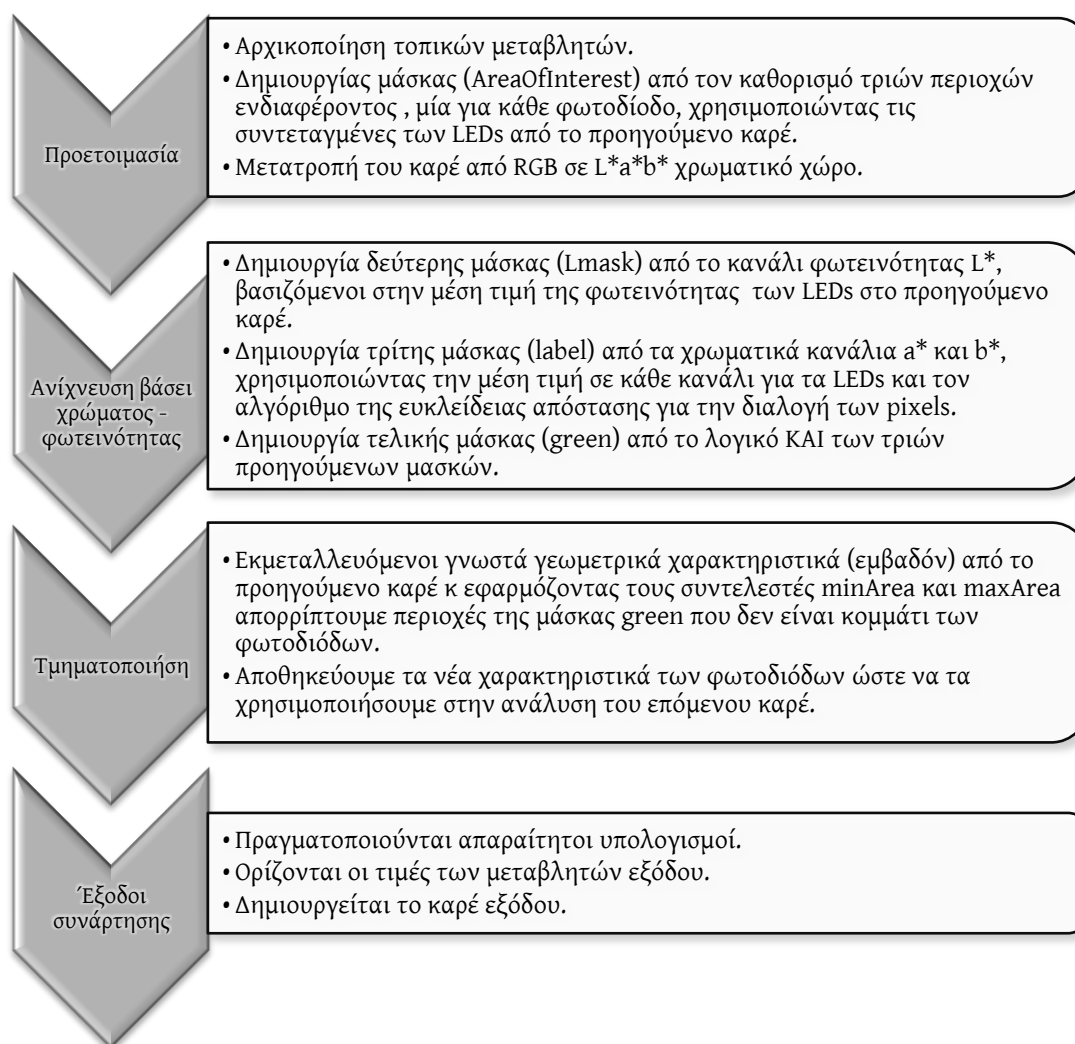
Centroid :	Οι συντεταγμένες του κέντρου βάρους της κάθε φωτοδιόδου στο συγκεκριμένο καρέ.
Area :	Το εμβαδόν κάθε φωτοδιόδου.
HDistance :	Η οριζόντια απόσταση μεταξύ των δύο πάνω φωτοδιόδων.
Cspot :	Το σημείο ενδιαφέροντος, όπως το έχουμε ορίσει και προηγουμένως, είναι το μέσο της απόστασης μεταξύ των δύο πάνω LEDs.
ratio :	Ο λόγος της οριζόντιας προς την κάθετη απόσταση όπως αυτή ορίζεται στο Mainscript.m.
BinaryImage :	Το δυαδικό καρέ εξόδου μετά την επεξεργασία του καρέ εισόδου, στο οποίο έχουμε ανιχνεύσει τις φωτοδιόδους.
VDistance :	Η απόσταση μεταξύ του κάτω LED και του Cspot.

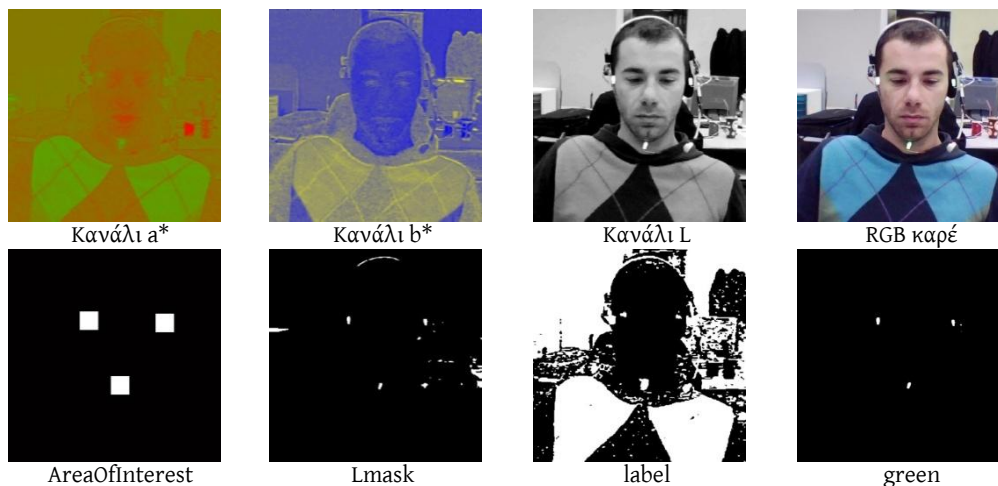
3.3.3.2 Διαδικασία ιχνηλάτησης

Η συνάρτηση Tracker.m είναι υπεύθυνη για την ιχνηλάτηση των φωτοδίοδων και την εκτέλεση μετρήσεων απαραίτητες για την ανίχνευση της στάσης του κεφαλιού. Εκτελείται μέσα σε μία for – loop από το Mainscript.m και ενημερώνει τους πίνακες εξόδου του προγράμματος κάθε φορά που ολοκληρώνει την επεξεργασία ενός καρέ του βίντεο εισόδου. Ως εισόδους δέχεται αρχικά τις πληροφορίες για τις φωτοδιόδους που έχουμε εξάγει από το πρώτο καρέ, και έπειτα τις πληροφορίες από το προηγούμενο του τρέχοντος καρέ. Η ανάλυση που γίνεται είναι πανομοιότυπη με αυτήν που πραγματοποιούμε για το πρώτο καρέ στο Mainscript.m, με την διαφορά ότι δεν χρειάζεται από τον χρήστη να ορίσει την περιοχή ενός LED και την περιοχή ενδιαφέροντος. Η διαδικασία ιχνηλάτησης είναι πλήρως αυτοματοποιημένη. Παρακάτω αναφέρουμε συνοπτικά τα βήματα εκτέλεσης του αλγόριθμου.



Εικ. 3.10 Το καρέ εξόδου της συνάρτησης Tracker.m





Εικ. 3.11 Τα κανάλια και οι μάσκες που δημιουργούνται από την συνάρτηση Tracker.m

3.3.4 Estimation.m

3.3.4.1 Περιγραφή μεταβλητών συνάρτησης

>> Μεταβλητές Εισόδου <<

- distanceR :** Πίνακας με τους λόγους των αποστάσεων $\frac{LEDHDistances}{LEDVDistances}$ από όλα τα καρέ του βίντεο.
- centerLed :** Πίνακας με τις συντεταγμένες του σημείου ενδιαφέροντος Cspot από όλα τα καρέ του βίντεο.
- numfram :** Ο συνολικός αριθμός των καρέ του βίντεο.
- threas :** Ένας δεκαδικός αριθμός, παράμετρος της συνάρτησης, που ρυθμίζει την ευαισθησία του αλγορίθμου κατά την διαδικασία διαχωρισμού της περιστροφικής από την μεταφορική κίνηση.

>> Μεταβλητές Εξόδου <<

- Movement :** Τρισδιάστατος πίνακας, που το πρώτο του στοιχείο στην διάσταση Z περιέχει την περιστροφή του κεφαλιού σε εικονοστοιχεία στον άξονα X, ενώ το δεύτερο περιστροφή στον άξονα Y.

3.3.4.2 Εκτίμηση της θέσης του κεφαλιού μετρούμενη σε εικονοστοιχεία

Η συνάρτηση Estimation.m επεξεργάζεται τον λόγο της οριζόντιας απόστασης μεταξύ των δύο πάνω φωτοδιόδων και την απόσταση του κάτω LED με το σημείο ενδιαφέροντος. Το βασικό κριτήριο για τον διαχωρισμό της περιστροφικής από την μεταφορική κίνηση είναι το μέγεθος της μεταβολής του λόγου των δύο αποστάσεων. Αν μεταξύ δύο διαδοχικών καρέ ο

λόγος μεταβάλλεται πάνω από ένα όριο (threas), τότε θεωρούμε ότι πρόκειται για περιστροφική κίνηση και σημειώνουμε στον πίνακα Movement την απόσταση που διένυσε το σημείο Cspot από το προηγούμενο στο παρόν καρέ. Διαφορετικά θέτουμε την περιστροφή ίση με μηδέν και προχωρούμε στο επόμενο. Στον πρώτο πίνακα του τρισδιάστατου πίνακα Movement σημειώνουμε την απόσταση στον Χ άξονα. Στον δεύτερο πίνακα την απόσταση στον Υ άξονα. Για τον υπολογισμό της απόστασης στον κάθε άξονα, αφαιρούμε από την τιμή της αντίστοιχης συντεταγμένης του σημείου Cspot του παρόντος καρέ, την τιμή της συντεταγμένης του προηγούμενου. Το σημείο (0,0) σε μία εικόνα είναι το πρώτο πάνω αριστερά εικοστοιχείο της. Η αρίθμηση των εικονοστοιχείων γίνεται από το matlab κατά αύξουσα σειρά, για τον άξονα Χ από αριστερά προς τα δεξιά, για τον άξονα Υ από πάνω προς τα κάτω. Συνεπώς, τα πρόσημα των τιμών που προκύπτουν στον πίνακα Movement ακολουθεί την λογική του πίνακα 3.2.

Πίνακας 3.2 Επεξήγηση προσήμανσης των τιμών του πίνακα Movement.

Μετατόπιση :	→	←	↑	↓
Πρόσημο :	+	-	-	+

3.3.4.3 Ο αλγόριθμος της συνάρτησης

Παρακάτω φαίνεται συνοπτικά το λογικό διάγραμμα του αλγορίθμου της συνάρτησης Estimation.m. Πρόκειται ακριβώς για την διαδικασία που περιγράψαμε στην προηγούμενη ενότητα.

```

IF distanceR(frame) > distanceR(frame - 1) + theas
OR
IF distanceR(frame) < distanceR(frame - 1) - theas
THEN
Movement(frame) = centerLed(frame) - centerLed(frame - 1)
ELSE
Movement(frame) = 0

```

3.3.5 Angle.m

3.3.5.1 Περιγραφή μεταβλητών συνάρτησης

>> Μεταβλητές Εισόδου <<

- Hor :** Πίνακας με τις αποστάσεις των δύο πάνω φωτοδιόδων για όλα τα καρέ του βίντεο.
- Ver :** Πίνακας με τις αποστάσεις του κάτω LED και του σημείου Cspot στο ενδιάμεσο της απόστασης των δύο πάνω LEDs, για όλα τα καρέ του βίντεο.
- distanceR :** Πίνακας με τους λόγους των αποστάσεων $\frac{LEDHDistances}{LEVDistances}$ από όλα τα καρέ του βίντεο.
- frames :** Ο συνολικός αριθμός των καρέ του βίντεο.
- threas :** Ένας δεκαδικός αριθμός, παράμετρος της συνάρτησης, που ρυθμίζει την ευαισθησία του αλγορίθμου κατά την διαδικασία διαχωρισμού της περιστροφικής από την μεταφορική κίνηση.

>> Μεταβλητές Εξόδου <<

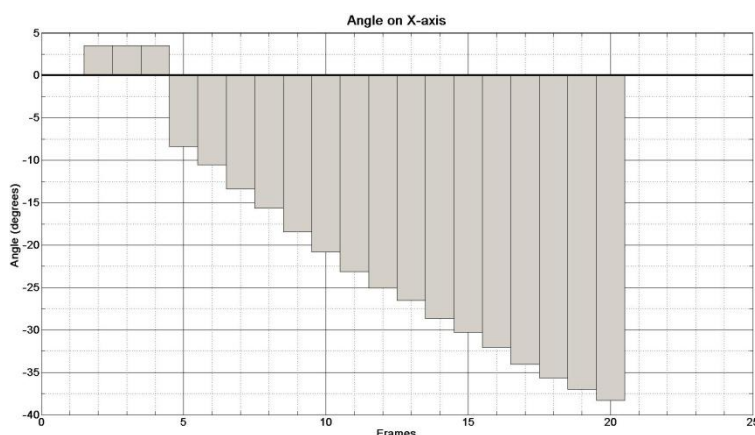
- angX :** Πίνακας που αποθηκεύονται οι τιμές της γωνίας της περιστροφής του κεφαλιού από την μετωπική στάση, στον οριζόντιο άξονα X, μετρώντας τις σε μοίρες για όλα τα καρέ του βίντεο.
- angY :** Πίνακας που αποθηκεύονται οι τιμές της γωνίας της περιστροφής του κεφαλιού από την μετωπική στάση, στον κατακόρυφο άξονα Y, μετρώντας τις σε μοίρες για όλα τα καρέ του βίντεο.

3.3.5.2 Εκτίμηση της στάσης του κεφαλιού μετρούμενη σε μοίρες

Με την βοήθεια την συνάρτησης `angle.m` εξάγουμε δύο πίνακες που περιέχουν τις γωνίες περιστροφής του κεφαλιού στους δύο άξονες κίνησης. Όπως και στην συνάρτηση `Estimation` χρησιμοποιούμε μία

παράμετρο (`threas`) που καθορίζει την ευαισθησία του προγράμματος στην διαδι-

κασία διαχωρισμού της περιστροφικής από την μεταφορική κίνηση. Μετά από δοκιμές βρήκαμε ότι έχουμε τα καλύτερα αποτελέσματα για `threas = 0,003`. Οπότε και την ορίζουμε ως προ-

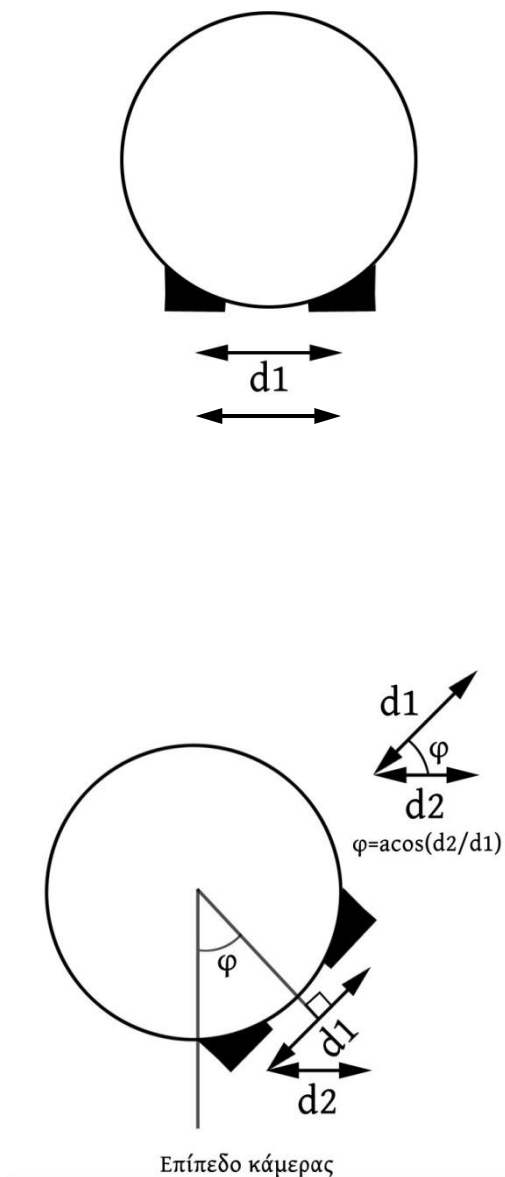


Εικ. 3.12 Υπολογισμός της γωνίας περιστροφής του κεφαλιού σε μοίρες στον οριζόντιο άξονα από ένα δοκιμαστικό βίντεο των 20 καρέ. Υπενθυμίζουμε ότι η αρνητική γωνία περιστροφής στον οριζόντιο άξονα σημαίνει ότι το κεφάλι βρίσκεται δεξιά από την μετωπική θέση.

καθορισμένη τιμή. Εν συνεχεία ακολουθώντας την ίδια λογική όπως αυτή περιγράφεται στην συνάρτηση `estimation.m` διακρίνουμε τα καρέ που πραγματοποιείται περιστροφή του κεφαλιού και αυτά που το κεφάλι παραμένει ακίνητο ή πραγματοποιεί μεταφορική κίνηση.

3.3.5.3 Υπολογισμός γωνίας περιστροφής

Για τα καρέ με την περιστροφική κίνηση υπολογίζουμε την γωνία περιστροφής με την λογική που περιγράφεται σχηματικά στην εικ. 3.12.



Εικ. 3.13 Σχηματική απεικόνιση του υπολογισμού της γωνίας περιστροφής του κεφαλιού σε μοίρες.

Στα σχήματα ως θεωρήσουμε ότι πρόκειται για την κάτοψη του κεφαλιού του ανθρώπου (κοιτάμε από πάνω προς τα κάτω). Στο πρώτο σχήμα θεωρούμε ότι το κεφάλι βρίσκεται στην μετωπική θέση, δηλαδή η γωνία φ με τον φακό της κάμερα και στους δύο άξονες είναι ίση με μηδέν. Η απόσταση μεταξύ των δύο πάνω φωτοδιόδων είναι $d1$. Η προβολή της απόστασης $d1$ πάνω στο επίπεδο του φακού της κάμερας είναι επίσης ίση με την $d1$, γιατί τα δύο ευθύγραμμα τμήματα είναι παράλληλα.

Στο δεύτερο σχήμα ο άνθρωπος έχει περιστρέψει το κεφάλι του προς τα αριστερά του σώματος του. Συνεπώς η γωνία στον άξονα X είναι διάφορη του μηδενός ενώ στον άξονα Y παραμένει μηδέν. Η απόσταση μεταξύ των φωτοδιόδων παραμένει $d1$ διότι δεν μετακινούνται από την θέση τους πάνω στο κεφάλι. Ωστόσο, η προβολή της απόστασης $d1$ πάνω στο νοητό επίπεδο του φακού της κάμερας είναι η απόσταση $d2$. Όπως φαίνεται και σχηματικά η απόσταση $d2$ είναι μικρότερη από την $d1$. Όταν λοιπόν έχουμε μεταφορική κίνηση η απόσταση $d1$ και η προβολή της $d2$ στο επίπεδο της κάμερας είναι ίσες, ενώ όταν έχουμε περιστροφική, η προβολή της είναι μικρότερη. Η ζητούμενη γωνία περιστροφής του κεφαλιού είναι η φ , η οποία είναι ίση με την γωνία που σχηματίζουν η απόσταση $d1$ με την προβολή της $d2$, γιατί όπως φαίνεται από το σχήμα οι δύο γωνίες είναι κάθετες μεταξύ τους. Οπότε υπολογίζοντας την γωνία μεταξύ των $d1$ & $d2$ έχουμε βρει την γωνία περιστροφής του κεφαλιού. Οπότε: $\varphi = \cos^{-1} \frac{d2}{d1}$

3.3.6 SimpleSign.m

3.3.6.1 Περιγραφή μεταβλητών συνάρτησης

>> Μεταβλητές Εισόδου <<

- Rot :** Πίνακας με τις περιστροφές του κεφαλιού από την αρχική μετωπική θέση. Στο πρώτο στοιχείο του πίνακα είναι αποθηκευμένες οι τιμές για την περιστροφή στον άξονα X, ενώ στο δεύτερο οι τιμές για την περιστροφή στον άξονα Y. Οι τιμές είναι σε εικονοστοιχεία.
- AngX :** Η έξοδος της συνάρτησης angle.m με τις γωνίες περιστροφής στον άξονα X χρησιμοποιείται εδώ ως είσοδος.
- AngY :** Η έξοδος της συνάρτησης angle.m με τις γωνίες περιστροφής στον άξονα Y χρησιμοποιείται εδώ ως είσοδος.

>> Μεταβλητές Εξόδου <<

- sgnX:** Πίνακας που αποτελείται μόνο από τιμές +1 και -1. Με -1 σημειώνουμε περιστροφή προς τα δεξιά από την μετωπική θέση και με +1 κίνηση προς τα αριστερά.
- sgnY :** Πίνακας που αποτελείται μόνο από τιμές +1 και -1. Με +1 σημειώνουμε την περιστροφή προς τα κάτω από την μετωπική θέση, ενώ με -1 την περιστροφή προς τα πάνω.

3.3.6.2 Καθορισμός της πλευράς περιστροφής με προσήμανση της γωνίας

Σκοπός της συνάρτησης simpleSign.m είναι να ανιχνεύει την θέση του κεφαλιού σε σχέση με την αρχική μετωπική στάση. Εκμεταλλευόμαστε τον πίνακα Rotation στον οποίο έχουμε αποθηκεύσει τις αποστάσεις που έχει μετακινηθεί το σημείο ενδιαφέροντος Cspot. Αν για παράδειγμα, στον άξονα X δούμε μετατόπιση από την αρχική θέση μηδέν πάνω από δύο εικονοστοι-

χεία (κίνηση προς τα δεξιά της εικόνας), τότε θεωρούμε ότι το κεφάλι βρίσκεται αριστερά της

μετωπικής θέσης. Αν η μετατόπιση είναι κάτω από μείον δύο εικονοστοιχεία (κίνηση προς τα αριστερά της εικόνας), τότε θεωρούμε ότι το κεφάλι βρίσκεται δεξιά της αρχικής θέσης. Επειδή είναι ανούσιο να εξετάζουμε αν το κεφάλι βρίσκεται αριστερά ή δεξιά σε μεγάλες τιμές



Εικ. 3.14 Περιστροφή του κεφαλιού αριστερά της αρχικής μετωπικής στάσης. Φαίνονται τα σημεία ενδιαφέροντος από τα δύο καρέ, το βελάκι της κατεύθυνσης μετατόπισης, η μετατόπιση σε εικονοστοιχεία, η γωνία περιστροφής και η αλλαγή της συντεταγμένης X.

γωνιών, χρησιμοποιούμε τους πίνακες AngleX και AngleY, με τις γωνίες περιστροφής στις δύο διαστάσεις, και περιορίζουμε τον έλεγχο για τιμές γωνιών κάτω των δέκα μοιρών. Για τιμές γωνιών από 10° και πάνω διατηρούμε στο τρέχον καρέ το πρόσημο του προηγούμενου. Στην εικ. 3.14 έχουμε απομονώσει δύο καρέ από ένα βίντεο της βάσης και έχουμε δημιουργήσει μία εικόνα με αυτά, σε επικάλυψη το ένα με το άλλο, ώστε να φαίνονται οι διαφορές τους. Για το μεταγενέστερο χρονικά εκ των δύο καρέ, η τιμή στον πίνακα Rotation είναι +17,34 εικονοστοιχεία και η γωνία στον άξονα X είναι $18,3^\circ$, οπότε δεν εξετάζουμε αν αλλάζει το πρόσημο και αποδίδουμε πάλι την τιμή +1 στον πίνακα sgnX, δηλαδή την ίδια τιμή που είχε το προηγούμενο καρέ.

Κεφάλαιο 4. Ανάλυση της βάσης HPEG

4.1 Εισαγωγή

Στο κεφάλαιο 4 γίνεται μία προσπάθεια να παρουσιαστούν με συνοπτικό αλλά περιεκτικό τρόπο όσο το δυνατόν καλύτερα τα αποτελέσματα του προγράμματος πάνω στην βάση προσώπων με ακολουθίες καρέ HPEG (Head Pose and Eye Gaze Dataset) που δημιουργήσαμε. Όπως γίνεται εύκολα αντιληπτό λόγω της μορφής των αποτελεσμάτων, δεν είναι εύκολο να αποτυπωθούν και να περιγραφούν σε έντυπη μορφή, γι' αυτό απαραίτητο συμπλήρωμα στην ανάλυση που ακολουθεί είναι το σύνολο των βίντεο που έχει ως έξοδο το πρόγραμμα και βρίσκονται στο συνοδευτικό υλικό. Σε αυτό το εισαγωγικό κομμάτι γίνεται παράθεση πληροφοριών σχετικά με την μορφή των αποτελεσμάτων, τις προκαθορισμένες ρυθμίσεις του προγράμματος και την ορολογία που χρησιμοποιούμε στην ανάλυση των βίντεο που ακολουθούν. Τέλος παρουσιάζονται οι παράγοντες που επηρεάζουν την απόδοση και αξιοπιστία του προγράμματος καθώς και κάποιες γενικές παρατηρήσεις.

4.1.1 Αρχεία εξόδου

Για κάθε βίντεο της βάσης υπάρχουν:

- Το **αρχικό βίντεο**, συμπίεσμένο με κωδικοποίηση XviD ISO MPEG-4 με ρυθμό δεδομένων 1006 kbps, ανάλυση 640x480 και ρυθμό καρέ 30 fps, όπως και στο ασυμπίεστο βίντεο. Με την κωδικοποίηση επιτυγχάνεται συμπίεση 220:1 ώστε να μπορούν όλα τα βίντεο να χωρέσουν στον συνοδευτικό οπτικό δίσκο.
- Το **βίντεο – έξοδος** του προγράμματος που φαίνονται τα LEDs, τα ευθύγραμμα τμήματα που μας ενδιαφέρουν και χρησιμεύουν στην ανάλυση καθώς και το τρέχων καρέ και η γωνία που εκτιμά ο αλγόριθμος εκτίμησης της γωνίας στον X και Y άξονα. Για να ξεχωρίζει από το αρχικό βίντεο το όνομά του έχει την μορφή VideoXXLEDs.avi όπου στα XX είναι ο αριθμός και το γράμμα του εκάστοτε βίντεο, π.χ. Video1aLEDs.avi.
- Ένα **αρχείο με κατάληξη .mat** και περιέχει τις μεταβλητές του matlab που έχουν αποθηκευμένα τα δεδομένα από την επεξεργασία των βίντεο από το πρόγραμμα. Συγκεκριμένα οι μεταβλητές είναι: DistanceRatios, LEDCoordinates, LEDCspots, LEDHDistances, LEDVDistances, Movemet, Rotation, frames, AnglesX, AnglesY, sgnX, sgnY.

4.1.2 Προκαθορισμένες ρυθμίσεις (Default Settings)

Στην ανάλυση των βίντεο που ακολουθεί γίνεται αναφορά στις προκαθορισμένες ρυθμίσεις των παραμέτρων του προγράμματος. Εδώ αναφέρουμε ποιες είναι αυτές οι παράμετροι και ποιες οι προκαθορισμένες τιμές τους, για τις οποίες δουλεύει χωρίς σφάλματα το πρόγραμμα στα περισσότερα βίντεο.

Πίνακας 4.1 Προκαθορισμένες τιμές των παραμέτρων του αλγόριθμου

Ρύθμιση χρωματικών επιπέδων	Απενεργοποιημένη
Συντελεστής ελάχιστου εμβαδού LED (<i>minArea</i>)	0,7
Συντελεστής μέγιστου εμβαδού LED (<i>maxArea</i>)	1,2
Περιοχή Ανίχνευσης (<i>sizeAreaOfInterest</i>)	20 pixel

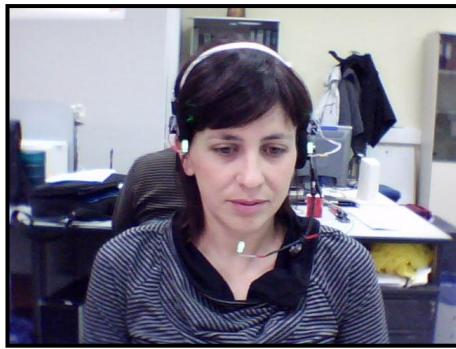
Επίσης για όλα τα βίντεο για την παράμετρο *threas* της συνάρτησης *Estimation* έχει επιλεγεί η τιμή 0,003. Υπενθυμίζουμε ότι η συγκεκριμένη παράμετρος ρυθμίζει το πόσο εύκολα η συνάρτηση που διακρίνει την μεταφορική από την περιστροφική κίνηση απορρίπτει τα καρέ της μεταφορικής.

4.1.3 Επεξήγηση όρων

Περιοχή (όρια) ασάφειας	Ως περιοχή ασάφειας θεωρούμε μια περιοχή τιμών στις εκτιμήσεις των γωνιών από -10° έως 10° . Στην περιοχή αυτή ο αλγόριθμος δεν δίνει πάντα με ακρίβεια την γωνία, κυρίως το πρόσημό της. Αυτό οφείλεται κατά κύριο λόγο στον τρόπο που γίνεται η εκτίμηση της πλευράς περιστροφής αλλά και στον τρόπο υπολογισμού της γωνίας. Υπάρχουν διάφοροι παράγοντες που εισάγουν σφάλματα στα μεγέθη που χρησιμοποιούν οι αλγόριθμοι για την εκτίμηση της γωνίας, υποβαθμίζοντας την αξιοπιστία των αποτελεσμάτων στις μικρές γωνίες. Και οι δύο αλγόριθμοι μετά από πολλές δοκιμές που πραγματοποιήσαμε απέδιδαν αξιόπιστα πάνω από 10° περιστροφής προς οποιαδήποτε πλευρά, γι' αυτό αξιολογούμε τα αποτελέσματα μόνο εκτός των ορίων, αν και μπορεί σε μερικές περιπτώσεις να είναι σωστά και εντός των ορίων.
--------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

<p>Καρέ αποτυχίας ανίχνευσης (Bad Frames)</p>	<p>Όπως έχουμε αναφέρει και στο κεφάλαιο της περιγραφής του προγράμματος υπάρχει ειδική μέριμνα για τις περιπτώσεις που ο αλγόριθμος χάνει ένα ή δύο από τα LEDs σε κάποιο καρέ. Το πρόγραμμα δεν θα καταρρεύσει και μπορεί στα επόμενα καρέ να εντοπίσει πάλι και τα τρία και να συνεχίσει χωρίς σφάλματα. Αν συμβεί μία τέτοια αποτυχία εντοπισμού των LEDs, το πρόγραμμα σημειώνει το καρέ που εντοπίστηκε το πρόβλημα σε μία λίστα που έχουμε ονομάσει Bad Frames και εμφανίζει σε ξεχωριστό παράθυρο το προβληματικό καρέ. Συνεπώς όταν γίνεται λόγος για bad frames, αναφερόμαστε σε καρέ που το πρόγραμμα δεν μπόρεσε να ανιχνεύσει και τα τρία LEDs, και βρίσκονται στον ομώνυμο πίνακα. Στο διπλανό πινακάκι φαίνεται ένα χαρακτηριστικό παράδειγμα που στο 89 καρέ το πρόγραμμα εντοπίζει μόνο τα δύο από τα τρία LED, και το σημειώνει ως bad frame, ενώ στο αμέσως επόμενο εντοπίζει και τα τρία και συνεχίζει κανονικά την εκτέλεσή του.</p> <div data-bbox="861 212 1324 689" style="border: 1px dotted black; padding: 10px;"> <pre> >>>>>>> Frame 89: <<<<<<<<< Areas Of LEDs in Frame 89: 36 32 Centroids Of LEDs in Frame 89: 298.2500 294.7222 373.0625 176.9063 >>>>>>> Frame 90: <<<<<<<<< Areas Of LEDs in Frame 90: 43 28 40 </pre> </div>
<p>Άξονες περιστροφής</p>	<p>Στα διαγράμματα που ακολουθούν κάθε βίντεο, τα δύο από αυτά παρουσιάζουν τις εκτιμήσεις της γωνίας περιστροφής του κεφαλιού σε δύο άξονες, στον X και τον Y (εκτίμηση πόζας σε 2 βαθμούς ελευθερίας). Ως X άξονα θεωρούμε τον οριζόντιο, ενώ ως Y τον κατακόρυφο. Στην αρχική θέση του υποκειμένου και εφόσον τα LEDs έχουν τοποθετηθεί προσεκτικά και με ακρίβεια, πρέπει η νοητή ευθεία που ενώνει τα δύο πάνω να είναι παράλληλη με τον οριζόντιο άξονα X, ενώ η νοητή ευθεία που ενώνει το μέσο της απόστασης μεταξύ των δύο πάνω με το τρίτο κάτω LED πρέπει να είναι παράλληλη με τον άξονα Y. Το μηδέν και στους δύο άξονες θεωρούμε ότι επιτυγχάνεται όταν το υποκείμενο βρίσκεται σε μετωπική πόζα (αρχική θέση).</p>
<p>Πρόσημο στα γραφήματα</p>	<p>Για τον προσδιορισμό της κατεύθυνσης περιστροφής χρησιμοποιούμε προσημασμένες τιμές, και στις εκτιμήσεις σε εικονοστοιχεία (πρώτο διάγραμμα σε κάθε βίντεο) αλλά και στις εκτιμήσεις της γωνίας σε Y και X άξονα (δεύτερο και τρίτο διάγραμμα). Συγκεκριμένα για τον X άξονα όταν</p>

	<p>το υποκείμενο έχει περιστρέψει το κεφάλι του αριστερά ο αλγόριθμος επιστρέφει θετικό πρόσημο, ενώ όταν το έχει περιστρέψει δεξιά αρνητικό. Στον Υ άξονα, αν το κεφάλι βρίσκεται προς τα πάνω από την μετωπική αρχική πόζα χρησιμοποιούμε αρνητικό, ενώ όταν βρίσκεται προς τα κάτω θετικό πρόσημο. Διευκρινίζουμε ότι το πρόσημο δεν έχει να κάνει με την στιγμιαία κατεύθυνση περιστροφής, αλλά με τον προσανατολισμό του κεφαλιού σε σχέση με την μετωπική πόζα, που είναι και το σημείο μηδέν για τις εκτιμήσεις (εικ.4.2).</p>
<p>Ρύθμιση χρωματικών επιπέδων</p>	<p>Με την ρύθμιση χρωματικών επιπέδων μπορούμε να βελτιώσουμε την αντίθεση των LEDs ώστε να διακρίνονται πιο εύκολα από τον ιχνηλάτη του προγράμματος. Είναι πολύ χρήσιμη επιλογή σε βίντεο ιδιαίτερα φωτεινά ή σε βίντεο με πολλές φωτεινές επιφάνειες (εικ.4.1).</p> <div data-bbox="505 844 1394 1176" data-label="Image"> </div> <p>Εικ. 4.1 Επάνω: Απενεργοποιημένη ρύθμιση χρωματικών επιπέδων. Κάτω: Ενεργοποιημένη ρύθμιση χρωματικών επιπέδων</p>
<p>minArea, maxArea, sizeAreaOfInterest</p>	<p>Όπως έχουμε αναφέρει και στην παράγραφο των προκαθορισμένων τιμών, πρόκειται για παραμέτρους του προγράμματος. Οι παράμετροι minArea και maxArea είναι συντελεστές που πολλαπλασιάζουμε το εμβαδό κάθε LED που έχει μετρηθεί από το προηγούμενο καρέ, για να θέσουμε ένα ελάχιστο και μέγιστο όριο αντίστοιχα του εμβαδού που πρέπει να έχει μία περιοχή ώστε να θεωρηθεί ότι είναι LED στο παρών καρέ. Όσο πιο κοντά στην μονάδα είναι οι συντελεστές, τόσο πιο αυστηρά είναι τα όρια. Η παράμετρος sizeAreaOfInterest καθορίζει σε pixel το μέγεθος της περιοχής ανίχνευσης γύρω από κάθε φωτοδίοδο, την ορίζουμε χειροκίνητα πριν την εκτέλεση του προγράμματος και δεν μεταβάλλεται καθόλη την διάρκεια της εκτέλεσης.</p>



(α)



(β)



(γ)



(δ)

Εικ. 4.2 (α) Αριστερά, Κάτω (+,+) (β) Δεξιά, Κάτω (-,+) (γ) Αριστερά, Πάνω (+,-) (δ) Δεξιά, Πάνω (-,-)

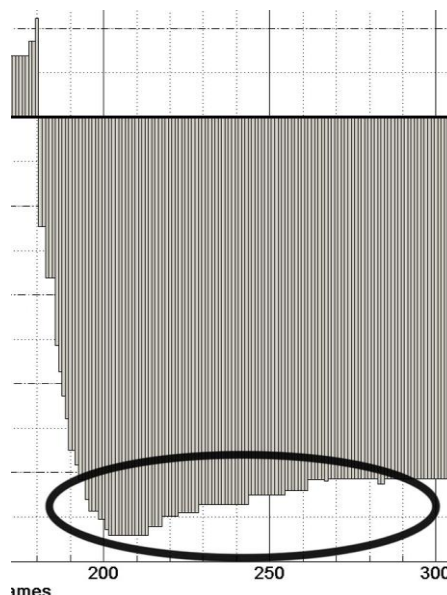
4.1.4 Παράγοντες που επηρεάζουν τα αποτελέσματα

Όπως έχουμε ήδη αναφέρει λόγω της ύπαρξης σφαλμάτων στις εκτιμήσεις του προγράμματος αναγκαστήκαμε να θεωρήσουμε μια περιοχή ασάφειας για τις μικρές γωνίες περιστροφής. Αυτό είναι μία σημαντική συνέπεια διαφόρων παραγόντων που επηρεάζουν την σωστή λειτουργία του προγράμματος και τους οποίους αναφέρουμε παρακάτω.

- **Τοποθέτηση των φωτοδιόδων.** Είναι σημαντικό τα LEDs να τοποθετηθούν με ακρίβεια στην σωστή θέση. Πρέπει η ευθεία που ενώνει τα δύο πάνω LEDs να είναι παράλληλη με την ευθεία των ματιών και όσο το δυνατόν πλησιέστερα σε αυτά. Επίσης το κάτω LED πρέπει να βρίσκεται ακριβώς κάτω από το πιγούνι. Στις περισσότερες των περιπτώσεων τα LED δεν τοποθετούνται από το υποκείμενο με ακρίβεια αλλά στο περίπου, οπότε οι μετρήσεις χάνουν σε ευαισθησία και ο αλγόριθμος που αποφασίζει για το πρόσημο δίνει αποτελέσματα με αυξημένο αριθμό σφαλμάτων.
- **Μεταβολή εμβαδού φωτοδιόδου.** Στο πρώτο στάδιο του προγράμματος που γίνεται η ανίχνευση των LEDs υπάρχει το ενδεχόμενο η επιφάνεια του κάθε LED που ανιχνεύεται να διαφέρει από αυτή που είχε μετρηθεί στο προηγούμενο, ακόμα και όταν το υποκείμενο είναι ακίνητο. Επειδή η επιφάνεια που υπολογίζεται για το κάθε LED έχει

σημασία για τους υπολογισμούς που γίνονται στην συνέχεια, δεν πρέπει να μεταβάλλεται σημαντικά από καρέ σε καρέ αν δεν γίνεται κάποια κίνηση. Ωστόσο αυτό δεν είναι εύκολο να συμβεί, ειδικά όταν υπάρχει ανομοιογένεια στον φωτισμό του φόντου, οπότε υπάρχουν πάντα σφάλματα στον υπολογισμό με αποτέλεσμα να μετακινείται ελαφρά το κέντρο του κάθε LED σε διαδοχικά καρέ και ο αλγόριθμος να το εκλαμβάνει εσφαλμένα ως κίνηση.

- **Ακούσια κίνηση του κεφαλιού.** Είναι πολύ συχνό φαινόμενο όταν θέλει κάποιος να εστιάσει το βλέμμα του προς μία κατεύθυνση να μετακινεί ασυναίσθητα και το κεφάλι του προς την ίδια κατεύθυνση. Το ίδιο φαινόμενο παρατηρείται και από τα άτομα που έλαβαν μέρος στην βάση μας. Ειδικά στο β' μέρος της βάσης παρατηρείται πιο έντονα, που το υποκείμενο πρέπει να αλλάζει την κατεύθυνση του βλέμματος χωρίς να μετακινεί το κεφάλι του. Για αυτόν τον λόγο παρατηρούμε στα γραφήματα των γωνιών περιστροφής να υπάρχουν μικρές διακυμάνσεις των τιμών τους (Εικ. 4.3). Το υποκείμενο παρόλο που πρέπει να μην περιστρέψει το κεφάλι του και να αλλάξει μόνο την κατεύθυνση του βλέμματος, παρατηρούμε ότι το περιστρέφει ακούσια προς την πλευρά που γυρίζει το βλέμμα του.

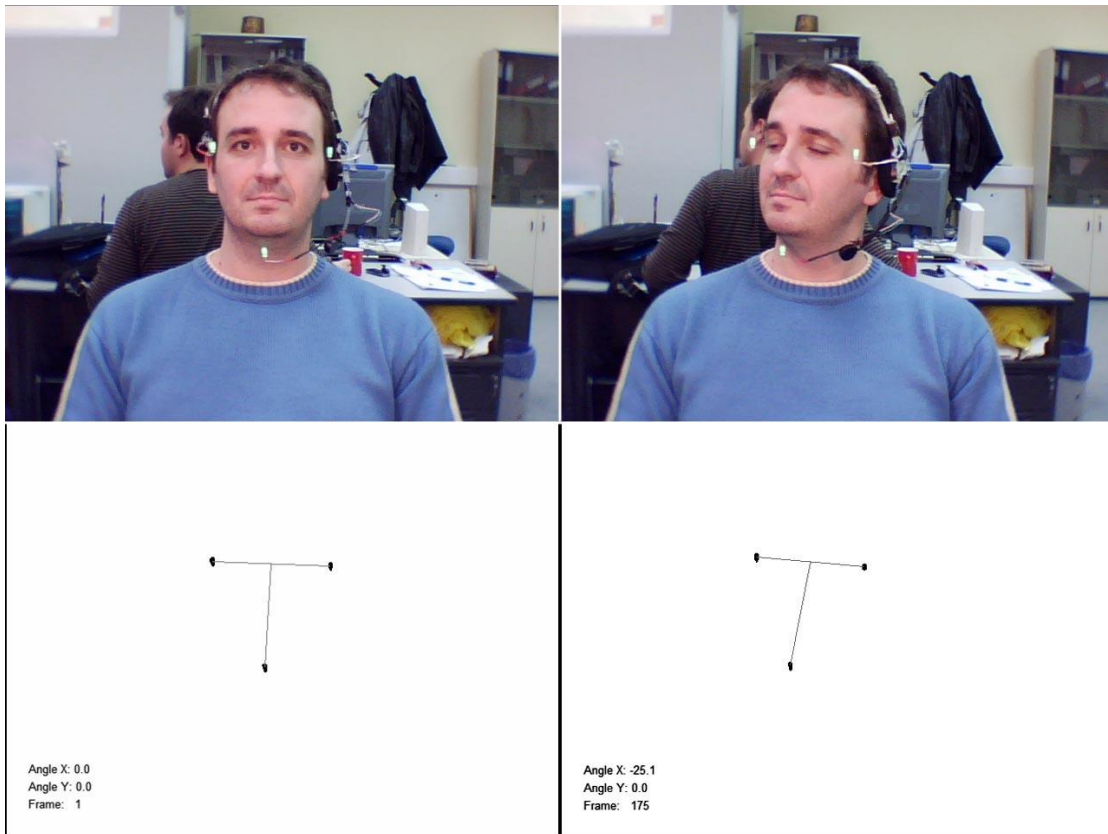


Εικ. 4.3 Ακούσια περιστροφή του κεφαλιού

4.2 Ανάλυση των βίντεο της βάσης

Στις σελίδες που ακολουθούν κάθε υποενότητα παρουσιάζει και ένα βίντεο της βάσης. Αρχικά παραθέτουμε δύο στιγμιότυπα από το αρχικό βίντεο μαζί με τα αντίστοιχα στιγμιότυπα από το βίντεο εξόδου. Στην συνέχεια παραθέτουμε σε πίνακα κάποια τεχνικά χαρακτηριστικά του και σημειώνουμε αν έχει γίνει χρήση της ρύθμισης χρωματικών επιπέδων και αν έχουν προκύψει προβληματικά καρέ, και αν ναι σημειώνουμε ποια ήταν αυτά. Στην συνέχεια γίνεται ένας σύντομος σχολιασμός και αξιολόγηση των αποτελεσμάτων και στο τέλος παραθέτουμε τρία γραφήματα σχεδιασμένα από το matlab. Στο πρώτο γράφημα απεικονίζονται τα εικονοστοιχεία περιστροφής από την αρχική μετωπική θέση του κεφαλιού και στους δύο άξονες ανίχνευσης της κίνησης. Στο δεύτερο, έχουν σχεδιαστεί οι μοίρες της γωνία περιστροφής του κεφαλιού στον άξονα Y για όλα τα καρέ του βίντεο. Τέλος στο τρίτο γράφημα, παρουσιάζονται οι μοίρες της γωνίας περιστροφής του κεφαλιού στον άξονα X για όλα τα καρέ του βίντεο.

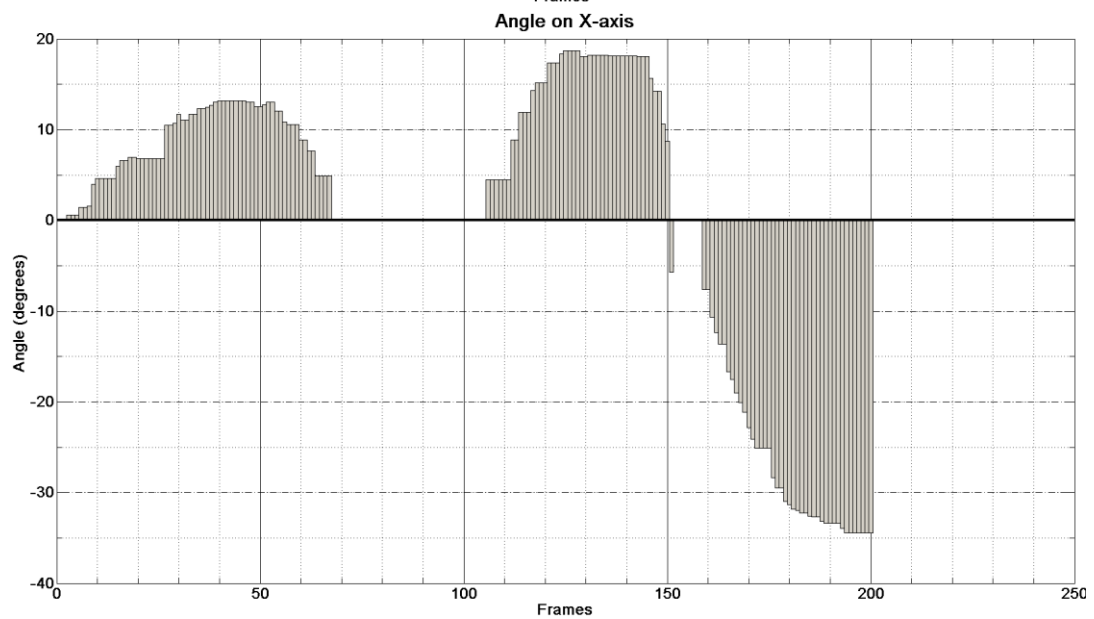
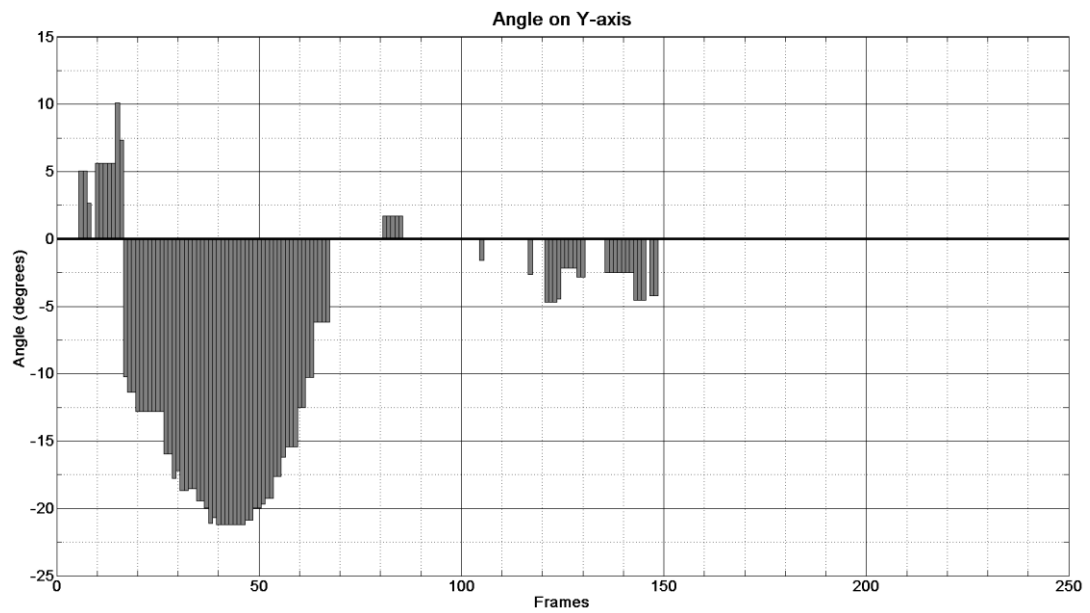
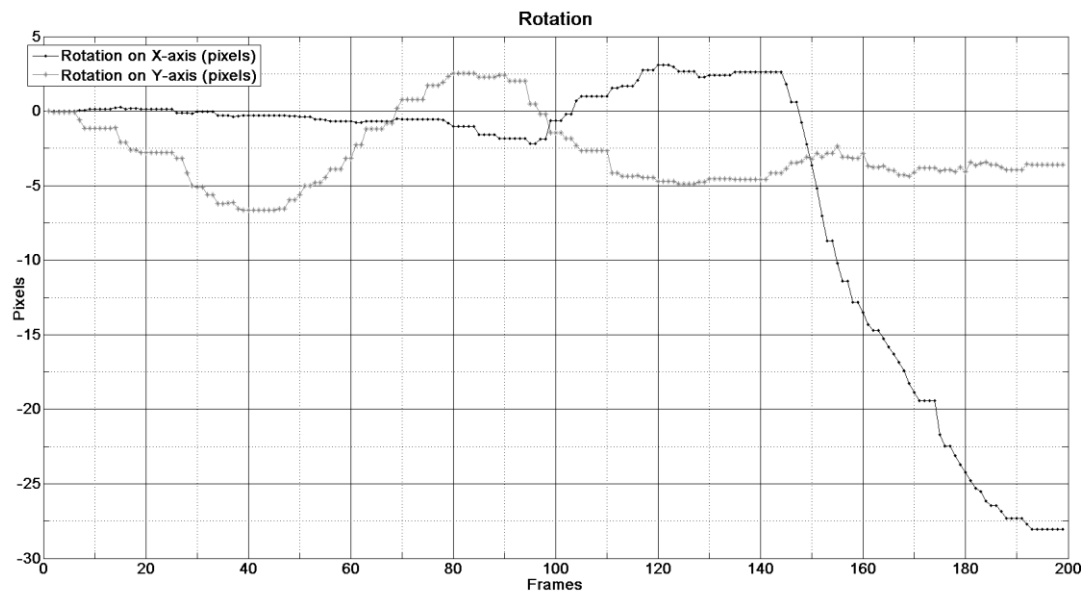
4.2.1 Video1a



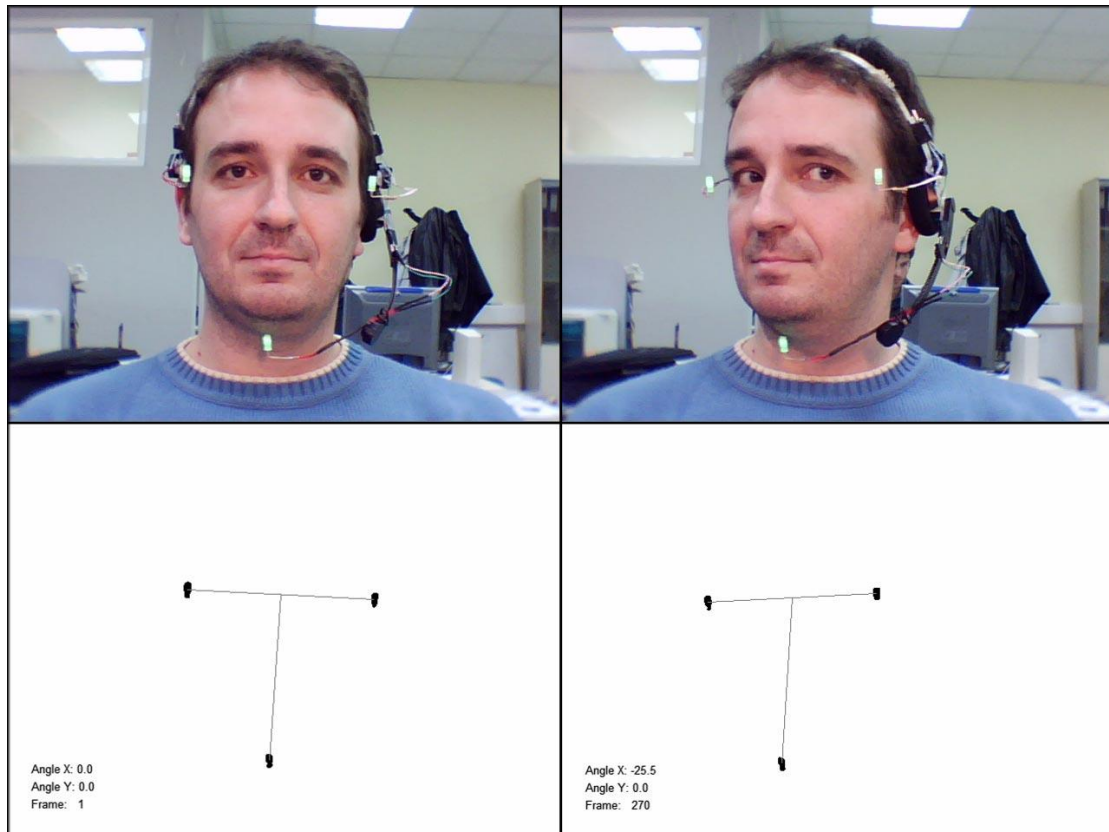
Ανάλυση	Χρωματική παλέτα	Συμπίεση	Αριθμός καρέ	Frames per sec	Ρύθμιση χρωμ. Επιπέδων (levels adjustment)	Καρε αποτυχίας ανίχνευσης (Bad Frames)
640x480	RGB24	όχι	200	30fps	×	-

Παρατηρήσεις:

Το Video1a παρουσίαζε αρκετά bad frames σε διάφορα σημεία του λόγω της ελαφρώς μεγαλύτερης διακύμανσης του εμβαδού του κάθε LED που παρατηρήθηκε σε σχέση με την πλειονότητα των βίντεο της βάσης. Με κατάλληλη ρύθμιση των παραμέτρων του ανιχνευτή εξαλείφτηκαν όλα. Συγκεκριμένα, ορίσαμε τον συντελεστή $\text{minArea} = 0,58$ και τον $\text{maxArea} = 1,4$ ενώ οι υπόλοιπες παράμετροι έμειναν στις προκαθορισμένες ρυθμίσεις τους. Παρατηρούμε στο διάγραμμα της γωνίας περιστροφής στον y άξονα υπάρχει μία απότομη αλλαγή προσήμου στα πρώτα καρέ, αλλά είναι εντός των ορίων ασάφειας για την κατεύθυνση περιστροφής. Η σωστή κατεύθυνση είναι προς τα πάνω, και την αντιλαμβάνεται σωστά ο αλγόριθμος μόλις η τιμή της γωνίας υπερβεί περίπου τις 10° (αρνητικό πρόσημο) ενώ αρχικά την υπολόγιζε λάθος.



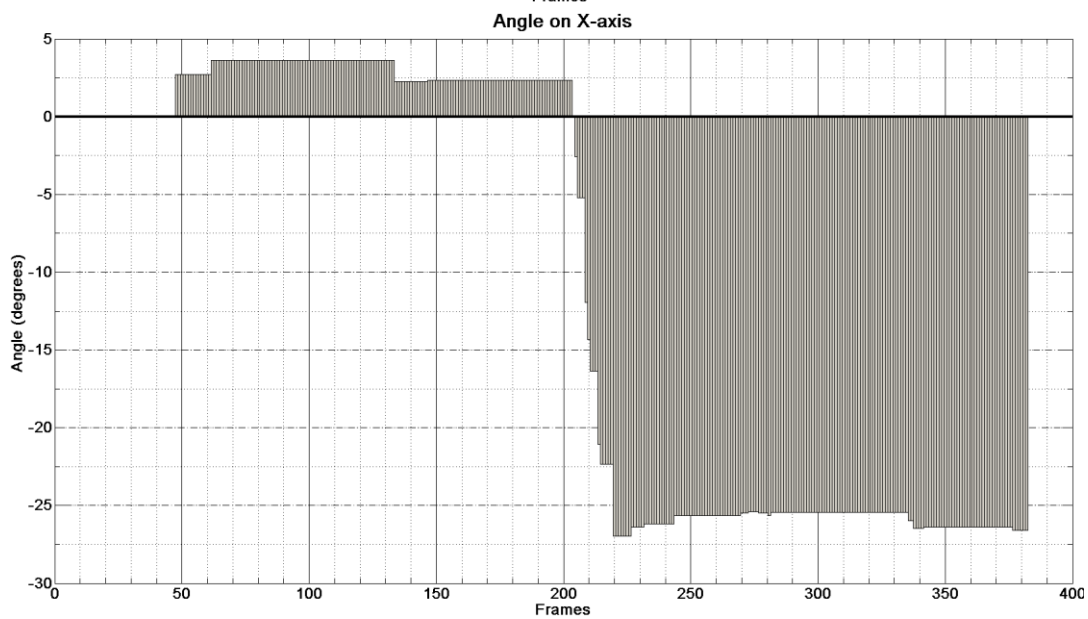
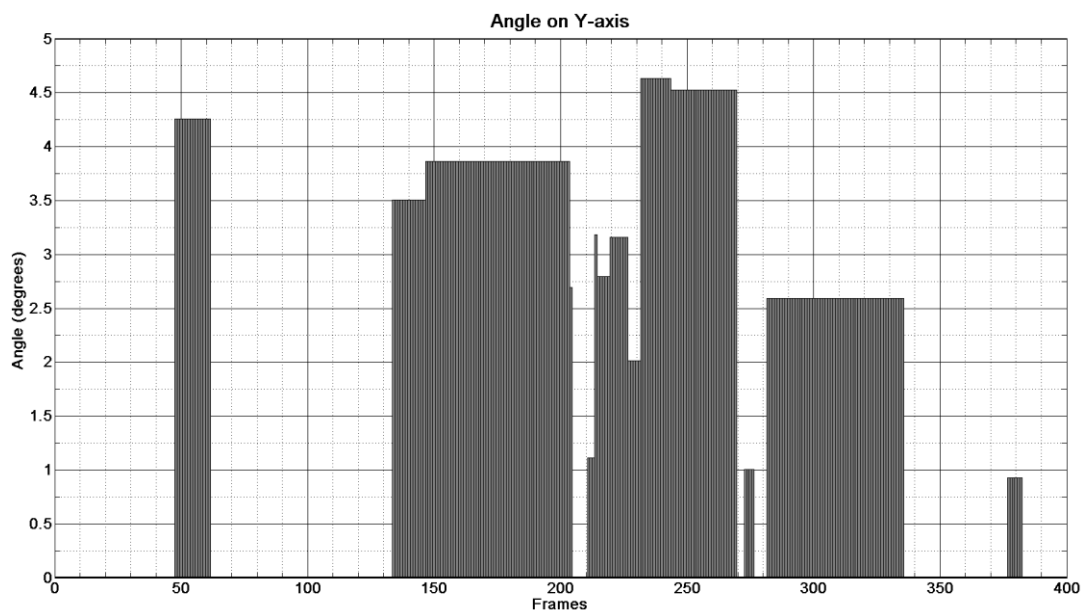
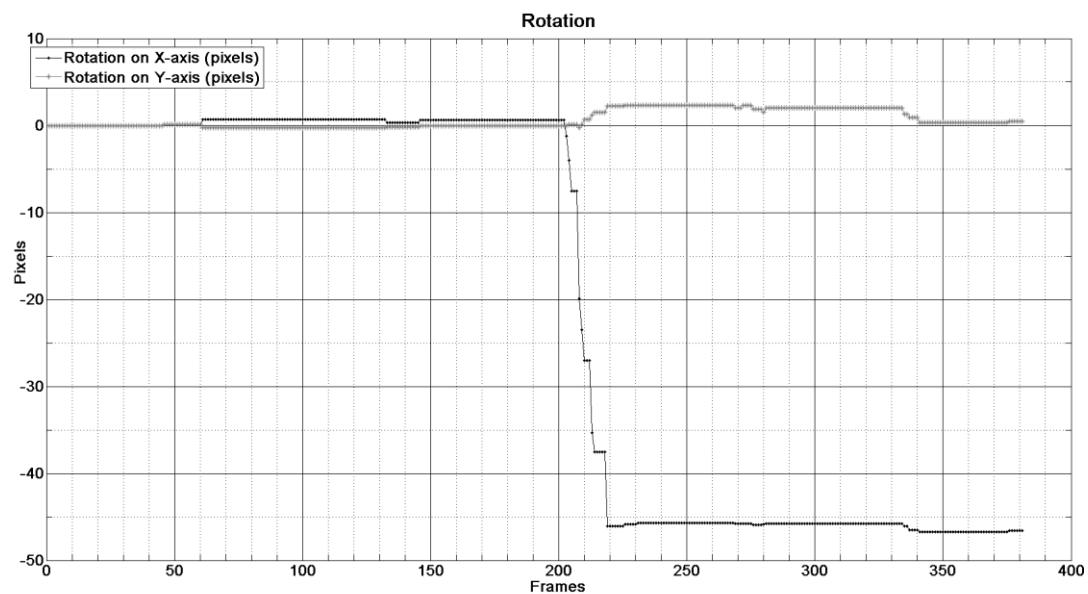
4.2.2 Video1b

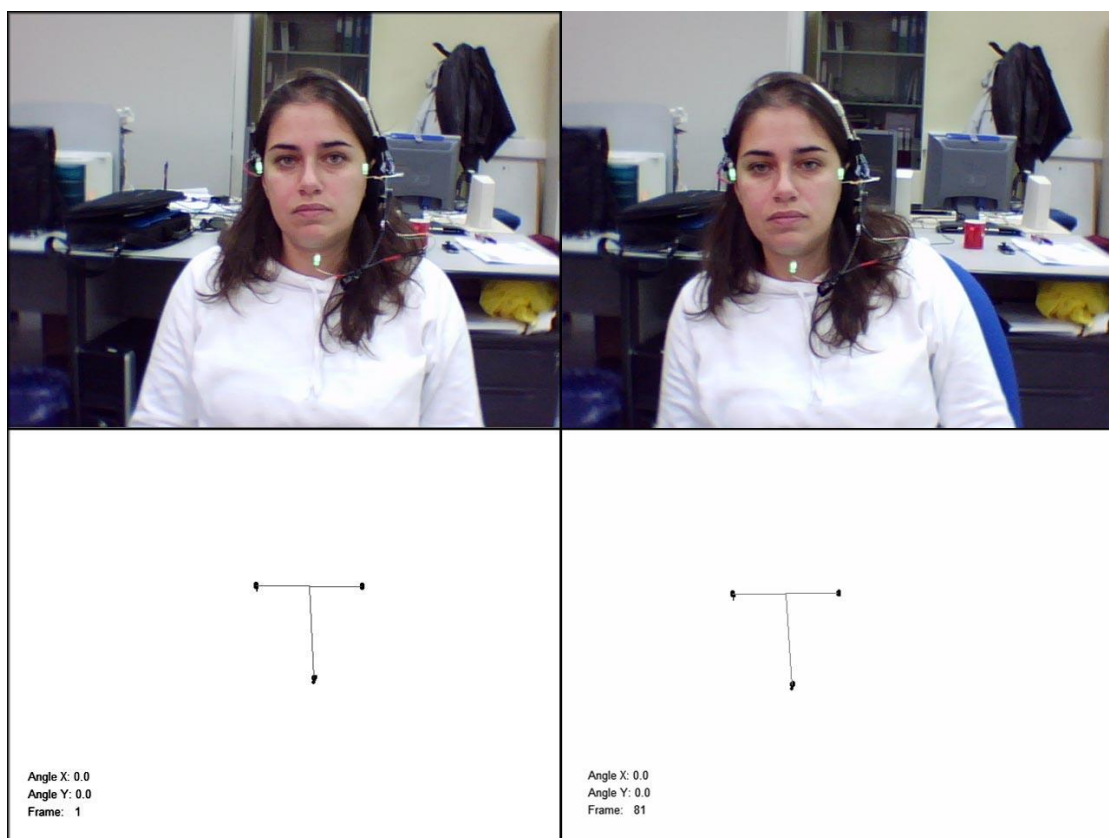


Ανάλυση	Χρωματική παλέτα	Συμπίεση	Αριθμός καρέ	Frames per sec	Ρύθμιση χρωμ. Επιπέδων (levels adjustment)	Καρε αποτυχίας ανίχνευσης (Bad Frames)
640x480	RGB24	όχι	382	30fps	x	-

Παρατηρήσεις:

Στο video2b δεν χρειάστηκε να γίνει καμία αλλαγή στις τιμές των παραμέτρων και η ρύθμιση χρωματικών επιπέδων ήταν απενεργοποιημένη. Το υποκείμενο περιστρέφει το κεφάλι του προς τα δεξιά στο 204 καρέ. Ο αλγόριθμος εκτιμά την κατεύθυνση χωρίς σφάλματα και στην μέτρηση σε εικονοστοιχεία και στην μέτρηση της γωνίας στους 2 άξονες, όπως φαίνεται και στα διαγράμματα που ακολουθούν. Κάποιες πολύ μικρές γωνίες που καταγράφονται όταν το κεφάλι του υποκειμένου είναι στην μετωπική πόζα είναι εντός των ορίων ανοχής, ενώ η μέγιστη γωνία περιστροφής είναι $26,5^\circ$ δεξιά στον οριζόντιο άξονα και $4,6^\circ$ πάνω στον κάθετο άξονα(εντός των ορίων ασάφειας).



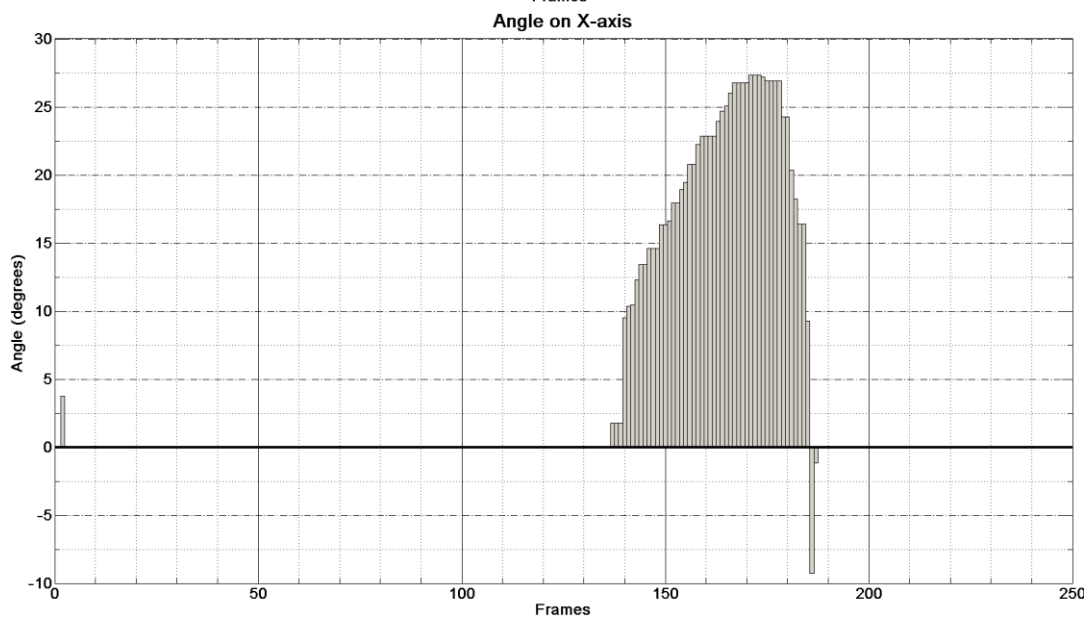
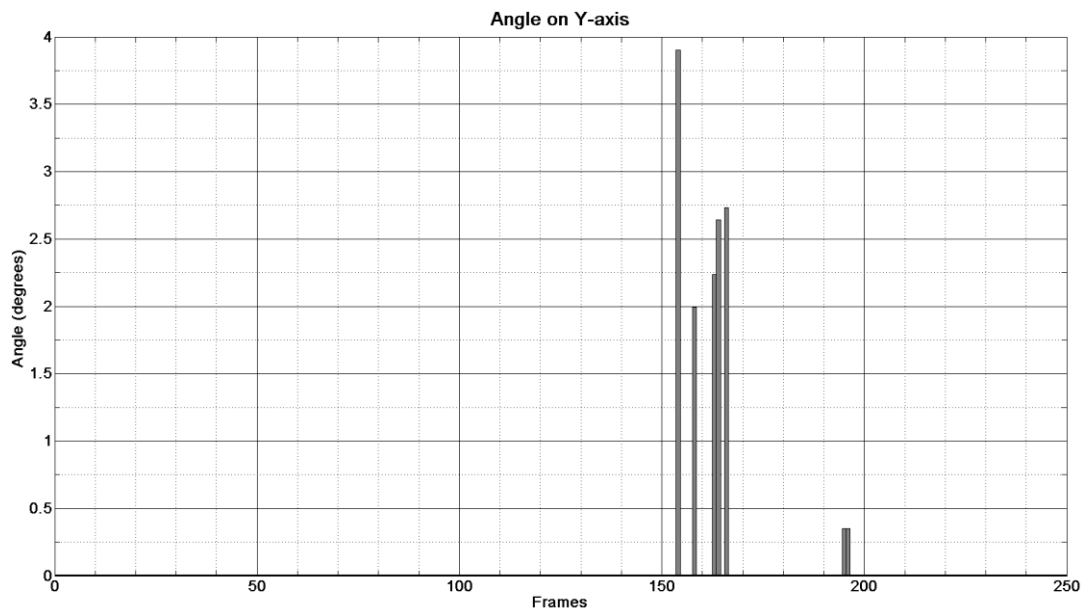
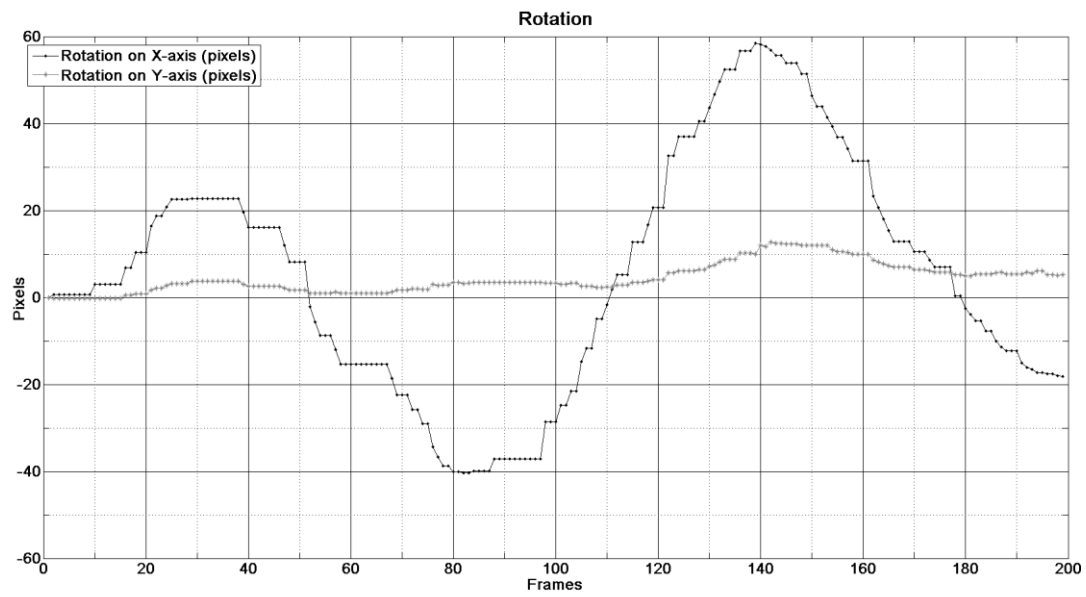


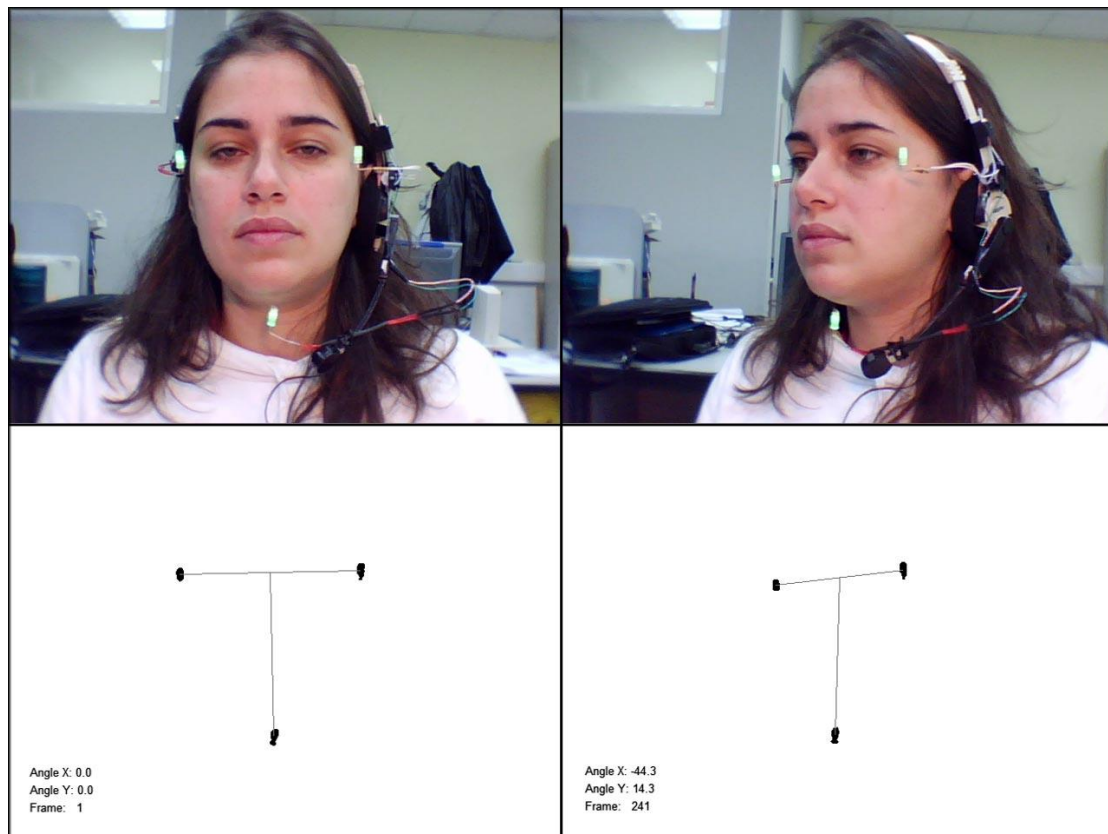
Ανάλυση	Χρωματική παλέτα	Συμπίεση	Αριθμός καρέ	Frames per sec	Ρύθμιση χρωμ. Επιπέδων (levels adjustment)	Καρέ αποτυχίας ανίχνευσης (Bad Frames)
640x480	RGB24	όχι	200	30fps	✓	-

Παρατηρήσεις:

Λόγω ύπαρξης μεγάλων περιοχών με λευκό χρώμα είναι πιθανό να μην γίνει σωστή ιχνηλάτηση των LEDs, γι' αυτό έγινε χρήση της ρύθμισης των χρωματικών επιπέδων (στην τιμή 0,65) ώστε να διακρίνονται καλύτερα τα σημεία. Επίσης περιορίσαμε την περιοχή ενδιαφέροντος (sizeAreaOfInterest) στα 18 pixel και τους συντελεστές minArea και maxArea στις τιμές 0,58 και 1,3 αντίστοιχα, στο κομμάτι ιχνηλάτησης του προγράμματος. Το αποτέλεσμα των ρυθμίσεων ήταν να μην υπάρχουν καρέ αποτυχίας ανίχνευσης. Το υποκείμενο στο συγκεκριμένο βίντεο πραγματοποιεί αρχικά μεταφορική κίνηση χωρίς να περιστρέφει το κεφάλι του, ενώ μετά το καρέ 138 μεταφορική και περιστροφική κίνηση ταυτόχρονα. Ενώ η μέθοδος

εκτίμησης σε εικονοστοιχεία αποτυγχάνει να διακρίνει την μεταφορική από την περιστροφική κίνηση όπως φαίνεται και στο διάγραμμα, η μέθοδος εκτίμησης της γωνίας διαχωρίζει με εξαιρετική ακρίβεια τις δύο κινήσεις. Από τα διαγράμματα παρατηρούμε ότι μέχρι το καρέ 138, που πραγματοποιείται καθαρά μεταφορική κίνηση, εκτιμάται μηδενική περιστροφή σε κατακόρυφο και οριζόντιο άξονα, ενώ μετά το 138 γίνεται άμεσα αντιληπτή η περιστροφή του κεφαλιού του αντικειμένου προς τα δεξιά με μέγιστη τιμή τις $27,2^\circ$.

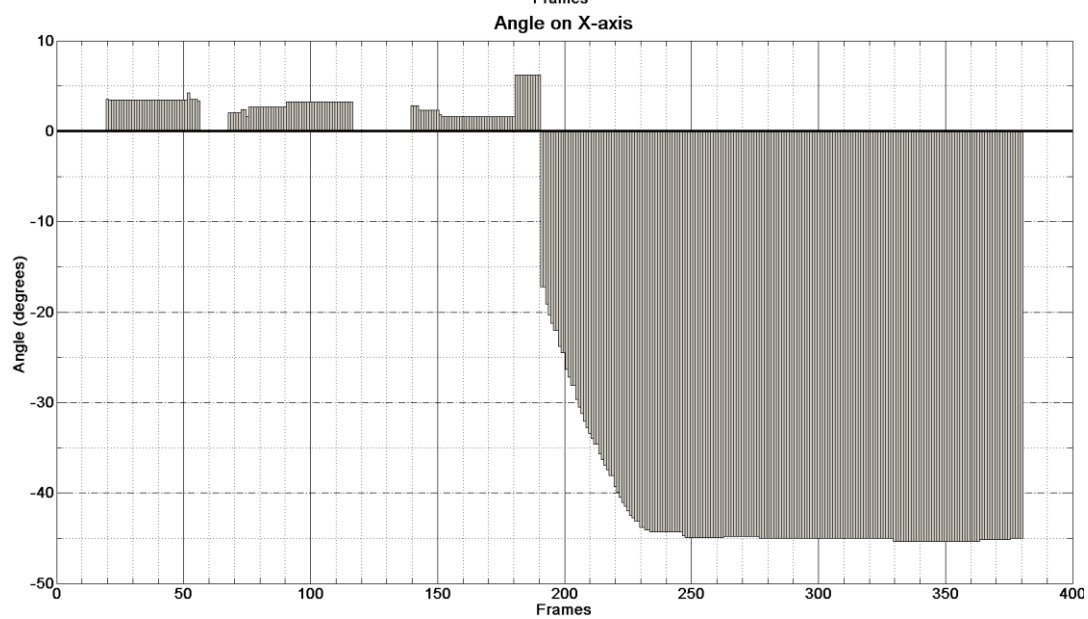
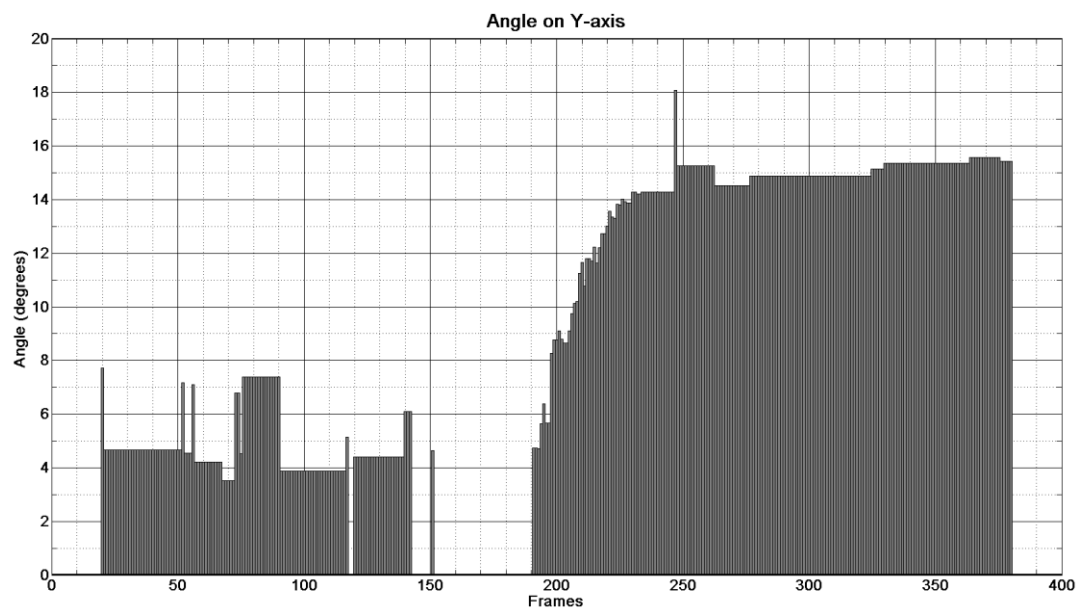
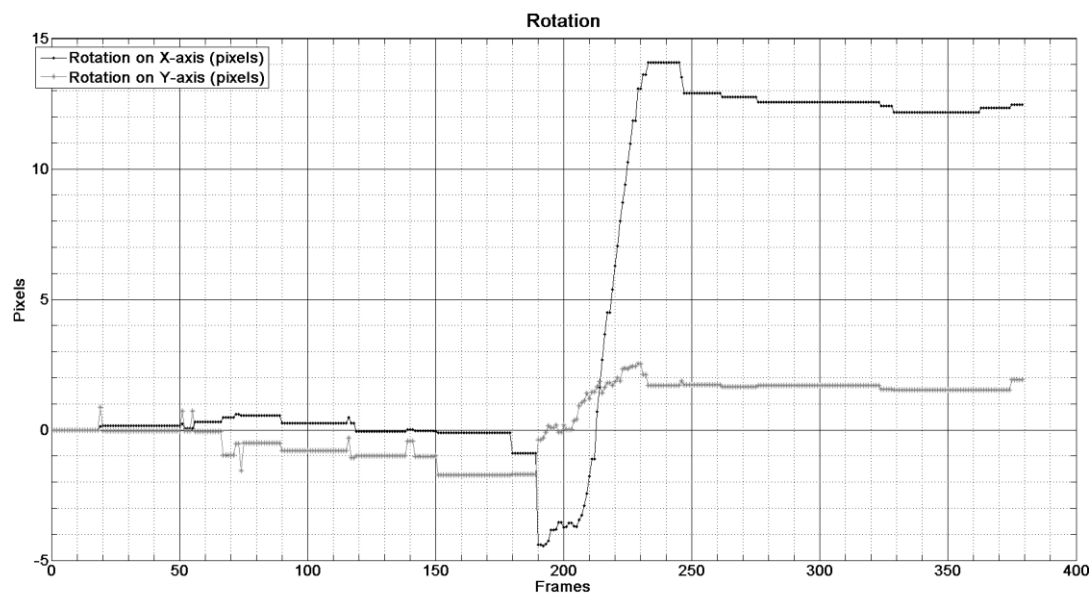


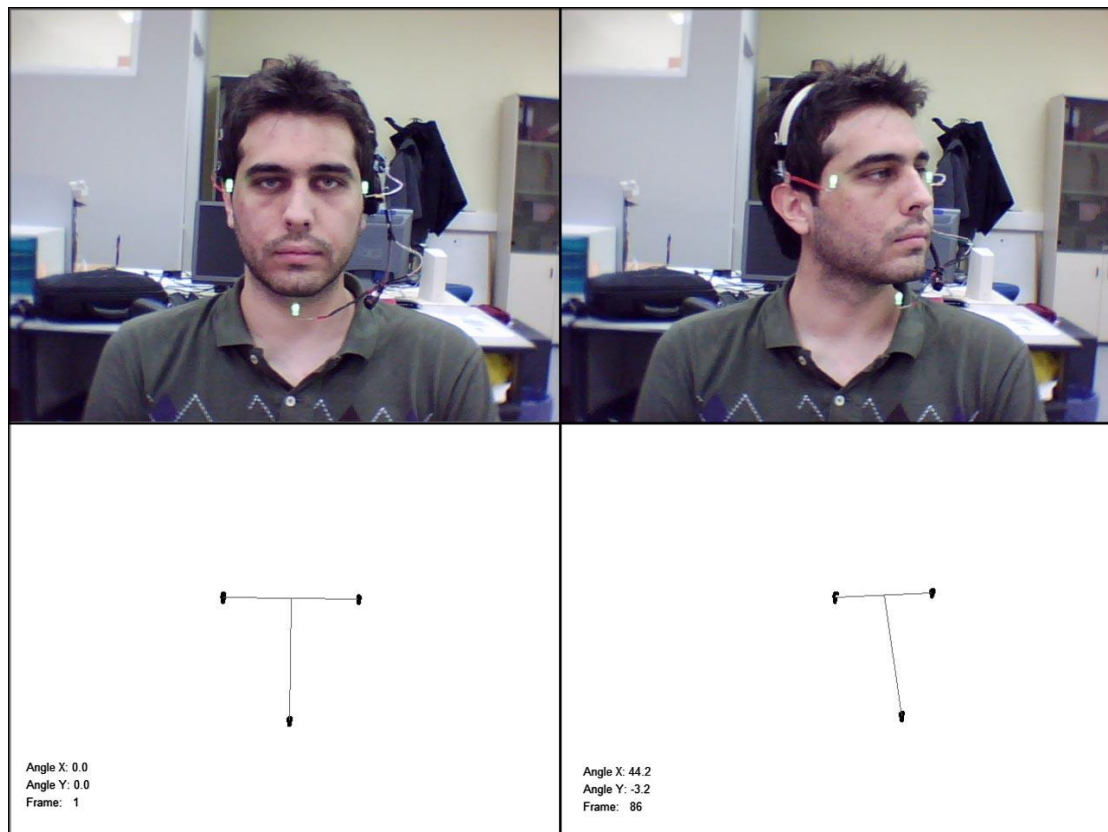


Ανάλυση	Χρωματική παλέτα	Συμπίεση	Αριθμός καρέ	Frames per sec	Ρύθμιση χρωμ. Επιπέδων (levels adjustment)	Καρε αποτυχίας ανίχνευσης (Bad Frames)
640x480	RGB24	όχι	380	30fps	×	-

Παρατηρήσεις:

Στο video2b δεν χρειάστηκε να γίνει καμία αλλαγή στις τιμές των παραμέτρων και η ρύθμιση χρωματικών επιπέδων ήταν απενεργοποιημένη. Το υποκείμενο περιστρέφεται προς τα δεξιά στο 176 καρέ με την διαφορά όμως από τα άλλα βίντεο ότι κάνει περιστροφή του σώματος και όχι μόνο του κεφαλιού. Αποτέλεσμα της λάθος εκτέλεσης της κίνησης από το υποκείμενο είναι ο αλγόριθμος να εκτιμά εσφαλμένα την κατεύθυνση στην μέτρηση σε εικονοστοιχεία. Ωστόσο ο αλγόριθμος εκτίμησης της γωνίας περιστροφής δίνει σωστά αποτελέσματα και δεν επηρεάζεται από την παράλληλη μεταφορική κίνηση του σώματος. Κάποιες πολύ μικρές γωνίες που καταγράφονται όταν το κεφάλι του υποκειμένου είναι στην μετωπική πόζα είναι εντός των ορίων ανοχής, ενώ η μέγιστη γωνία περιστροφής είναι 45,4° δεξιά στον οριζόντιο άξονα και 15,6° κάτω στον κάθετο άξονα.

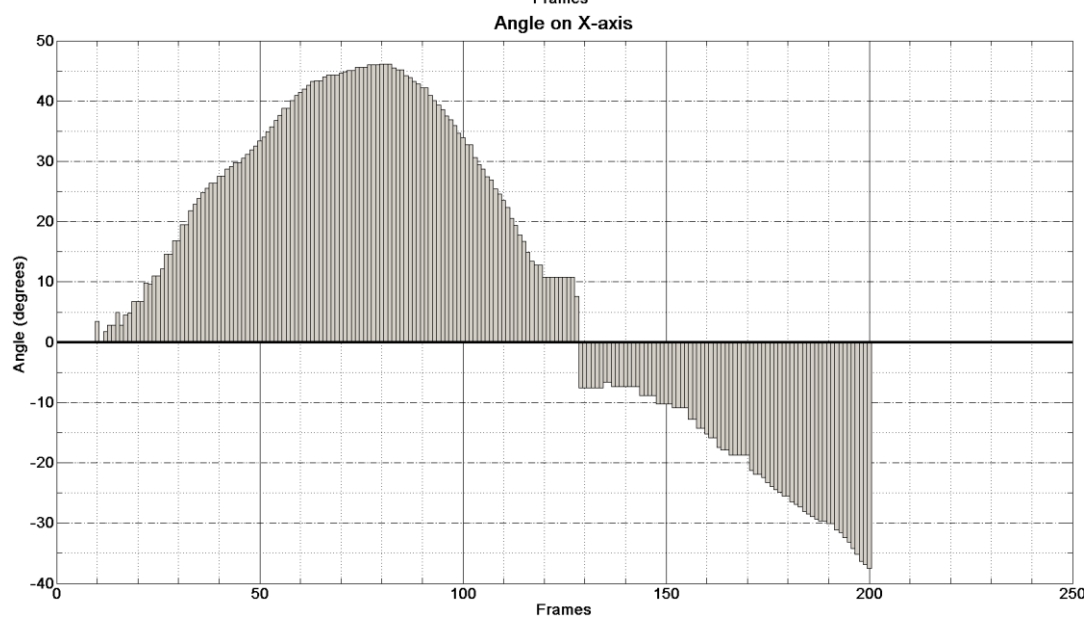
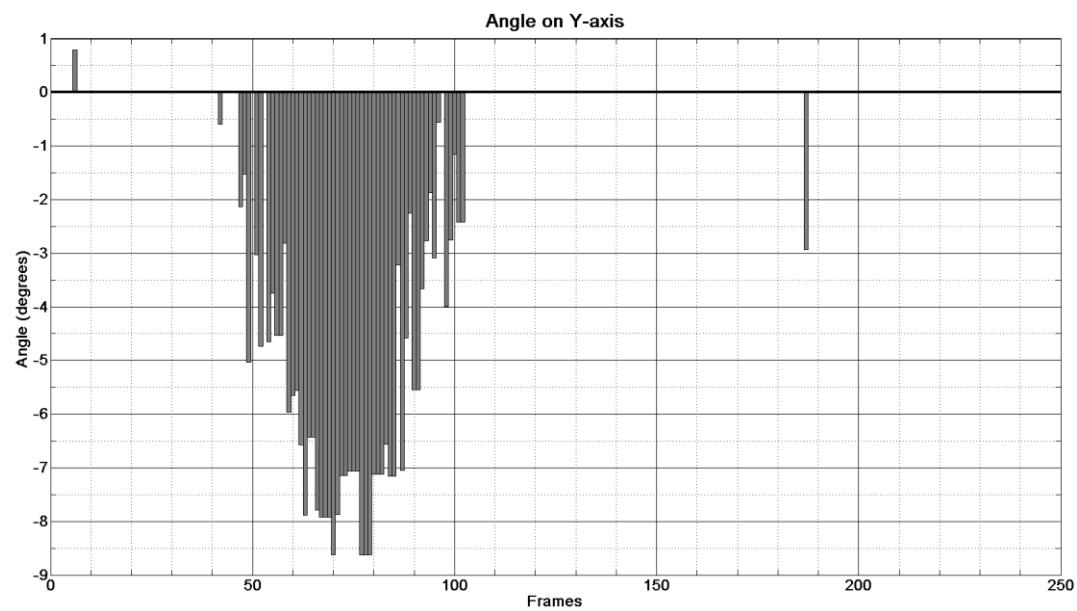
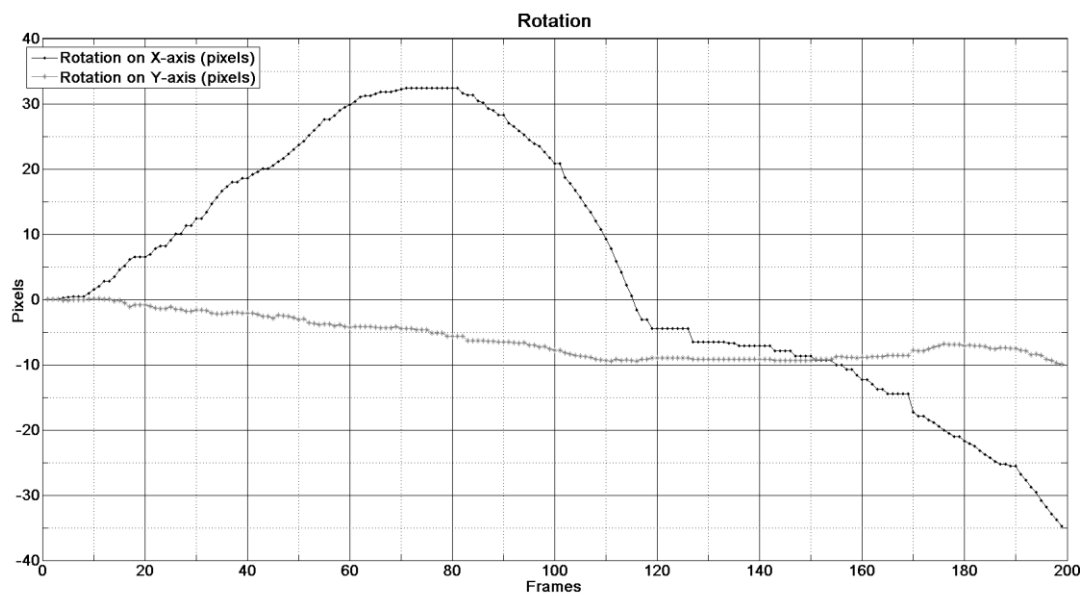


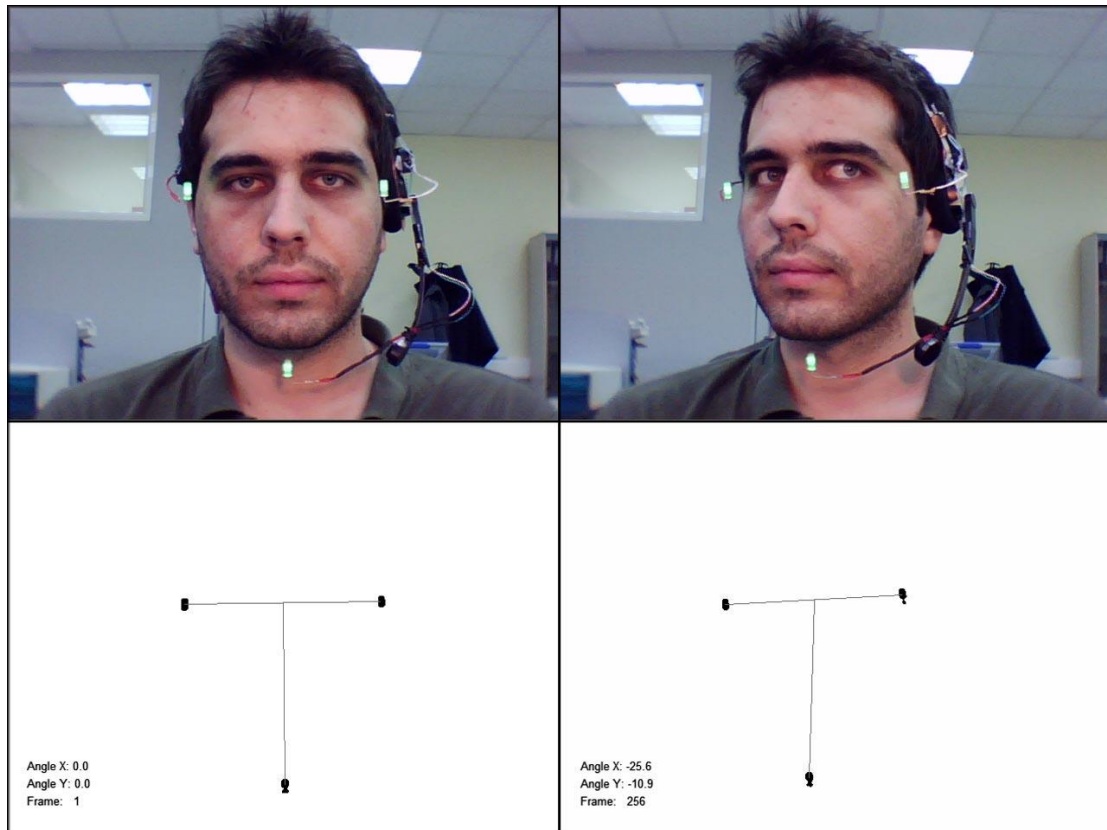


Ανάλυση	Χρωματική παλέτα	Συμπίεση	Αριθμός καρέ	Frames per sec	Ρύθμιση χρωμ. Επιπέδων (levels adjustment)	Καρε αποτυχίας ανίχνευσης (Bad Frames)
640x480	RGB24	όχι	200	30fps	×	-

Παρατηρήσεις

Δεν έγινε καμία ιδιαίτερη ρύθμιση των παραμέτρων του αλγορίθμου και δεν κρίθηκε απαραίτητη η χρήση της ρύθμισης των χρωματικών επιπέδων. Το υποκείμενο πραγματοποιεί ελεύθερη κίνηση και επιλέγει να περιστρέψει το κεφάλι του αρχικά δεξιά και στην συνέχεια αριστερά. Οι αλγόριθμοι εκτίμησης της πόζας υπολογίζουν σωστά τις κατευθύνσεις περιστροφής. Όπως φαίνεται και στα διαγράμματα η γωνία στον κάθετο άξονα είναι μηδέν στο μεγαλύτερο μέρος της κίνησης ενώ για περίπου 50 καρέ ανιχνεύεται μία πολύ μικρή κλίση του κεφαλιού προς τα πάνω με μέγιστη τιμή τις $8,5^\circ$, που είναι εντός των ορίων ασάφειας. Στον οριζόντιο άξονα X εκτιμά σωστά την ομαλή περιστροφή του κεφαλιού από την μετωπική πόζα σε στροφή αριστερά, επαναφορά στην μετωπική και τέλος σε στροφή αριστερά με μέγιστες γωνίες 46° αριστερά και $37,5^\circ$ δεξιά.

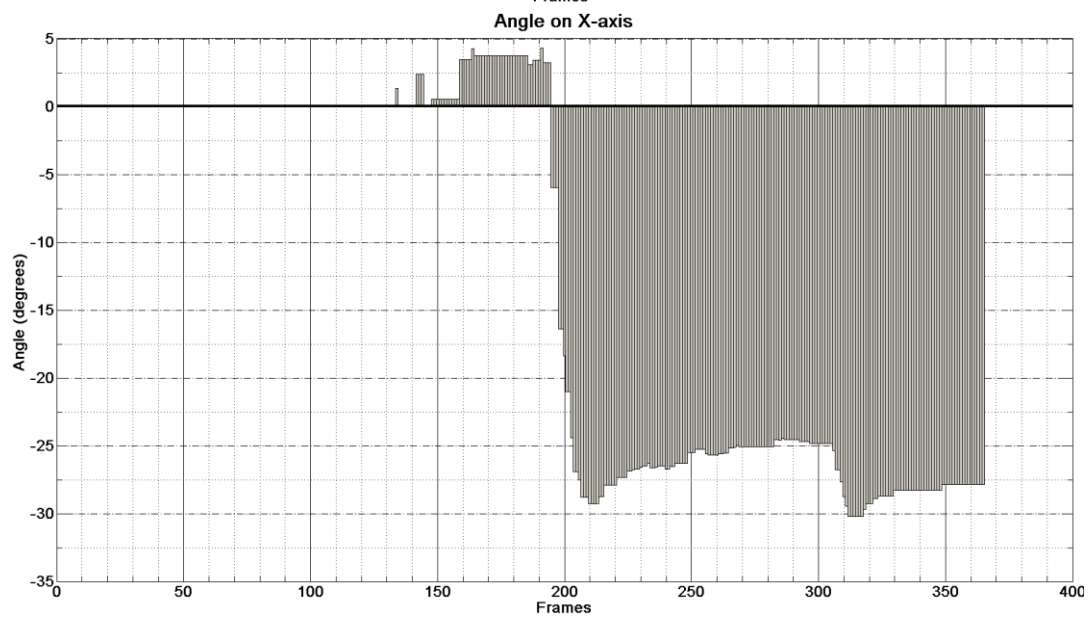
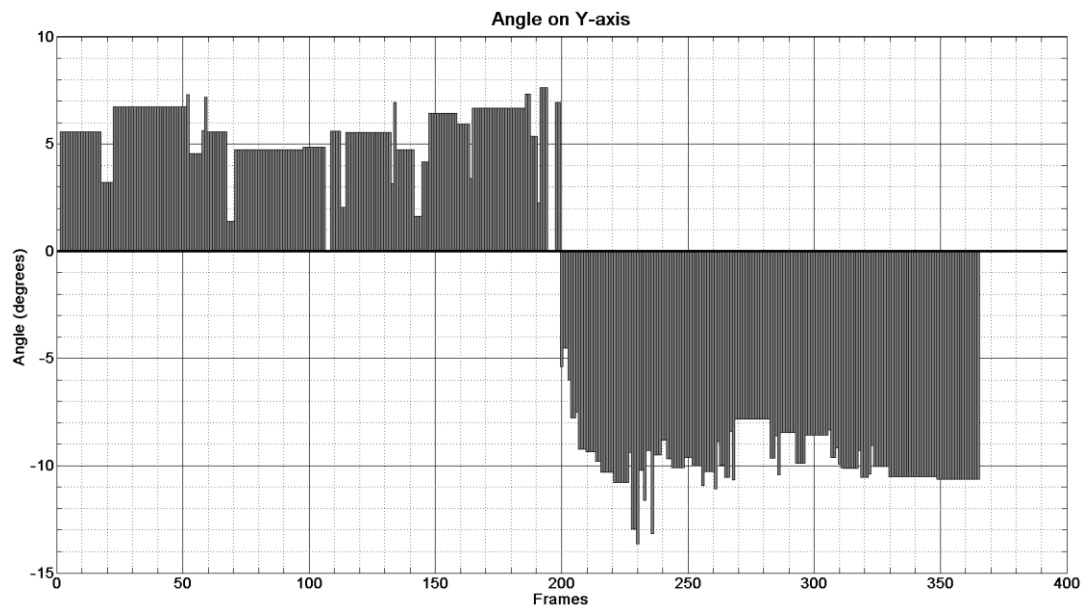
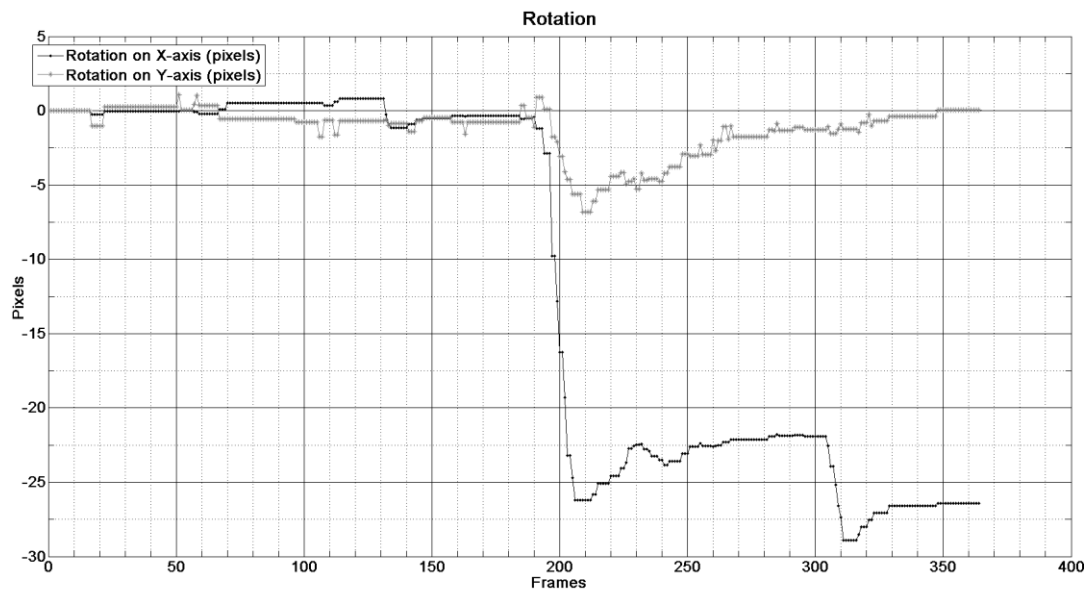


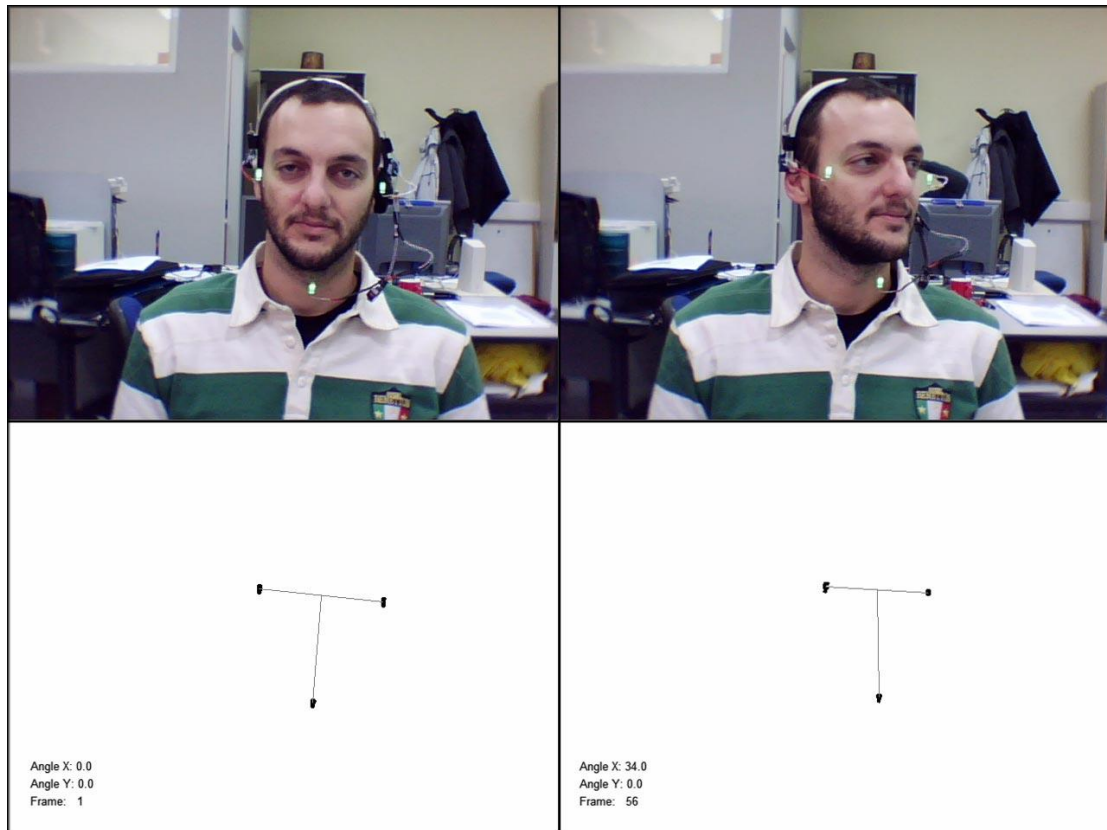


Ανάλυση	Χρωματική παλέτα	Συμπίεση	Αριθμός καρέ	Frames per sec	Ρύθμιση χρωμ. Επιπέδων (levels adjustment)	Καρε αποτυχίας ανίχνευσης (Bad Frames)
640x480	RGB24	όχι	365	30fps	✖	-

Παρατηρήσεις:

Οι παράμετροι παρέμειναν στις προκαθορισμένες τιμές τους, ενώ η ρύθμιση χρωματικών επιπέδων ήταν απενεργοποιημένη. Το υποκείμενο στο καρέ 192 πραγματοποιεί περιστροφή του κεφαλιού προς τα δεξιά, η οποία εκτιμάται ορθά από τον αλγόριθμο με μέγιστη τιμή τις 30°. Κάποιες πολύ μικρές γωνίες που παρατηρούνται και στους δύο άξονες είναι εντός των ορίων ασάφειας και δεν αξιολογούνται.



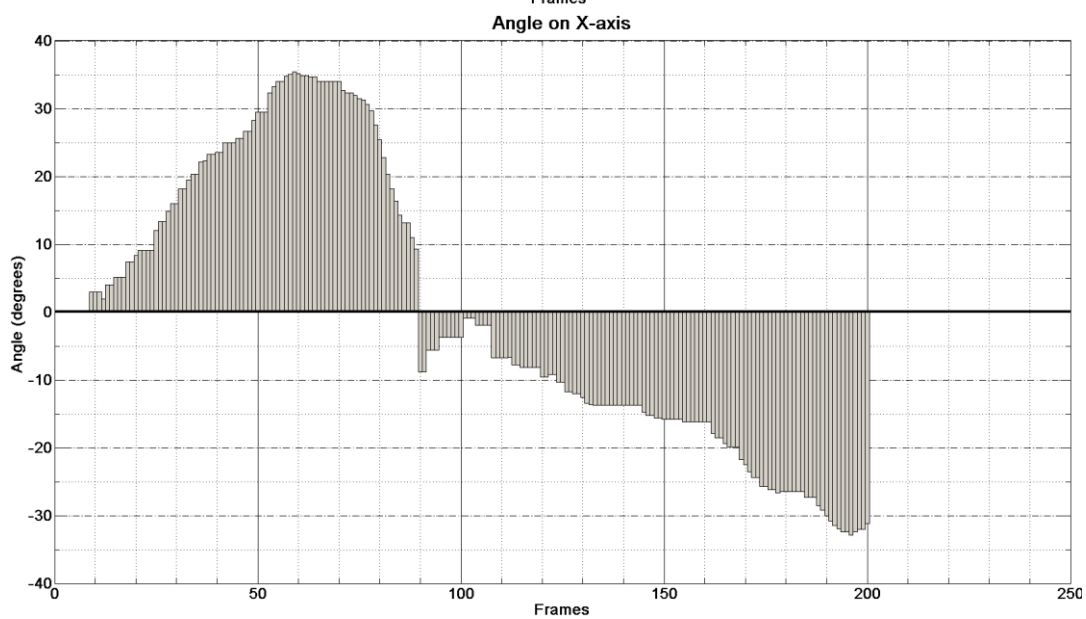
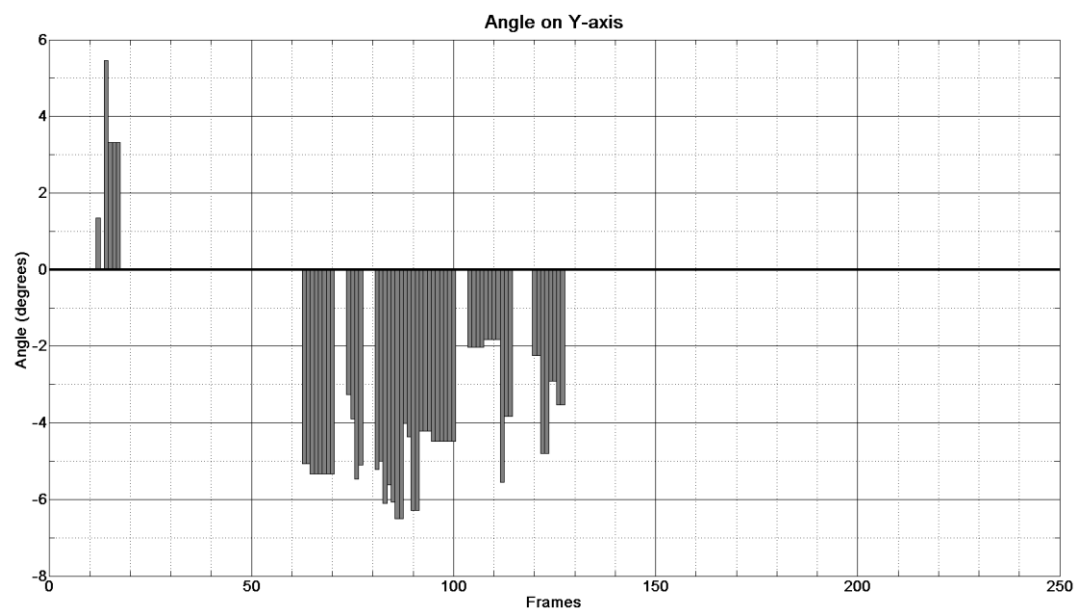
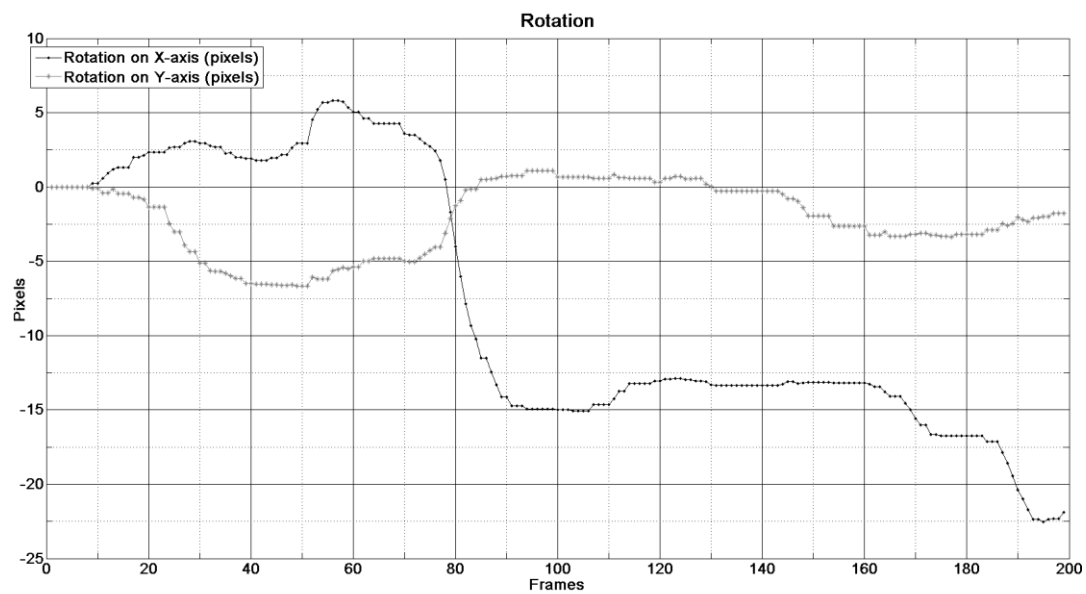


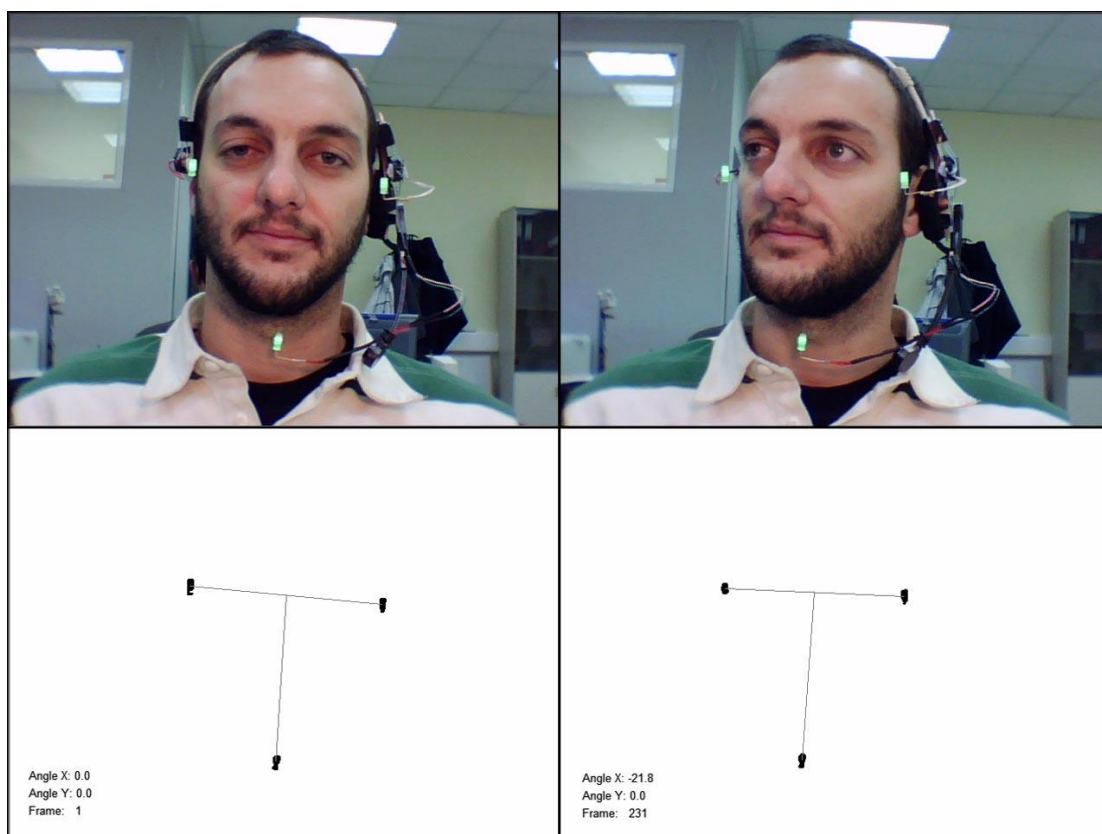
Ανάλυση	Χρωματική παλέτα	Συμπίεση	Αριθμός καρέ	Frames per sec	Ρύθμιση χρωμ. Επιπέδων (levels adjustment)	Καρε αποτυχίας ανίχνευσης (Bad Frames)
640x480	RGB24	όχι	200	30fps	✖	-

Παρατηρήσεις:

Στο video4a δεν πραγματοποιήθηκαν ειδικές ρυθμίσεις των παραμέτρων του αλγορίθμου ενώ η ρύθμιση χρωματικών επιπέδων ήταν απενεργοποιημένη. Το υποκείμενο πραγματοποιεί ελεύθερη κίνηση και επιλέγει να στρέψει το κεφάλι του αρχικά αριστερά και στην συνέχεια δεξιά με παράλληλες μικρές κλίσεις προς τα πάνω ή κάτω, ενώ αλλάζει συνεχώς την κατεύθυνση του βλέμματος του. Ο αλγόριθμος δεν επηρεάζεται από τις αλλαγές στην κατεύθυνση του βλέμματος και εκτιμά μόνο τις διαδοχικές στάσεις του κεφαλιού. Όπως παρατηρείται στα διαγράμματα οι μετρήσεις σε εικονοστοιχεία αλλά και σε μοίρες εκτιμούν σωστά τις κατευθύνσεις του κεφαλιού. Η μέθοδος υπολογισμού σε εικονοστοιχεία είναι πιο ευαίσθητη

στις μεταβολές στον κάθετο άξονα και όπως φαίνεται ανιχνεύει πιο εύκολα τις πολύ μικρές κλίσεις που παίρνει το κεφάλι του υποκειμένου. Στο διάγραμμα της γωνίας στον άξονα X η αλλαγή του πρόσημου δεν πραγματοποιείται στις 0 μοίρες αλλά στις 8° , τιμή όμως που βρίσκεται εντός των ορίων ασάφειας και συνεπώς δεν μπορεί να αξιολογηθεί ως λάθος εκτίμηση. Τέλος η μέγιστη γωνία περιστροφής στον X-άξονα είναι $35,4^{\circ}$ αριστερά και $32,9^{\circ}$ δεξιά, ενώ στον Y-άξονα οι τιμές που υπολογίζονται είναι εντός των ορίων ασάφειας.

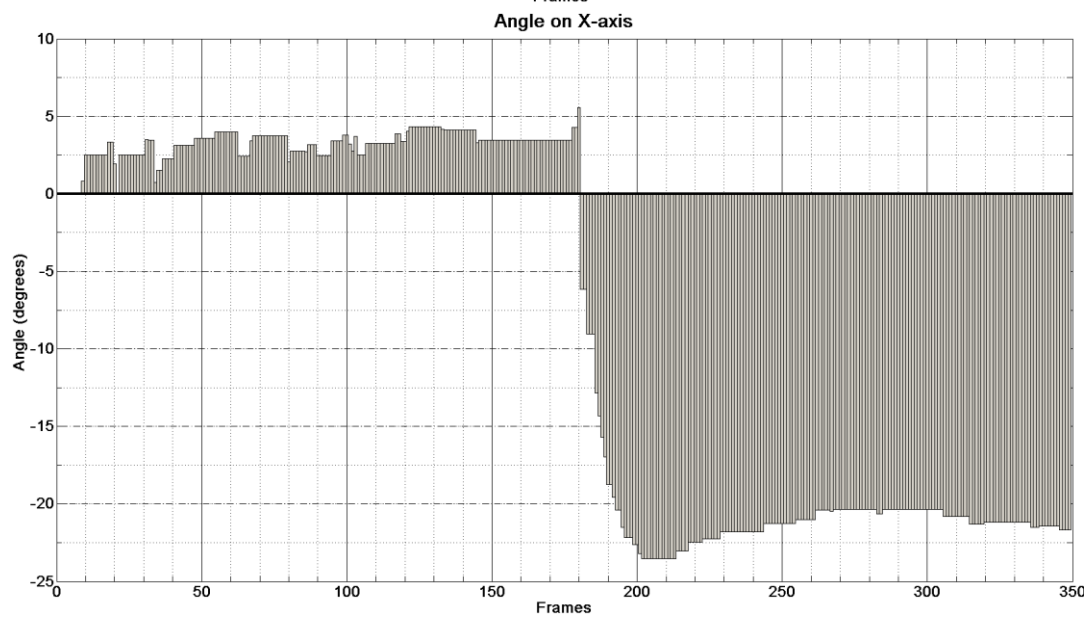
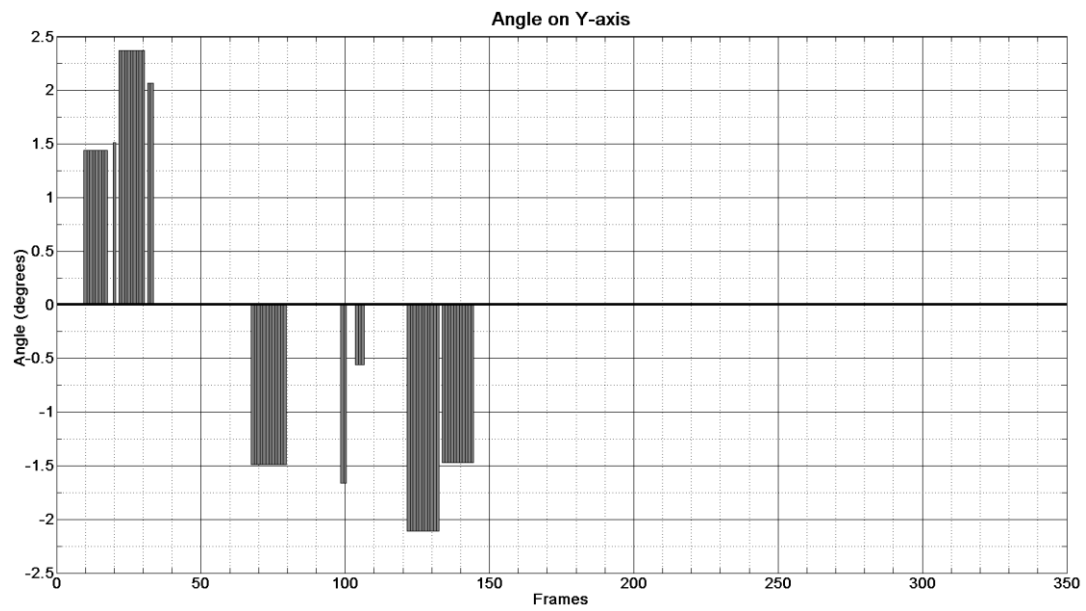
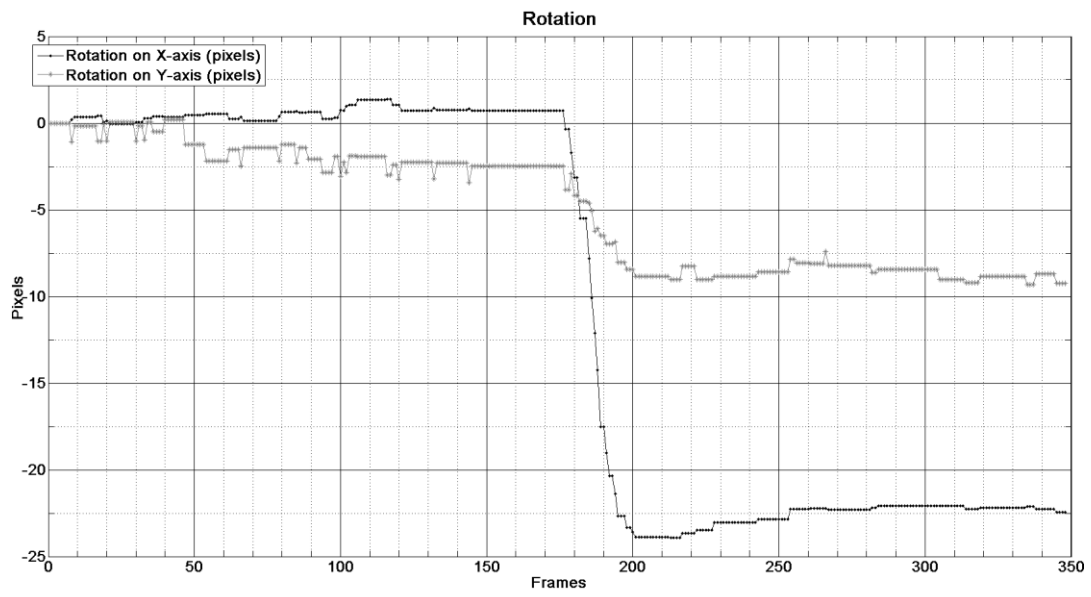


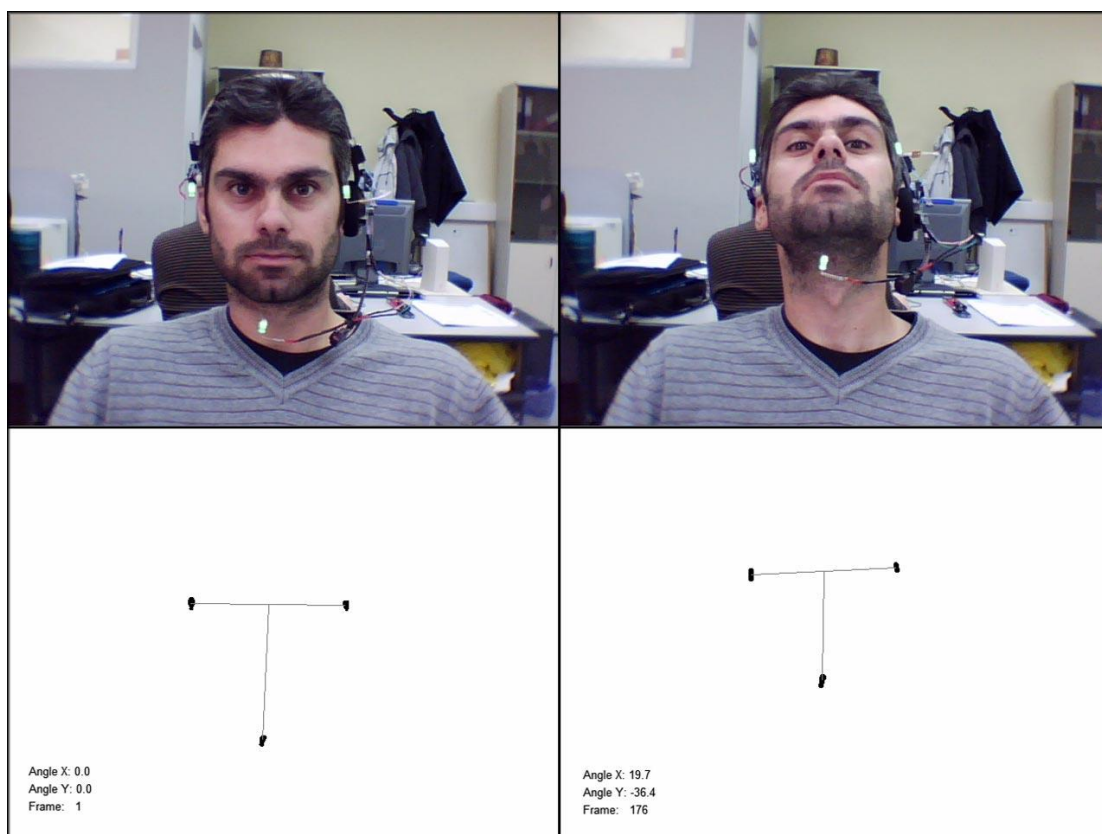


Ανάλυση	Χρωματική παλέτα	Συμπίεση	Αριθμός καρέ	Frames per sec	Ρύθμιση χρωμ. Επιπέδων (levels adjustment)	Καρε αποτυχίας ανίχνευσης (Bad Frames)
640x480	RGB24	όχι	349	30fps	×	-

Παρατηρήσεις:

Οι παράμετροι ρυθμίστηκαν στις προκαθορισμένες τιμές τους, ενώ η ρύθμιση χρωματικών επιπέδων ήταν απενεργοποιημένη. Το υποκείμενο στο 177 καρέ πραγματοποιεί περιστροφή του κεφαλιού προς τα δεξιά, η οποία εκτιμάται ορθά από τον αλγόριθμο με μέγιστη τιμή τις 23,5°. Κάποιες πολύ μικρές γωνίες που παρατηρούνται και στους δύο άξονες είναι εντός των ορίων ασάφειας και δεν αξιολογούνται. Τέλος, σημειώνεται ότι η θέση των LED στο κεφάλι του υποκειμένου θα έπρεπε να ήταν πιο ψηλά έτσι ώστε η νοητή ευθεία των 2 πάνω LEDs να ταυτίζεται με την ευθεία των ματιών. Παρά την σχετικά λάθος τοποθέτηση δεν παρατηρούνται σφάλματα στις εκτιμήσεις του αλγορίθμου.



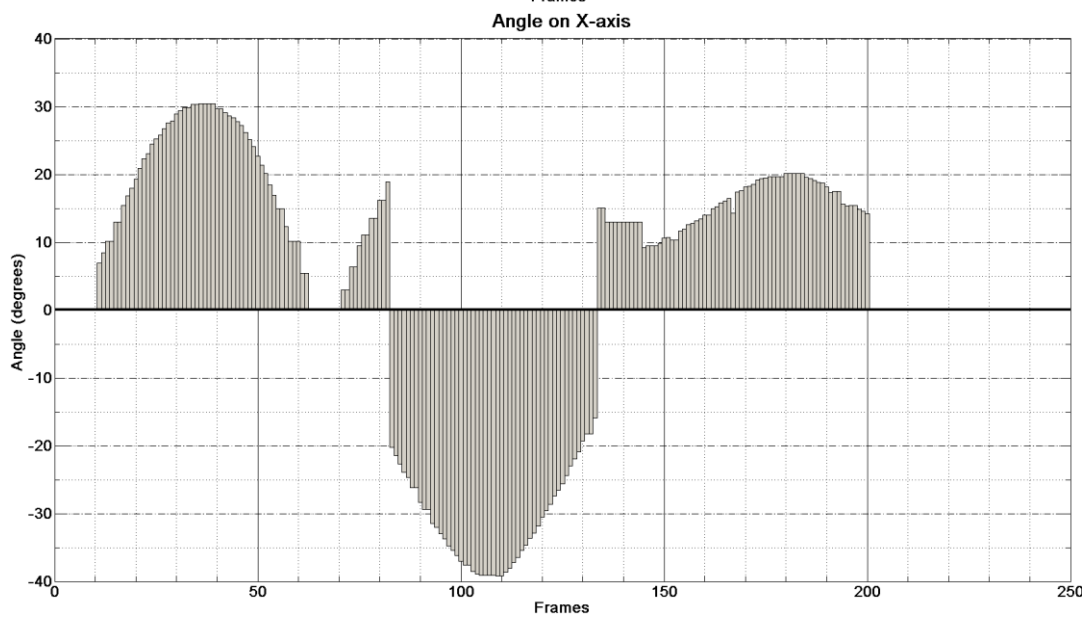
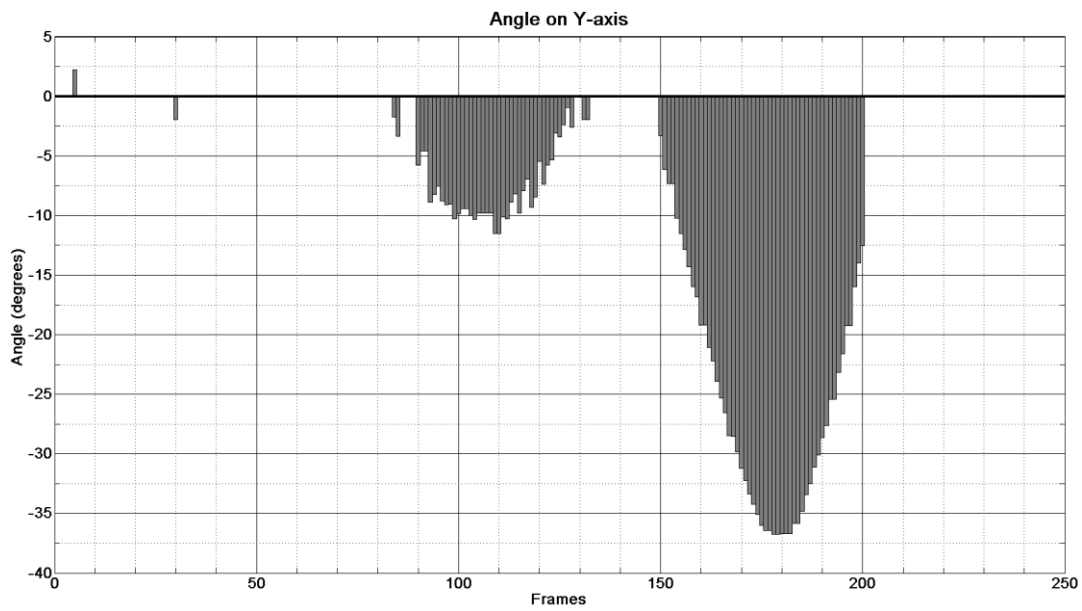
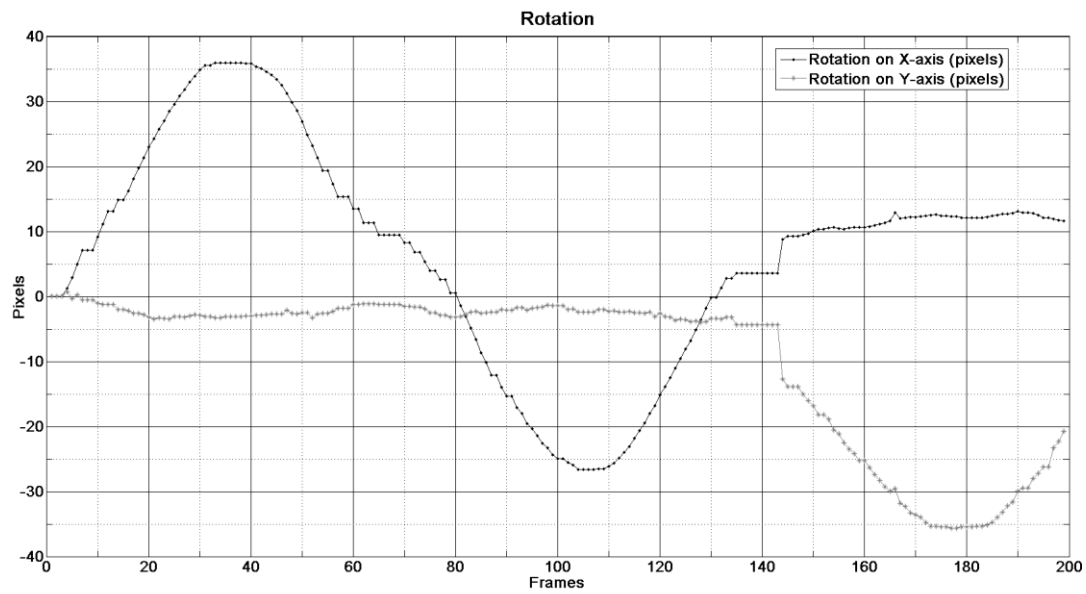


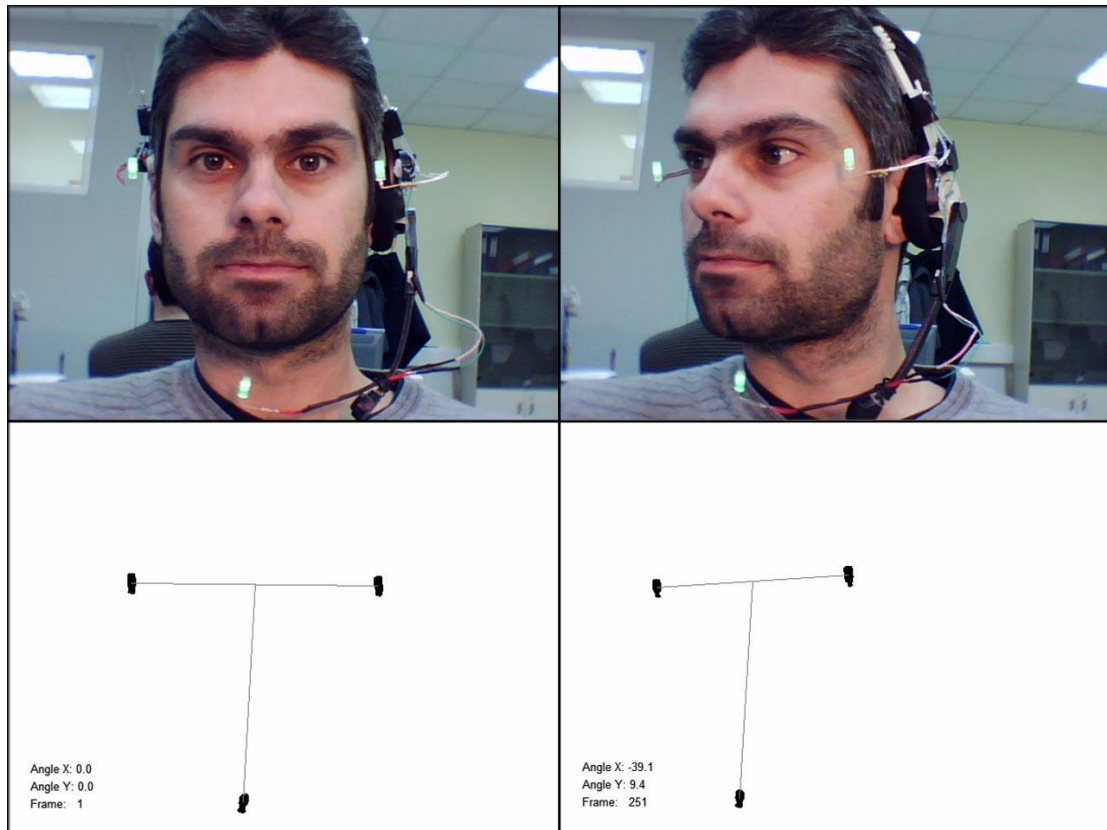
Ανάλυση	Χρωματική παλέτα	Συμπίεση	Αριθμός καρέ	Frames per sec	Ρύθμιση χρωμ. Επιπέδων (levels adjustment)	Καρε αποτυχίας ανίχνευσης (Bad Frames)
640x480	RGB24	όχι	200	30fps	×	-

Παρατηρήσεις:

Στο video5a οι εκτιμήσεις έγιναν με τις προκαθορισμένες ρυθμίσεις των παραμέτρων και απενεργοποιημένη την ρύθμιση χρωματικών επιπέδων. Το υποκείμενο πραγματοποιεί ελεύθερη κίνηση αρχικά στρέφοντας το κεφάλι του αριστερά, στην συνέχεια δεξιά και τέλος αφού έχει επανέλθει στην μετωπική πόζα το στρέφει προς το πάνω. Δεν πραγματοποιείται καμία μεταφορική κίνηση. Στο διάγραμμα της περιστροφής σε εικονοστοιχεία φαίνονται σωστές οι εκτιμήσεις των αλλαγών της στάσης και στους δύο άξονες. Στο διάγραμμα της γωνίας στον Y-άξονα παρατηρούμε την σωστή ανίχνευση της στροφής του κεφαλιού προς τα πάνω μετά το καρέ 150 με μέγιστη γωνία στις 36,8°. Ωστόσο στον X-άξονα υπάρχει μία καθυσ-

τέρηση στην σωστή εκτίμηση των πλευράς περιστροφής μετά το καρέ 70 και για περίπου 12 καρέ. Επίσης εσφαλμένη αλλαγή προσήμου γίνεται στο καρέ 165 που πραγματοποιείται σε τιμή γωνίας εκτός της περιοχής ασάφειας. Ο βασικός λόγος που παρατηρούνται τα παραπάνω σφάλματα είναι η κακή τοποθέτηση του LED κάτω από το σαγόνι, το οποίο δεν ισαπέχει από τα άλλα δύο LEDs. Οι μέγιστες γωνίες στροφής στον οριζόντιο άξονα είναι $30,4^\circ$ αριστερά και $39,2^\circ$ δεξιά.

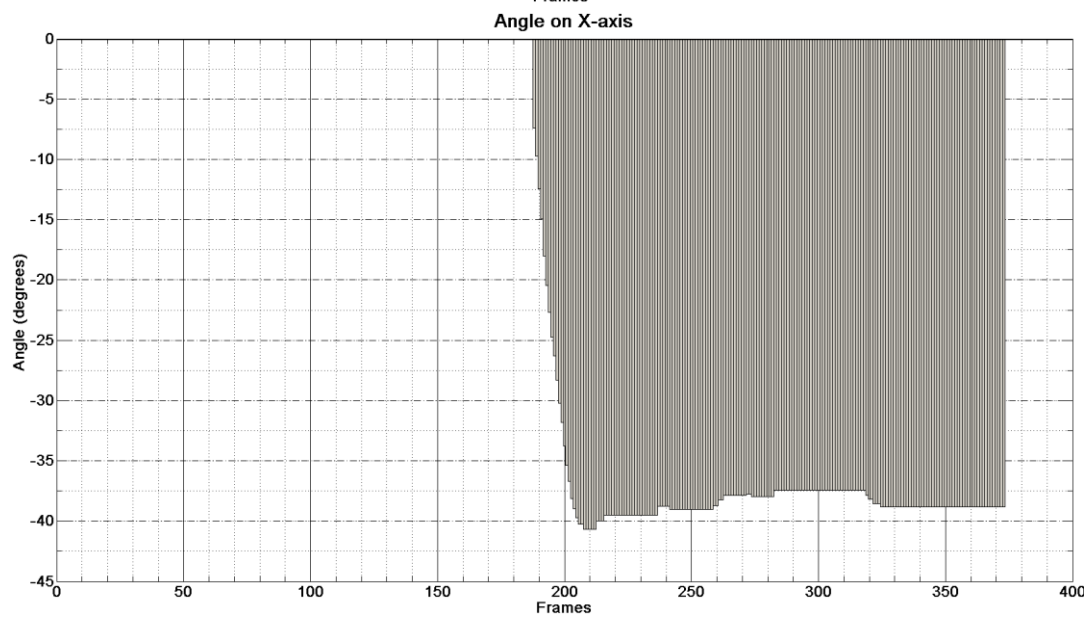
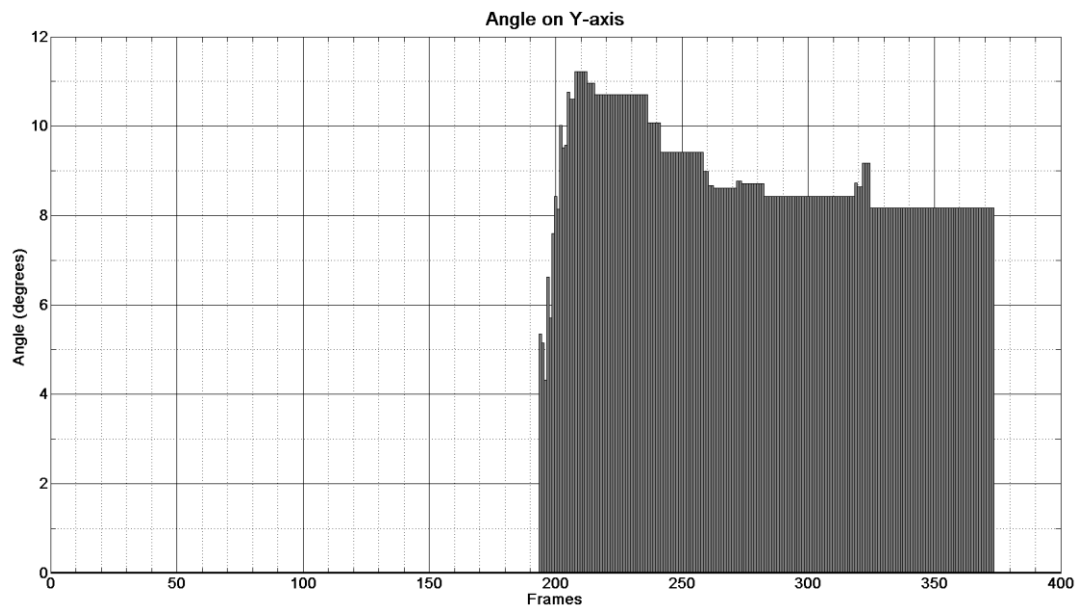
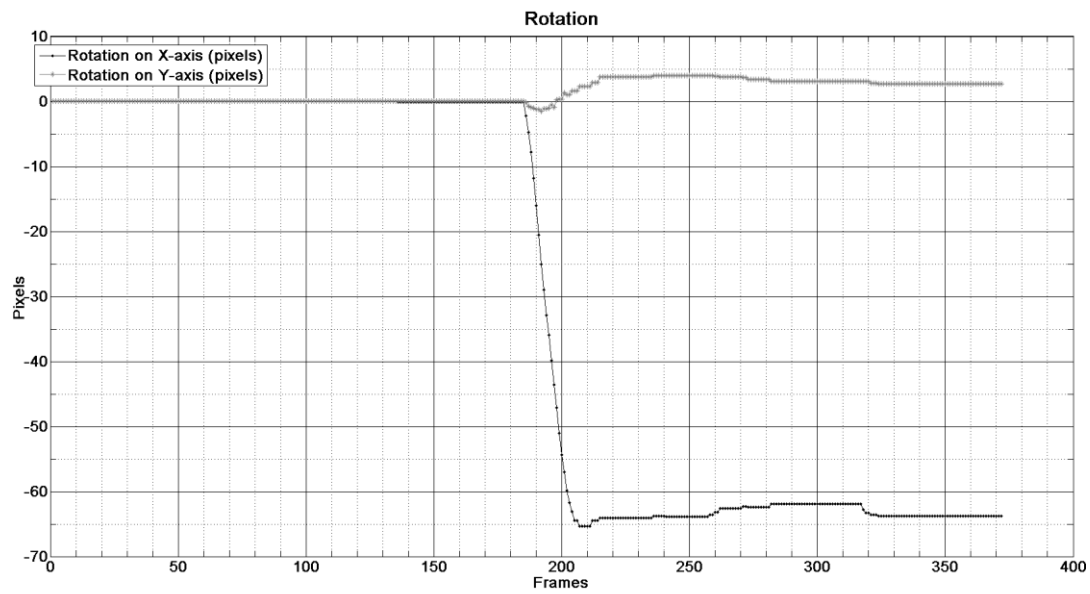




Ανάλυση	Χρωματική παλέτα	Συμπίεση	Αριθμός καρέ	Frames per sec	Ρύθμιση χρωμ. Επιπέδων (levels adjustment)	Καρε αποτυχίας ανίχνευσης (Bad Frames)
640x480	RGB24	όχι	373	30fps	✖	-

Παρατηρήσεις:

Στο video5b οι εκτιμήσεις έγιναν με τις προκαθορισμένες τιμές των παραμέτρων ενώ η ρύθμιση χρωματικών επιπέδων ήταν απενεργοποιημένη. Το υποκείμενο στο καρέ 185 πραγματοποιεί περιστροφή του κεφαλιού προς τα δεξιά, η οποία εκτιμάται ορθά από τον αλγόριθμο με μέγιστη τιμή τις 40,7°. Η θέση του κάτω LED δεν είναι η σωστή, όπως αναφέραμε και στο προηγούμενο βίντεο με το ίδιο άτομο, αλλά εδώ δεν παρατηρούνται σφάλματα και οι εκτιμήσεις πραγματοποιούνται με επιτυχία.



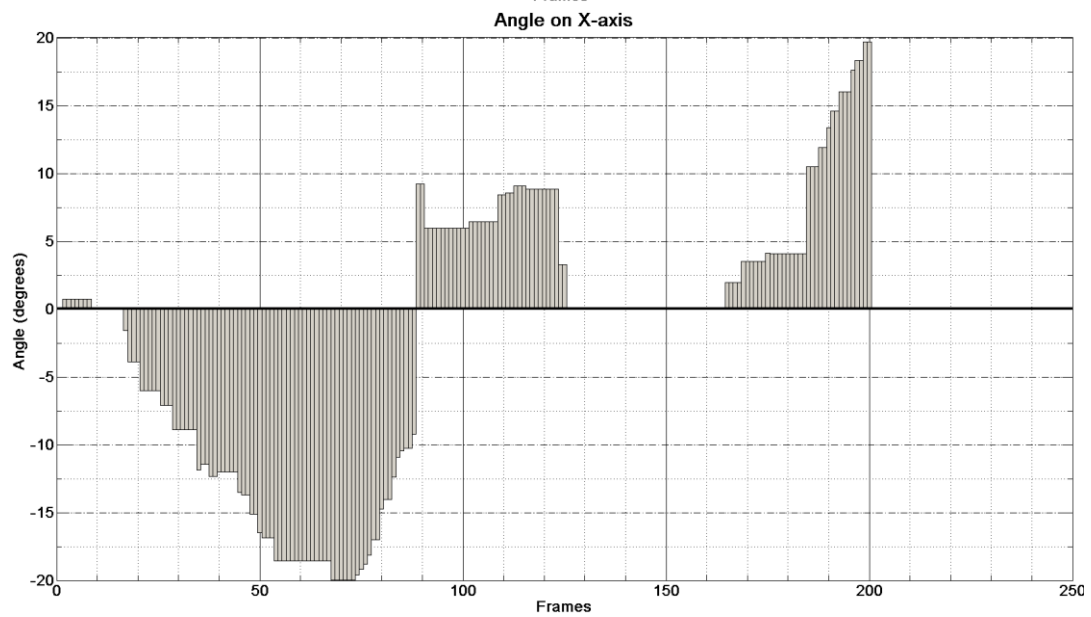
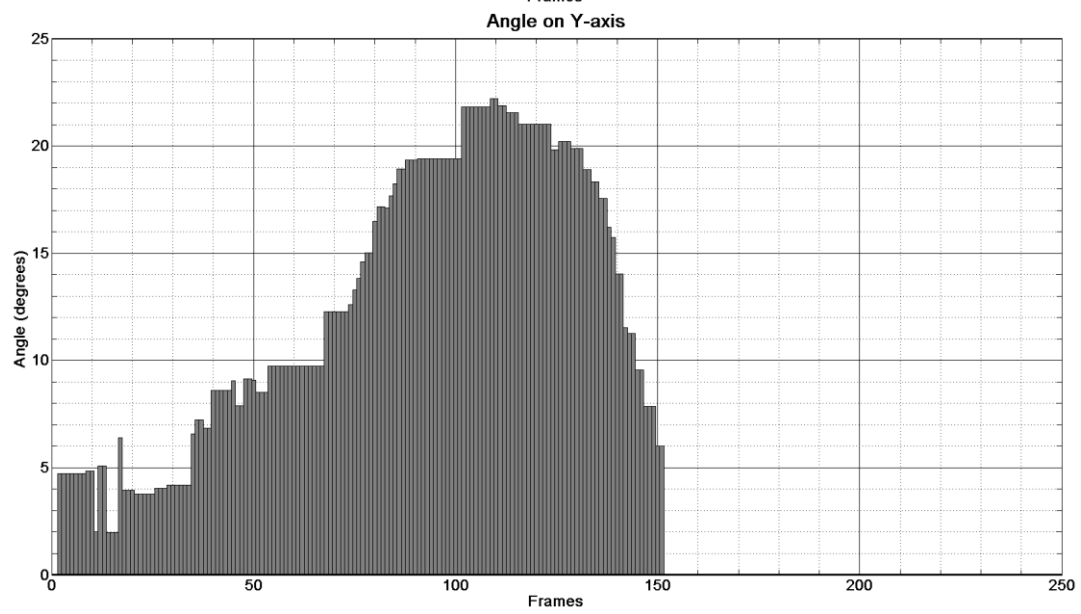
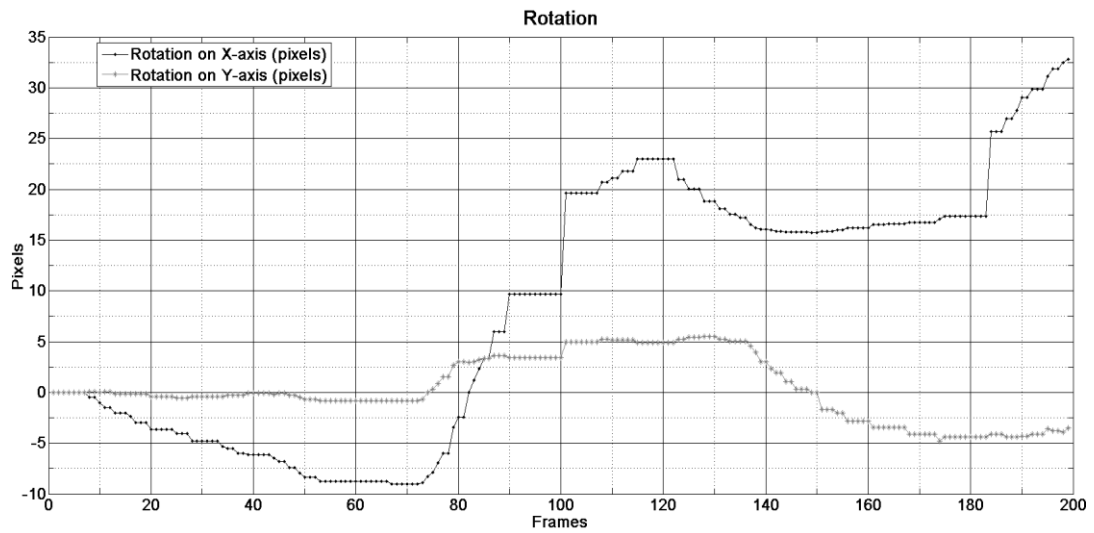


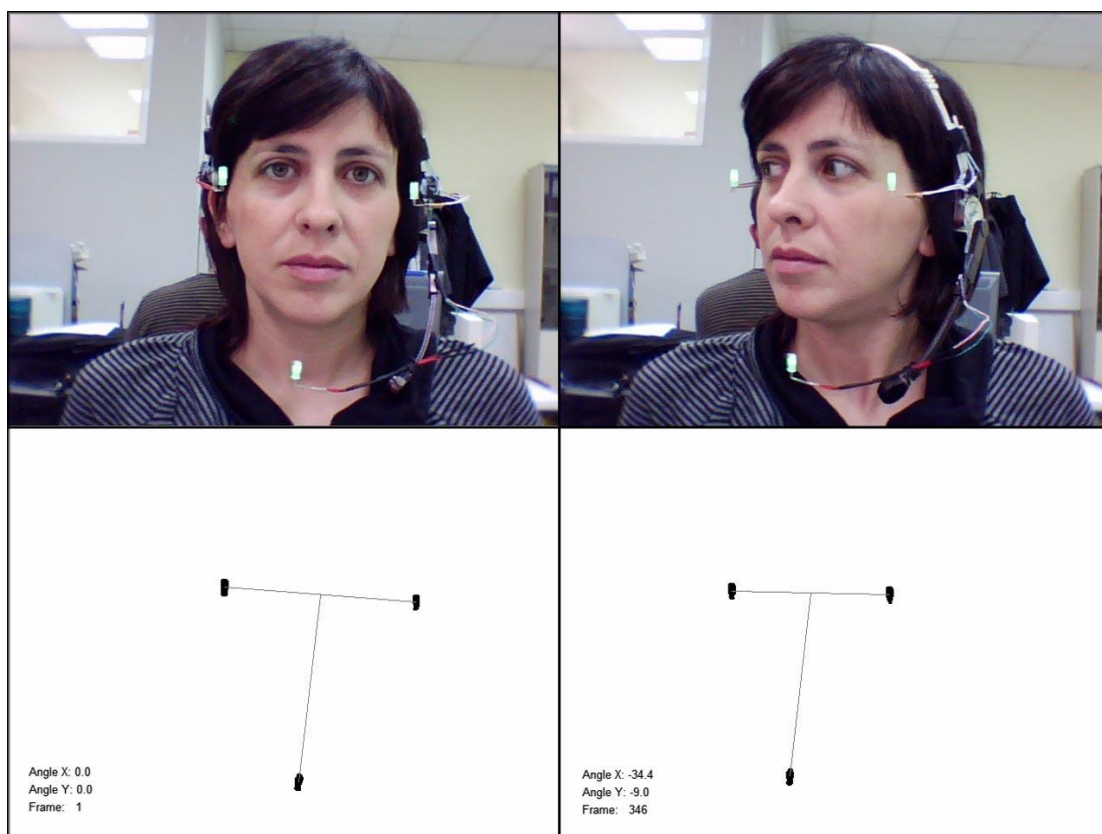
Ανάλυση	Χρωματική παλέτα	Συμπίεση	Αριθμός καρέ	Frames per sec	Ρύθμιση χρωμ. Επιπέδων (levels adjustment)	Καρε αποτυχίας ανίχνευσης (Bad Frames)
640x480	RGB24	όχι	200	30fps	*	-

Παρατηρήσεις:

Και στο video6a χρησιμοποιήθηκαν οι προκαθορισμένες τιμές των παραμέτρων και δεν έγινε χρήση της ρύθμισης χρωματικών επιπέδων. Το υποκείμενο πραγματοποιεί ελεύθερη κίνηση κατά βούληση και επιλέγει να στρέψει το κεφάλι του αρχικά δεξιά και κάτω, επανέρχεται μετωπικά με γωνία προς τα κάτω, δεξιά και κάτω, μετωπικά και τέλος στρέφεται αριστερά. Η μέθοδος εκτίμησης της πόζας σε εικονοστοιχεία ακολουθεί αρκετά καλά τις κινήσεις του υποκειμένου όπως φαίνεται και στο διάγραμμα περιστροφής. Οι γωνίες στον Y-άξονα εκτιμούνται σωστά, με σωστή κλίση προς τα κάτω, ομαλή διαβάθμιση και μέγιστη τιμή τις 21,9 μοίρες. Μετά το 150 καρέ το υποκείμενο επιστρέφει στις 0 μοίρες και παραμένει εκεί για το

υπόλοιπο του βίντεο. Στον Χ-άξονα παρατηρούμε σωστή ανίχνευση προσήμου και ομαλή διαβάθμιση με μόνη εξαίρεση μία απότομη αλλαγή προσήμου στο 89 καρέ, που είναι όμως εντός των ορίων ασάφειας. Από το 125 έως το 165 καρέ φαίνεται η μηδενική περιστροφή στον Χ άξονα και η σταδιακή επαναφορά του κεφαλιού του αντικειμένου στις 0 μοίρες στον Υ άξονα. Τέλος, οι μέγιστες γωνίες περιστροφής για τον Χ-άξονα είναι 20° αριστερά και 20° μοίρες δεξιά.

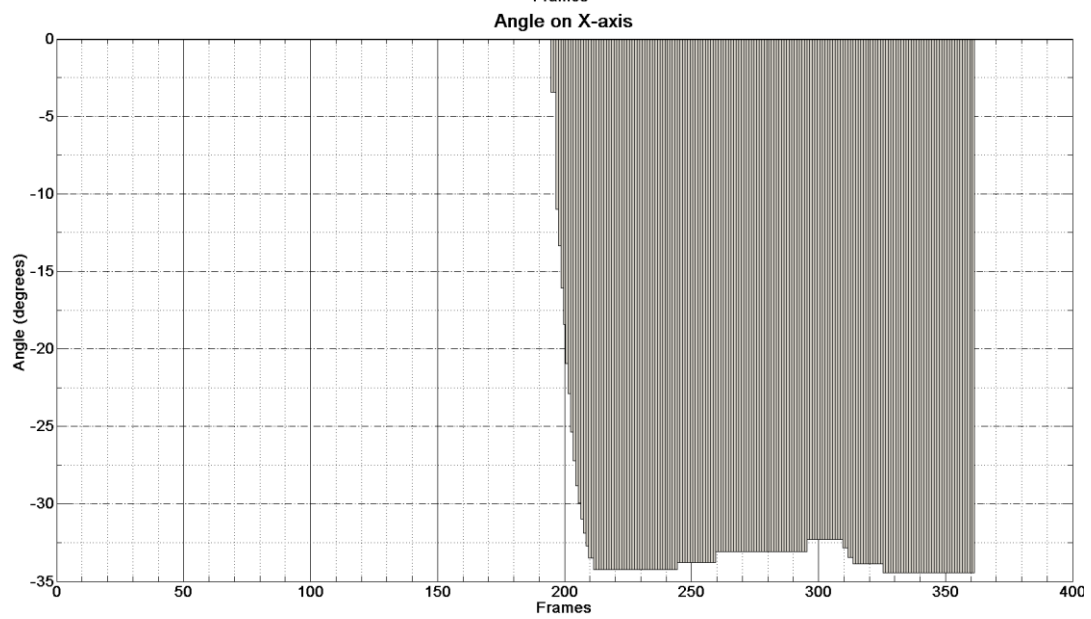
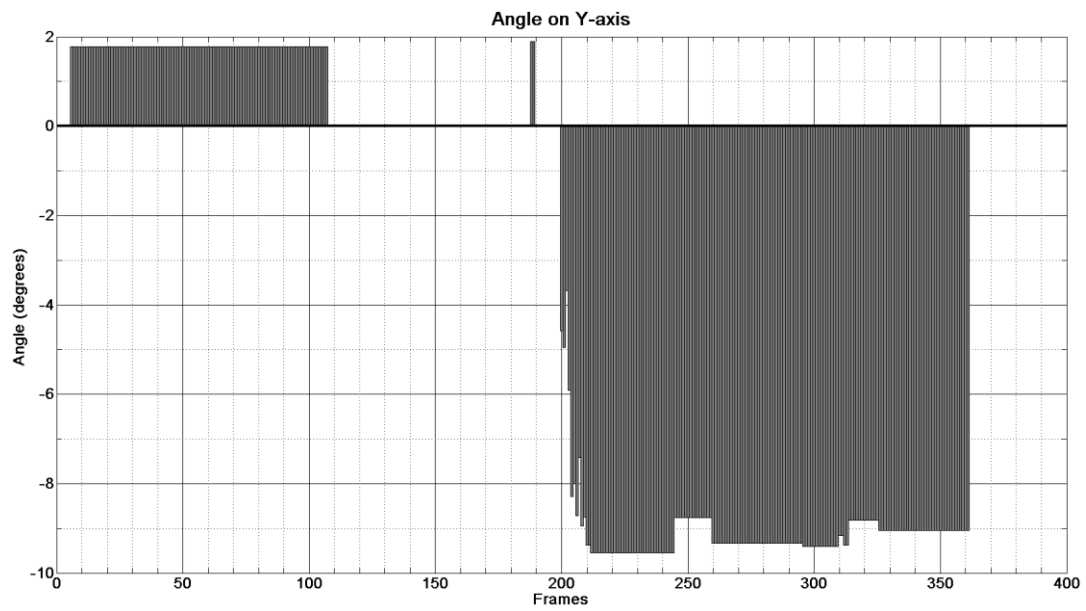
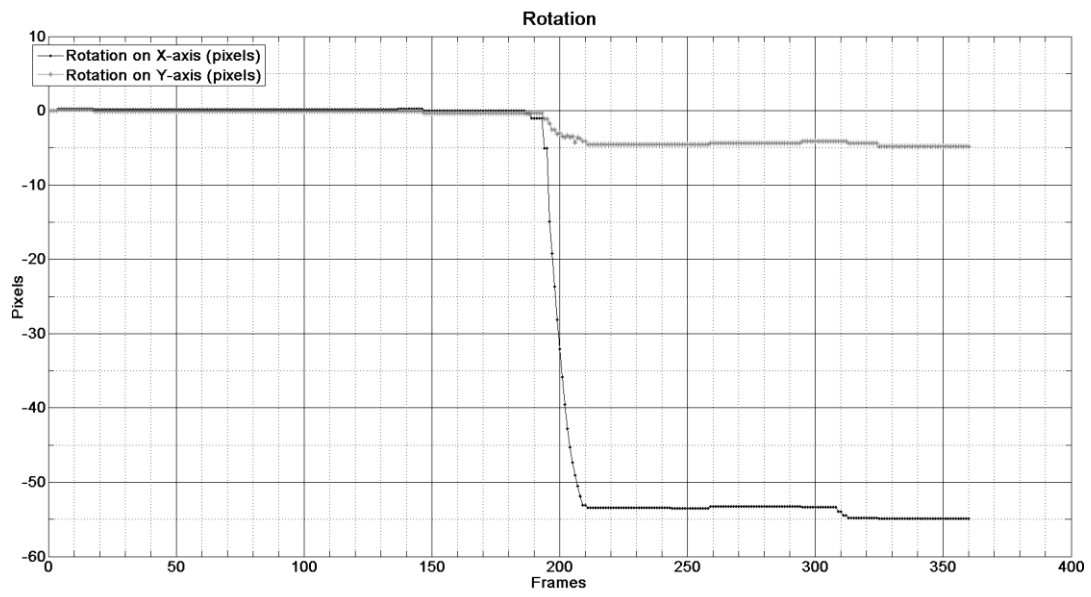


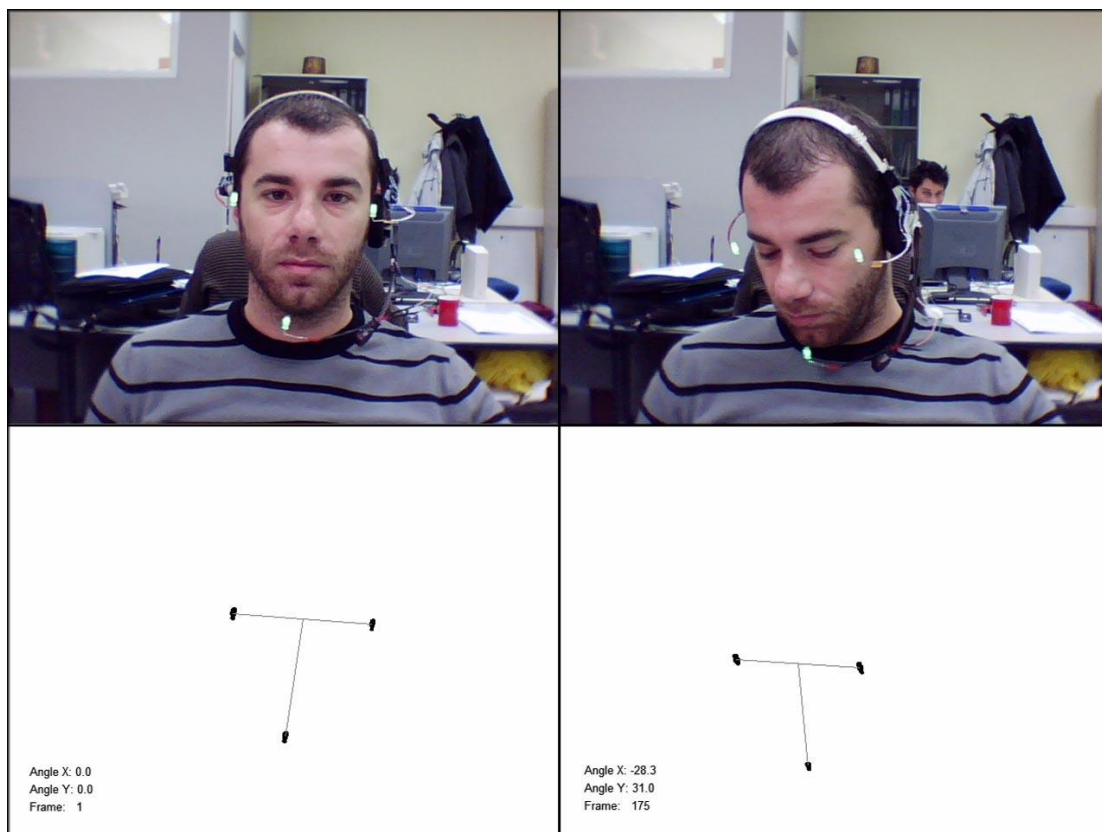


Ανάλυση	Χρωματική παλέτα	Συμπίεση	Αριθμός καρέ	Frames per sec	Ρύθμιση χρωμ. Επιπέδων (levels adjustment)	Καρε αποτυχίας ανίχνευσης (Bad Frames)
640x480	RGB24	όχι	360	30fps	×	-

Παρατηρήσεις:

Στο video6b δεν χρειάστηκε να γίνει καμία αλλαγή στις τιμές των παραμέτρων και η ρύθμιση χρωματικών επιπέδων ήταν απενεργοποιημένη. Το υποκείμενο περιστρέφει το κεφάλι του προς τα δεξιά στο 192 καρέ. Ο αλγόριθμος εκτιμά την κατεύθυνση χωρίς σφάλματα και στην μέτρηση σε εικονοστοιχεία και στην μέτρηση της γωνίας στους 2 άξονες, όπως φαίνεται και στα διαγράμματα που ακολουθούν. Κάποιες πολύ μικρές γωνίες που καταγράφονται όταν το κεφάλι του υποκειμένου είναι στην μετωπική πόζα είναι εντός των ορίων ανοχής, ενώ η μέγιστη γωνία περιστροφής είναι 34,4° δεξιά στον οριζόντιο άξονα και 9° πάνω στον κάθετο άξονα.



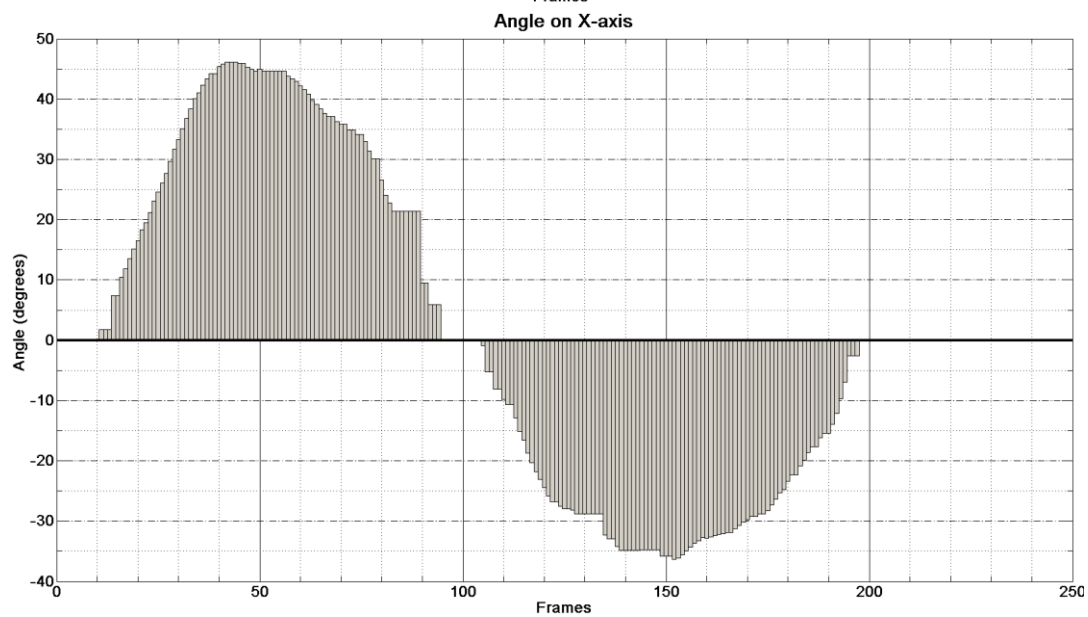
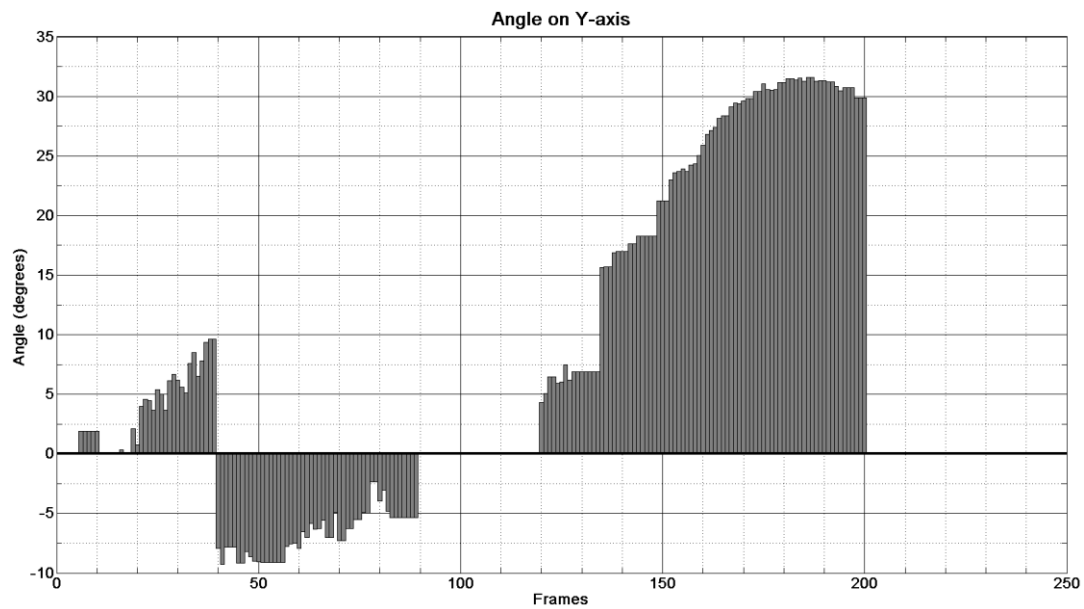
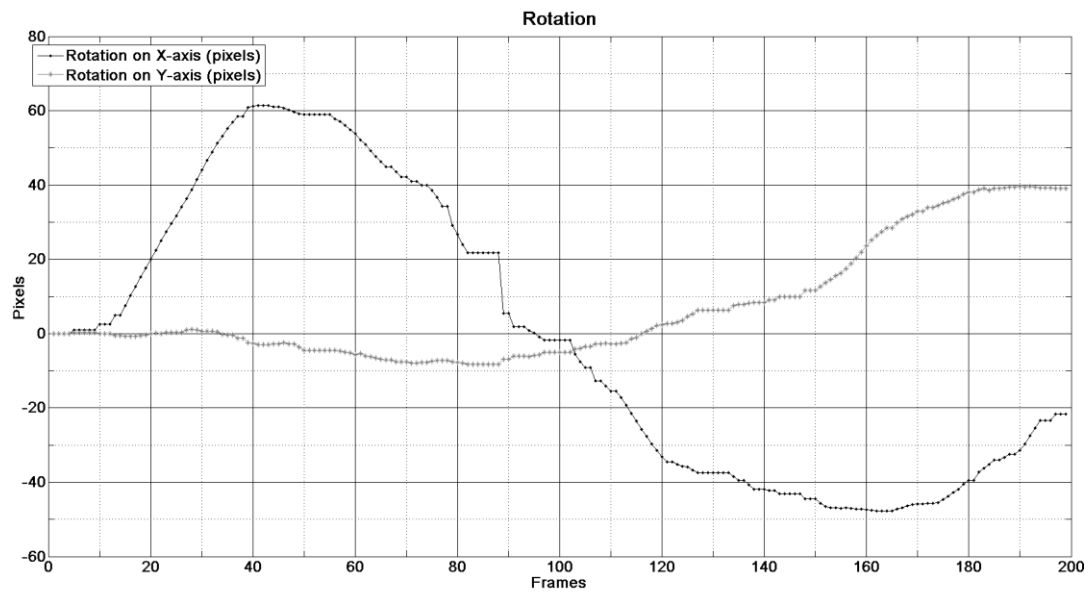


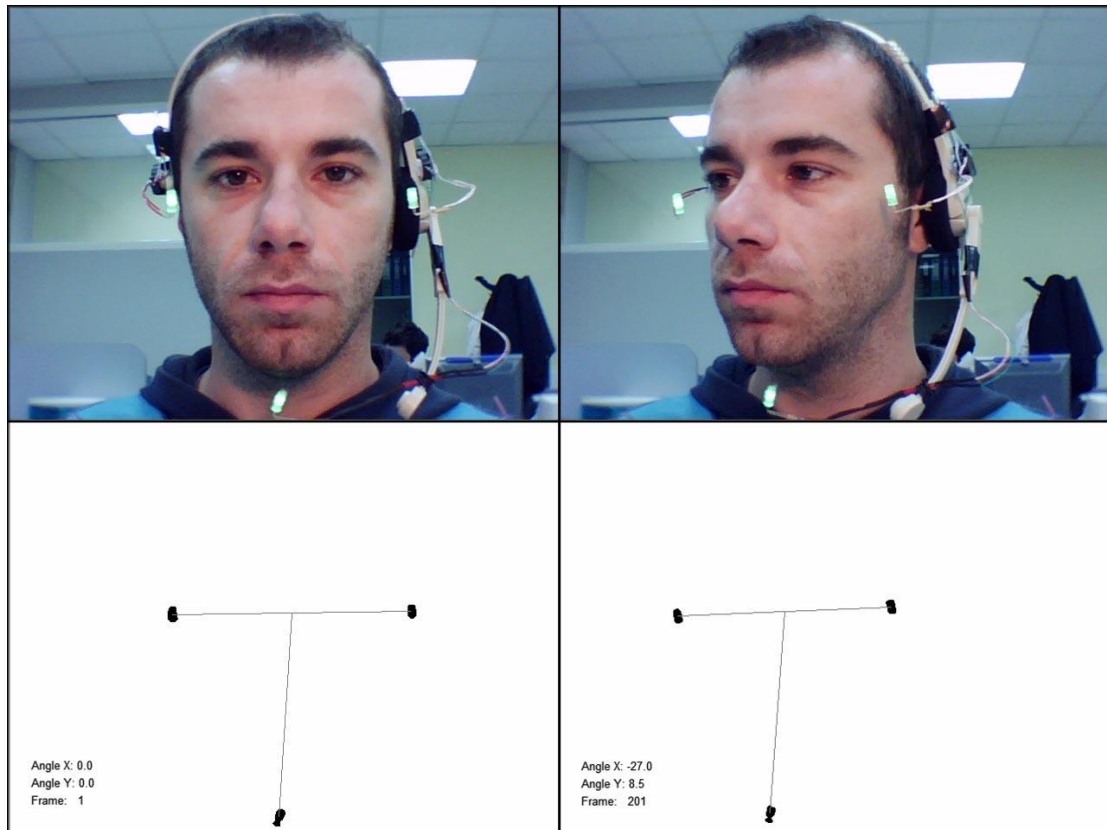
Ανάλυση	Χρωματική παλέτα	Συμπίεση	Αριθμός καρέ	Frames per sec	Ρύθμιση χρωμ. Επιπέδων (levels adjustment)	Καρε αποτυχίας ανίχνευσης (Bad Frames)
640x480	RGB24	όχι	200	30fps	×	-

Παρατηρήσεις:

Στο video7a αυξήσαμε την τιμή της περιοχής ενδιαφέροντος (sizeAreaOfInterest) από 20 pixel που είναι η προκαθορισμένη τιμή στα 35 για να μειώσουμε τα bad frames που είχαμε με τις προκαθορισμένες ρυθμίσεις. Όλες οι υπόλοιπες τιμές των παραμέτρων ήταν οι προκαθορισμένες και η ρύθμιση χρωματικών επιπέδων απενεργοποιημένη. Το υποκείμενο επιλέγει να περιστρέψει το κεφάλι του αρχικά αριστερά με κάποια μικρή κλίση προς τα πάνω και στην συνέχεια περιστρέφεται αριστερά και κάτω. Δεν πραγματοποιείται καμία μεταφορική κίνηση. Οι διαδοχικές στάσεις ανιχνεύονται χωρίς σημαντικά σφάλματα από τον αλγόριθμο. Συγκεκριμένα, στο διάγραμμα της γωνίας στον Y-άξονα παρατηρούμε μία απότομη αλλαγή προσήμου

σε μη μηδενική γωνία, αλλά επειδή πραγματοποιείται εντός της περιοχής ασάφειας δεν αξιολογείται ως σφάλμα. Μετά το 120 καρέ το υποκείμενο κλίνει το κεφάλι του προς τα κάτω με την μέγιστη γωνία να φτάνει τις $36,3^\circ$. Στο διάγραμμα της γωνίας στον Χ-άξονα παρατηρούμε ομαλά διαβαθμισμένη εκτίμηση των διαδοχικών στάσεων του κεφαλιού, χωρίς σφάλματα στην εκτίμηση της πλευράς περιστροφής και στο σημείο που αλλάζει. Οι μέγιστες γωνίες που καταγράφονται είναι $46,1^\circ$ αριστερά και $36,1^\circ$ δεξιά.

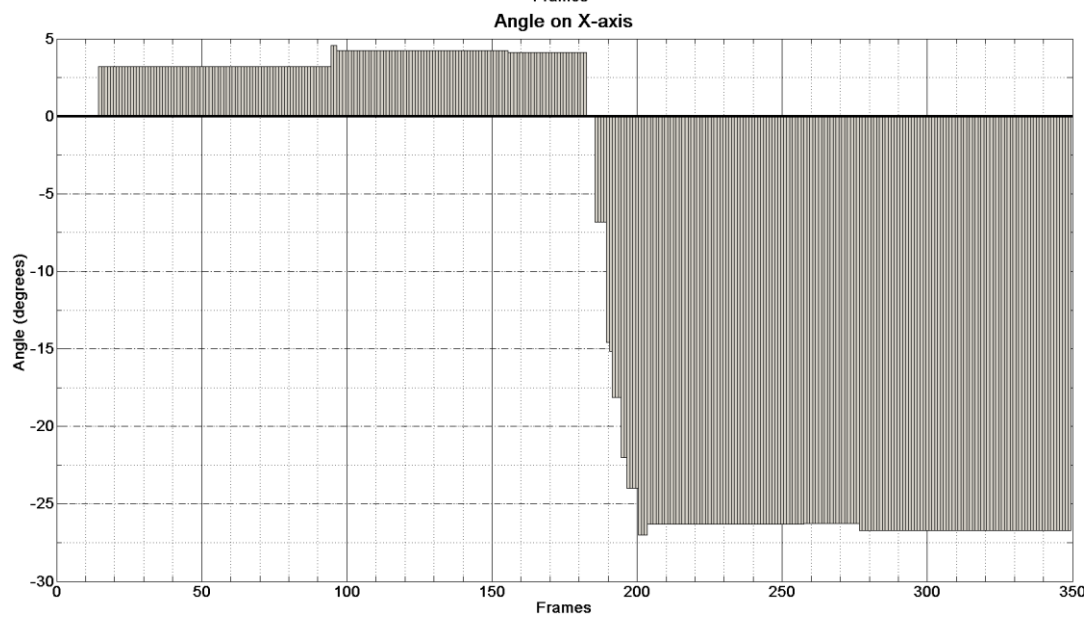
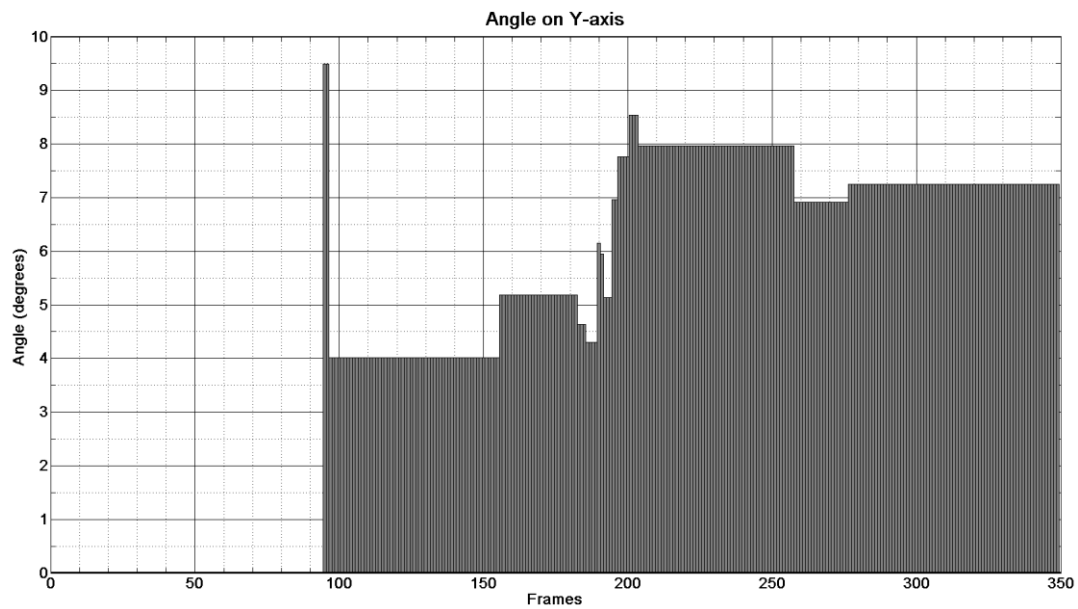
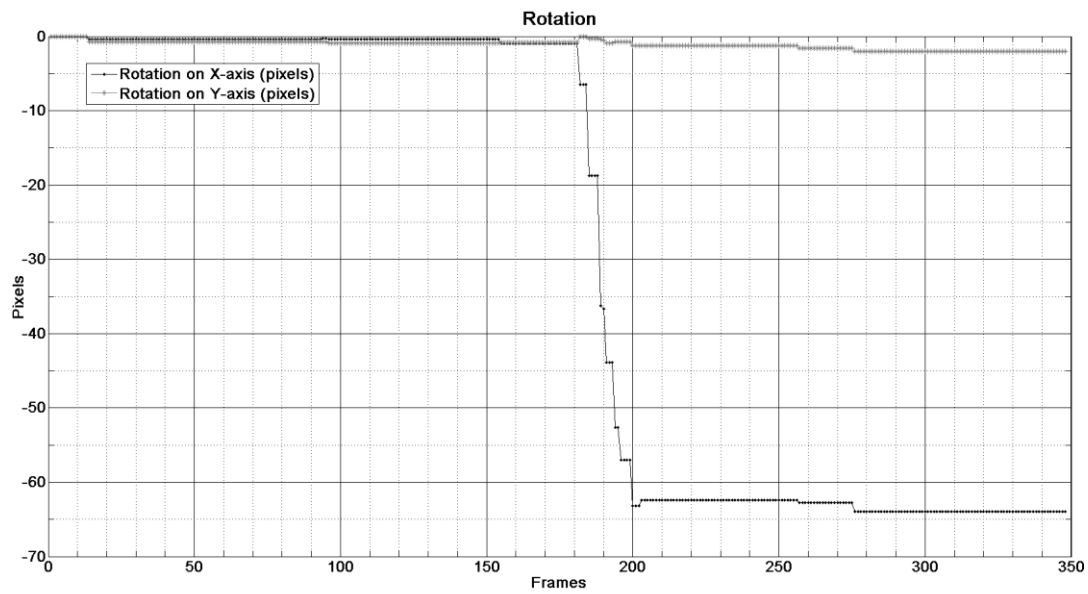


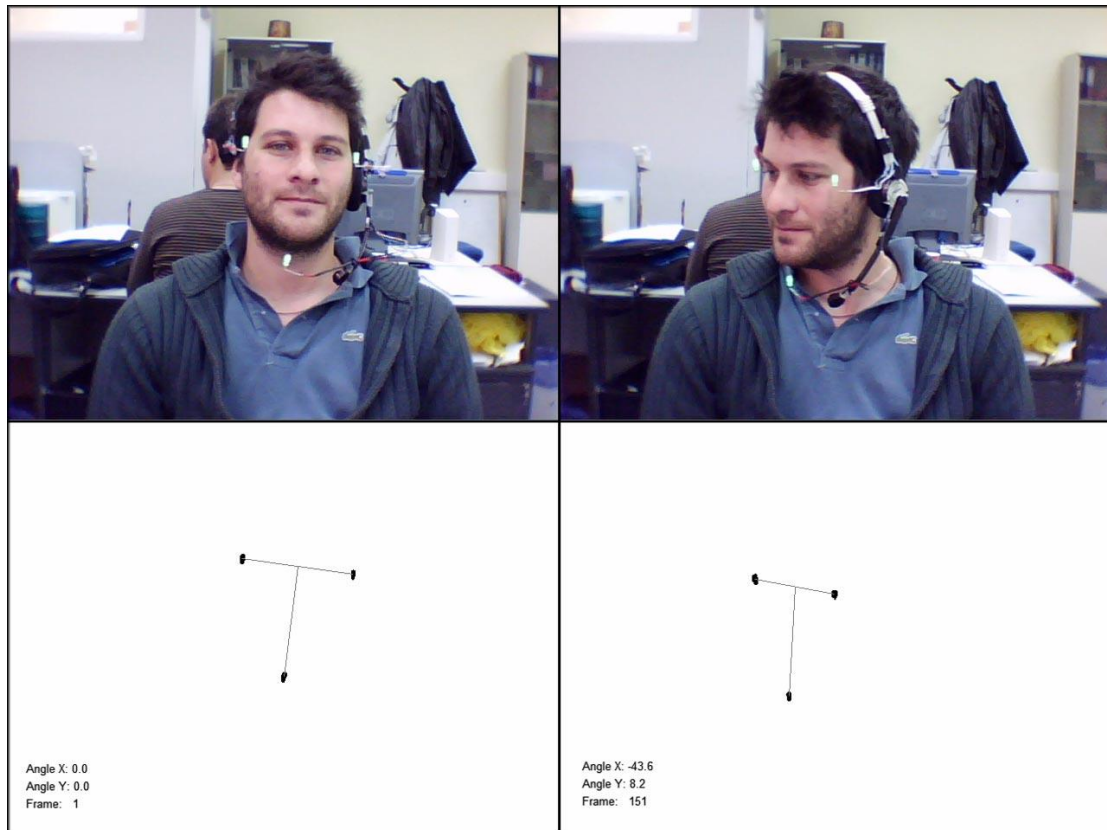


Ανάλυση	Χρωματική παλέτα	Συμπίεση	Αριθμός καρέ	Frames per sec	Ρύθμιση χρωμ. Επιπέδων (levels adjustment)	Καρε αποτυχίας ανίχνευσης (Bad Frames)
640x480	RGB24	όχι	349	30fps	*	-

Παρατηρήσεις:

Στο video7b χρειάστηκε να γίνει αλλαγή στην τιμή της περιοχής ανίχνευσης (sizeAreaOfInterest) στα 25 pixel ώστε να μην υπάρχει κανένα bad frame, ενώ οι υπόλοιπες παράμετροι παρέμειναν στις προκαθορισμένες τιμές τους και η ρύθμιση χρωματικών επιπέδων ήταν απενεργοποιημένη. Το υποκείμενο περιστρέφει το κεφάλι του προς τα δεξιά στο 177 καρέ. Ο αλγόριθμος εκτιμά την κατεύθυνση χωρίς σφάλματα και στην μέτρηση σε εικονοστοιχεία και στην μέτρηση της γωνίας στους 2 άξονες, όπως φαίνεται και στα διαγράμματα που ακολουθούν. Κάποιες πολύ μικρές γωνίες που καταγράφονται όταν το κεφάλι του υποκειμένου είναι στην μετωπική πόζα είναι εντός των ορίων ανοχής, ενώ η μέγιστη γωνία περιστροφής είναι 26,7° δεξιά στον οριζόντιο άξονα και 8° κάτω στον κάθετο άξονα, πάλι εντός των ορίων ασάφειας.



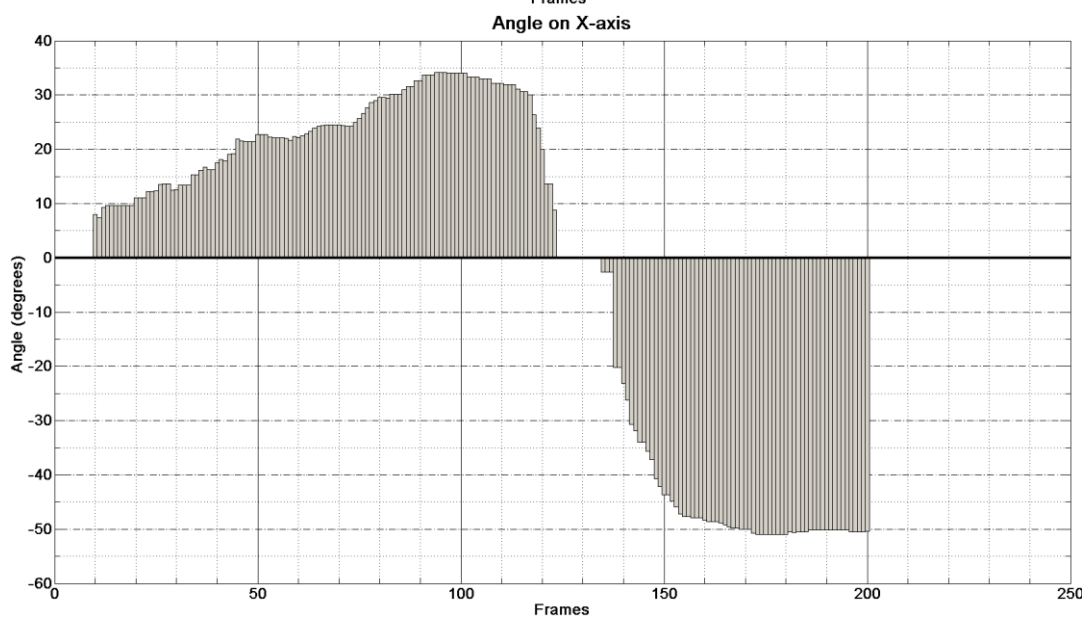
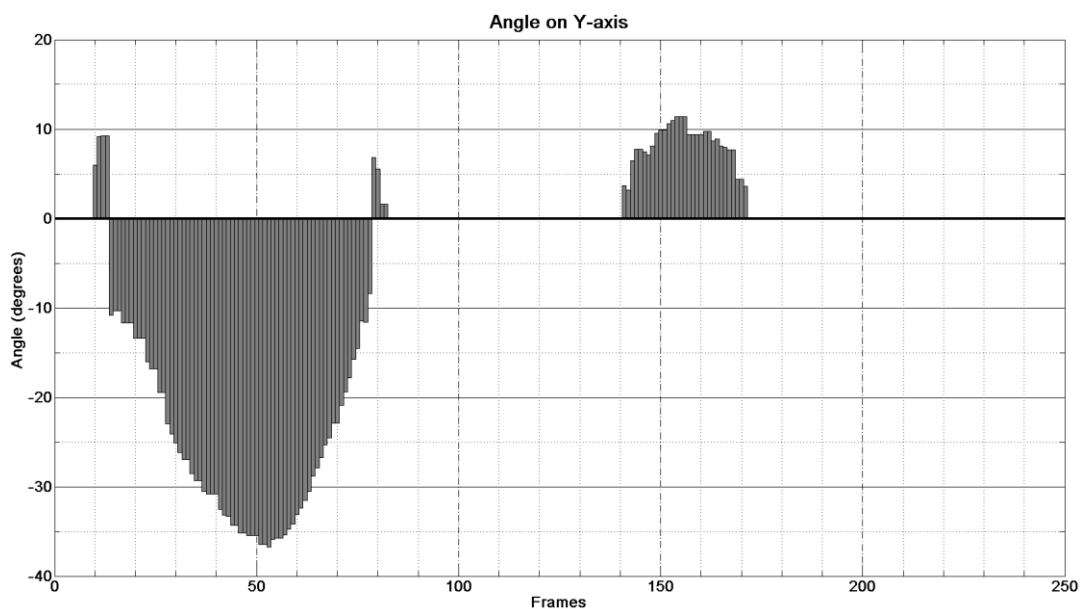
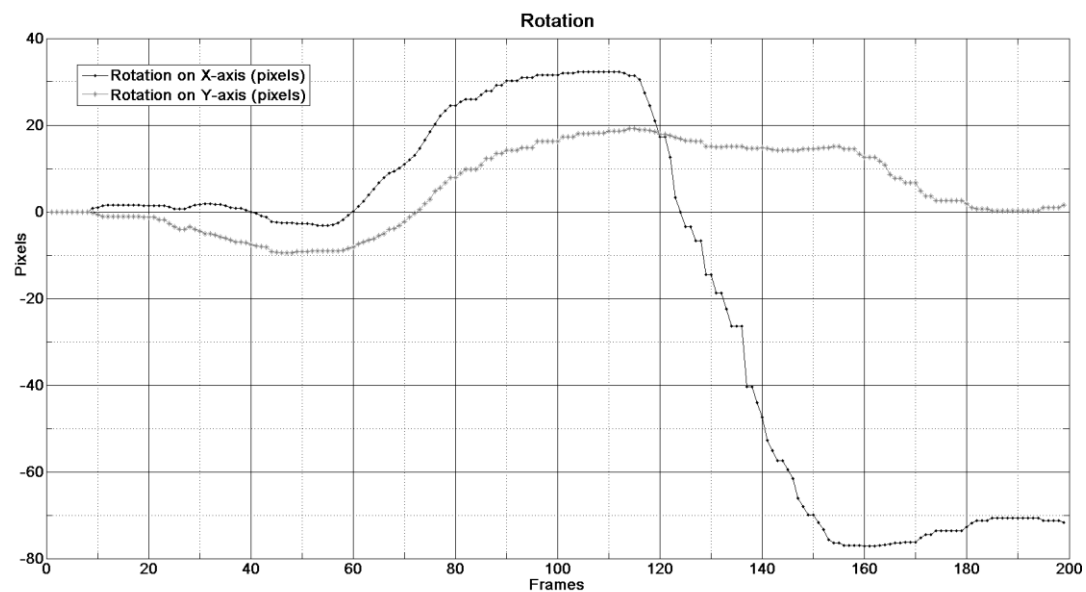


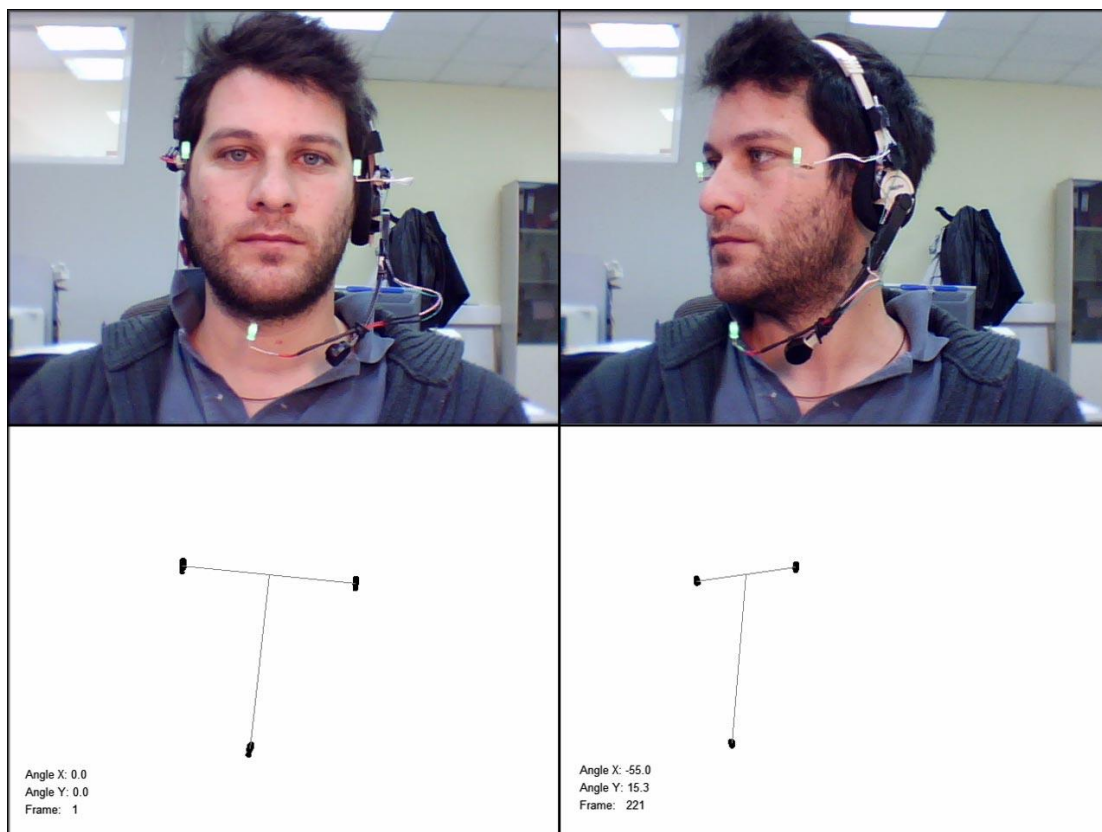
Ανάλυση	Χρωματική παλέτα	Συμπίεση	Αριθμός καρέ	Frames per sec	Ρύθμιση χρωμ. Επιπέδων (levels adjustment)	Καρε αποτυχίας ανίχνευσης (Bad Frames)
640x480	RGB24	όχι	200	30fps	✖	-

Παρατηρήσεις:

Στο video8a χρησιμοποιήθηκαν οι προκαθορισμένες τιμές των παραμέτρων και δεν έγινε χρήση της ρύθμισης χρωματικών επιπέδων. Το υποκείμενο πραγματοποιεί ελεύθερη κίνηση κατά βούληση και επιλέγει να στρέψει το κεφάλι του αρχικά αριστερά και πάνω, αριστερά, δεξιά και τέλος δεξιά και πάνω με μία συνεχόμενη κίνηση και διαφορετικές κατευθύνσεις στο βλέμμα του. Δεν πραγματοποιεί μεταφορική κίνηση. Η μέθοδος εκτίμησης της πόζας σε εικονοστοιχεία ακολουθεί αρκετά καλά τις κινήσεις του υποκειμένου όπως φαίνεται και στο διάγραμμα περιστροφής. Οι γωνίες στον Y-άξονα εκτιμούνται σωστά από ένα σημείο και μετά, με σωστή κλίση προς τα πάνω, ομαλή διαβάθμιση και μέγιστη τιμή τις 36,2 μοίρες. Μετά

το 170 καρέ ο αλγόριθμος εσφαλμένα δεν εκτιμά την κλίση προς τα πάνω που παίρνει σταδιακά το κεφάλι του υποκειμένου και την εκτιμά ίση με μηδέν. Στον Χ-άξονα παρατηρούμε σωστή ανίχνευση προσήμου και ομαλή διαβάθμιση με μόνη εξαίρεση μία απότομη αύξηση της τιμής της γωνίας στο 138 καρέ από 2° σε 20° . Ωστόσο, το άλμα αυτό είναι απολύτως φυσιολογικό γιατί η κάμερα παγώνει για μερικά καρέ και συνεπώς χάνει μέρος της κίνησης. Τέλος, οι μέγιστες γωνίες περιστροφής για τον Χ-άξονα είναι $34,2^{\circ}$ αριστερά και 51° μοίρες δεξιά.

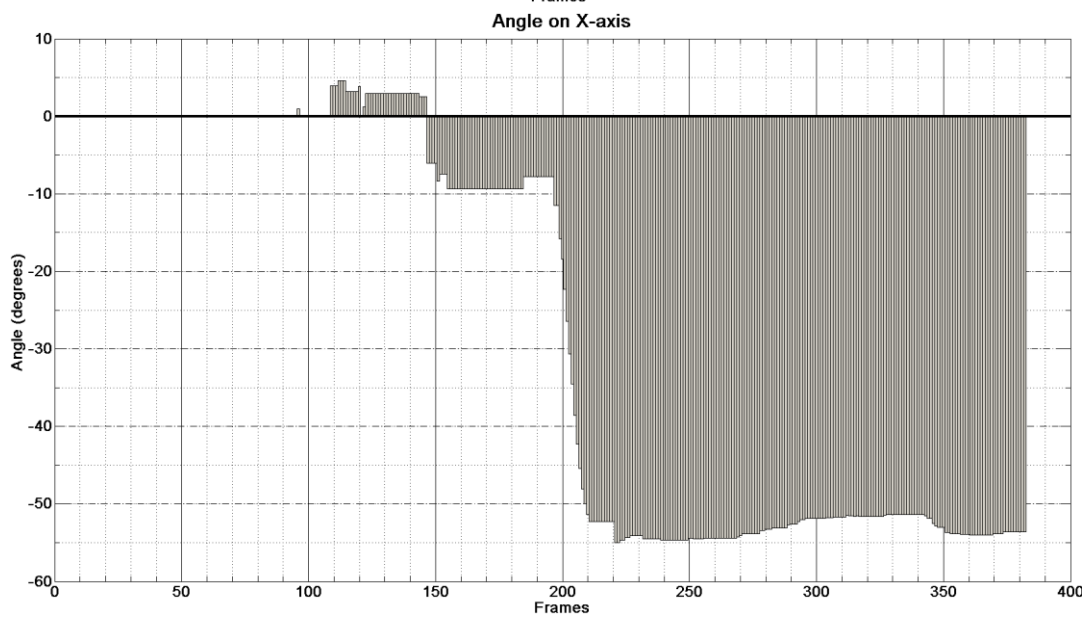
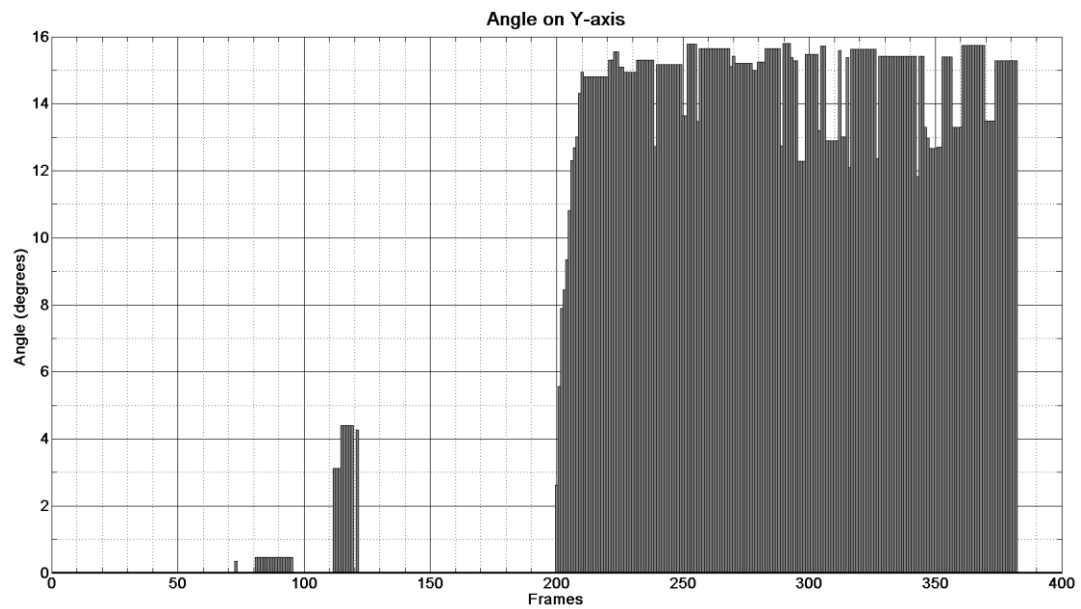
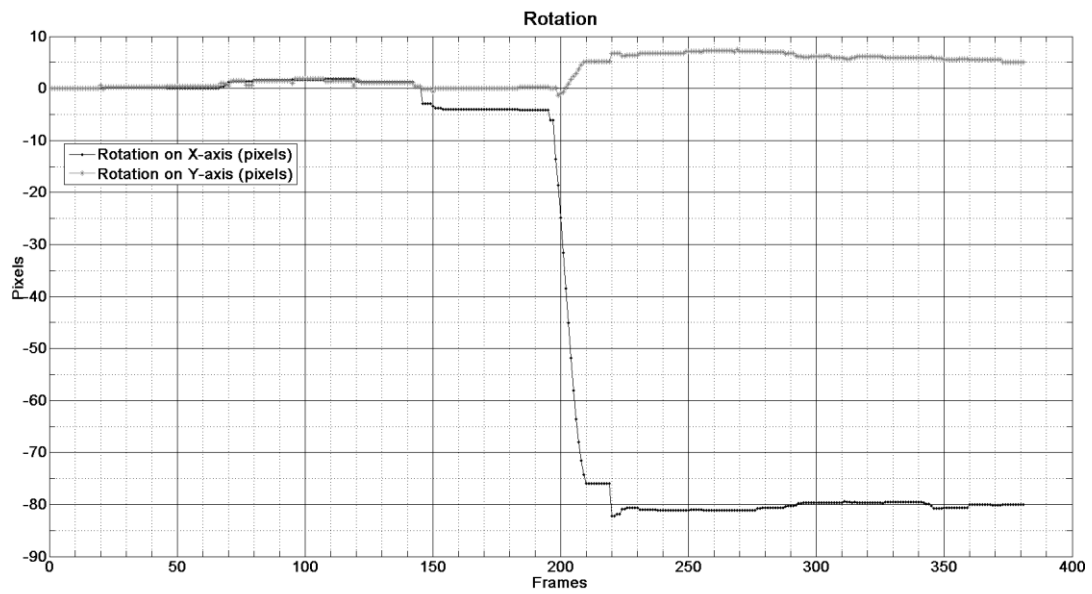


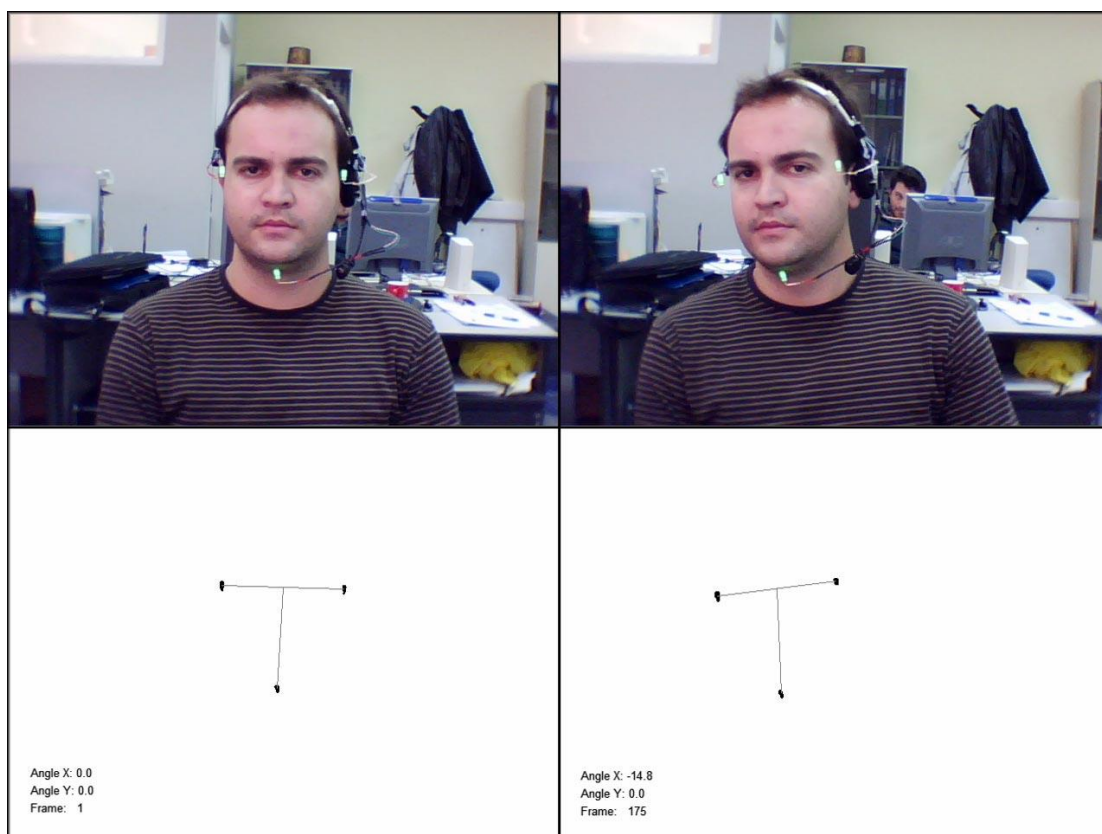


Ανάλυση	Χρωματική παλέτα	Συμπίεση	Αριθμός καρέ	Frames per sec	Ρύθμιση χρωμ. Επιπέδων (levels adjustment)	Καρε αποτυχίας ανίχνευσης (Bad Frames)
640x480	RGB24	όχι	382	30fps	×	-

Παρατηρήσεις:

Στο video8b δεν χρειάστηκε να γίνει καμία αλλαγή στις τιμές των παραμέτρων και η ρύθμιση χρωματικών επιπέδων ήταν απενεργοποιημένη. Το υποκείμενο περιστρέφει το κεφάλι του προς τα δεξιά στο 195 καρέ. Ο αλγόριθμος εκτιμά την κατεύθυνση χωρίς σφάλματα και στην μέτρηση σε εικονοστοιχεία και στην μέτρηση της γωνίας στους 2 άξονες, όπως φαίνεται και στα διαγράμματα που ακολουθούν. Κάποιες πολύ μικρές γωνίες που καταγράφονται όταν το κεφάλι του υποκειμένου είναι στην μετωπική πόζα είναι εντός των ορίων ανοχής, ενώ η μέγιστη γωνία περιστροφής είναι 54,7° δεξιά στον οριζόντιο άξονα και 15,6° κάτω στον κάθετο άξονα.



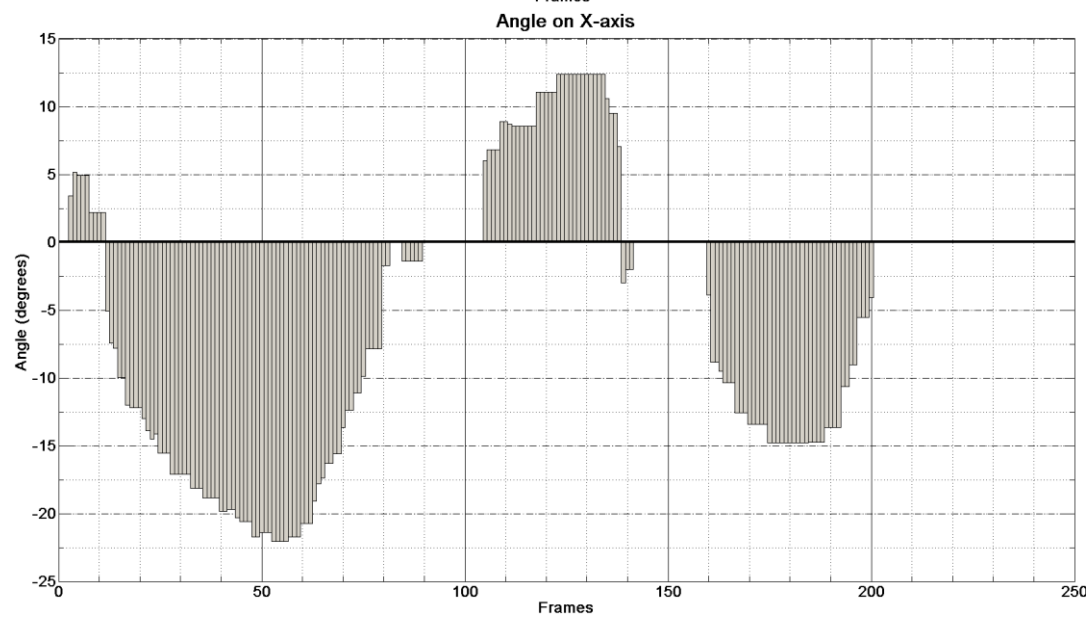
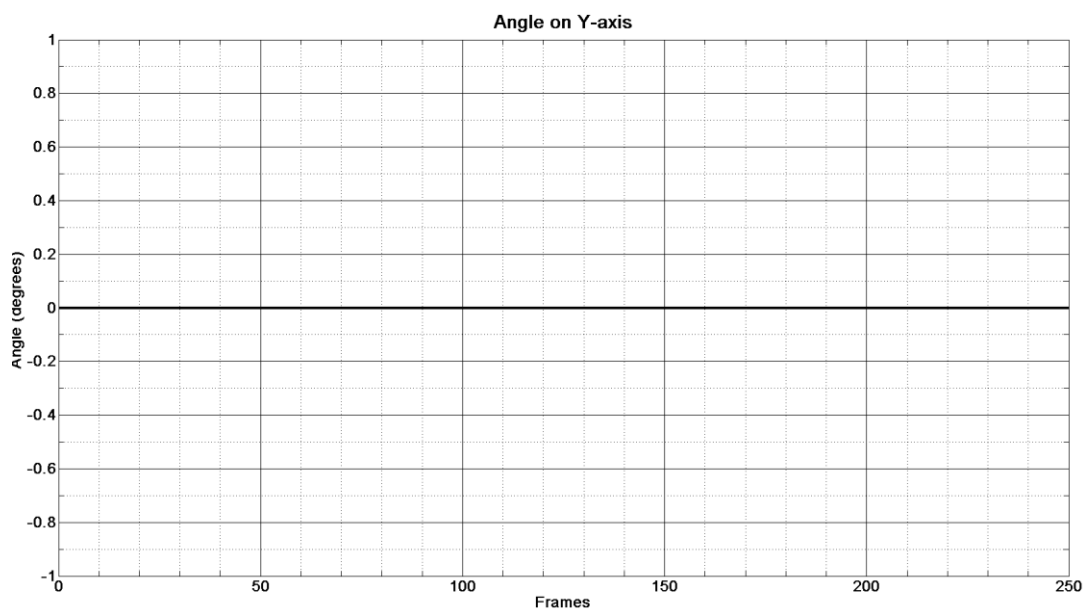
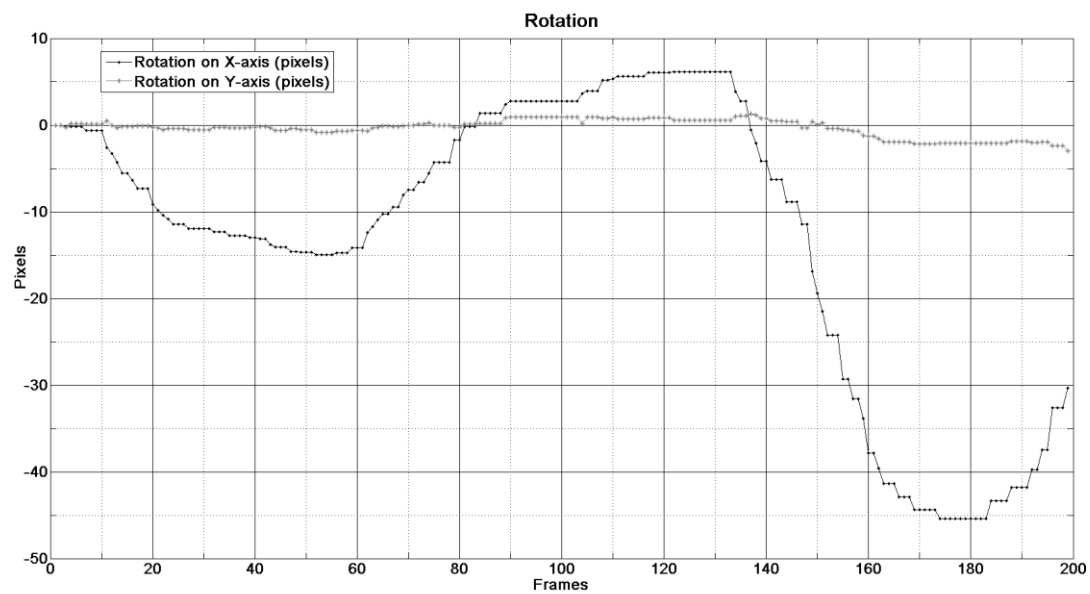


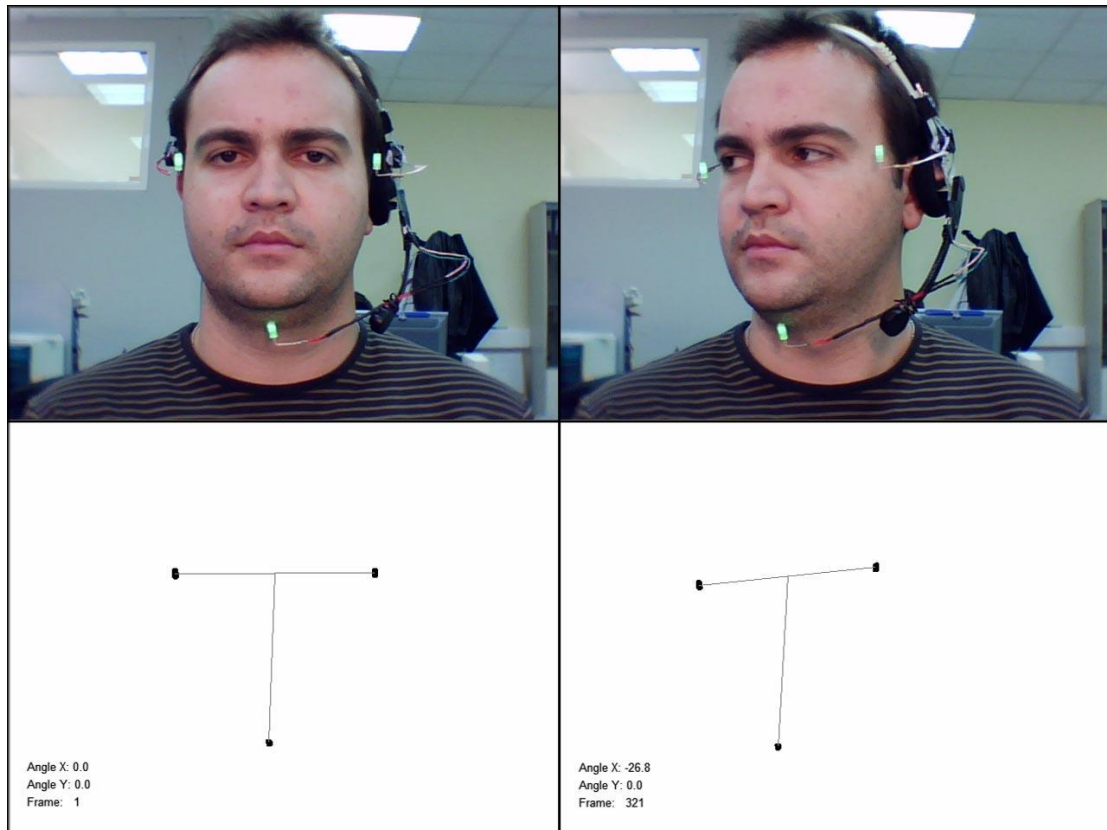
Ανάλυση	Χρωματική παλέτα	Συμπίεση	Αριθμός καρέ	Frames per sec	Ρύθμιση χρωμ. Επιπέδων (levels adjustment)	Καρε αποτυχίας ανίχνευσης (Bad Frames)
640x480	RGB24	όχι	200	30fps	×	-

Παρατηρήσεις:

Στο video9a έγινε αλλαγή της τιμής της περιοχής ανίχνευσης (sizeAreaOfInterest) σε 15 pixel ενώ για τις υπόλοιπες παραμέτρους χρησιμοποιήθηκαν οι προκαθορισμένες τιμές τους και δεν έγινε χρήση της ρύθμισης χρωματικών επιπέδων. Το υποκείμενο πραγματοποιεί ελεύθερη κίνηση κατά βούληση και επιλέγει να στρέψει το κεφάλι του αρχικά δεξιά, αριστερά και τέλος δεξιά με παράλληλη μεταφορική κίνηση προς τα δεξιά. Η κίνηση είναι συνεχόμενη και το υποκείμενο αλλάζει συχνά την κατεύθυνση του βλέμματός του. Η μέθοδος εκτίμησης της πόζας σε εικονοστοιχεία ακολουθεί αρκετά ορθά τις περιστροφές του κεφαλιού του υποκειμένου όπως φαίνεται και στο διάγραμμα περιστροφής, αλλά αδυνατεί να διαχωρίσει την

μεταφορική από την περιστροφική κίνηση μετά περίπου από το 140 καρέ. Οι γωνίες στον Y-άξονα εκτιμούνται σωστά μηδενικές. Στον X-άξονα παρατηρούμε σωστή ανίχνευση προσήμου και ομαλή διαβάθμιση. Μετά το 160 καρέ που γίνεται μαζί η μεταφορική και η περιστροφική κίνηση ο αλγόριθμος εκτιμά χωρίς σφάλματα την γωνία και μένει ανεπηρέαστος από την μεταφορική κίνηση. Τέλος, οι μέγιστες γωνίες περιστροφής για τον X-άξονα είναι $12,4^\circ$ αριστερά και $22,1^\circ$ μοίρες δεξιά.

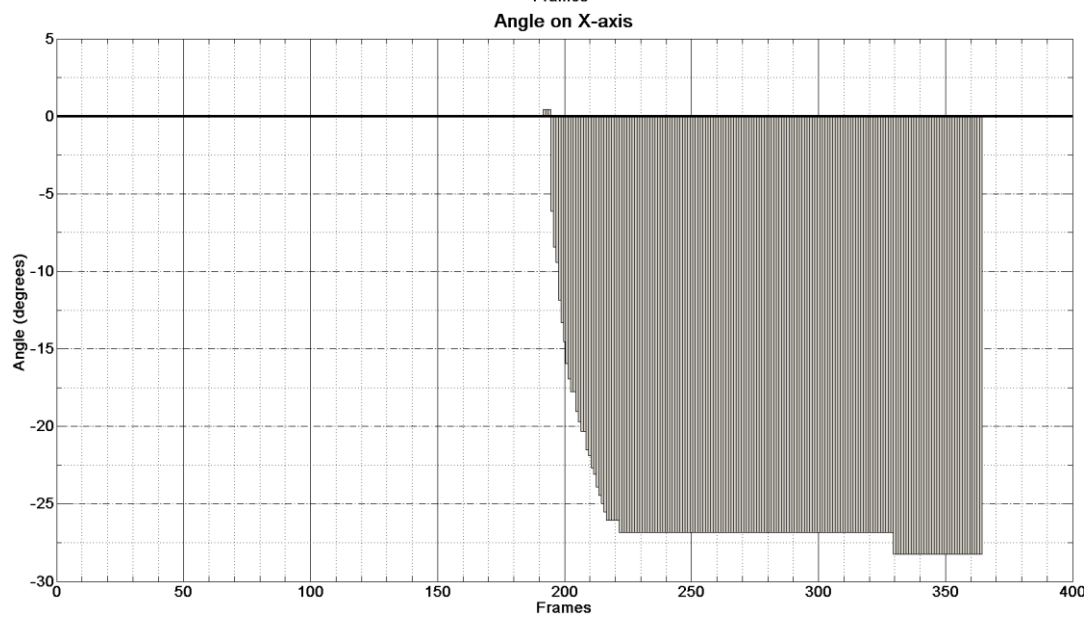
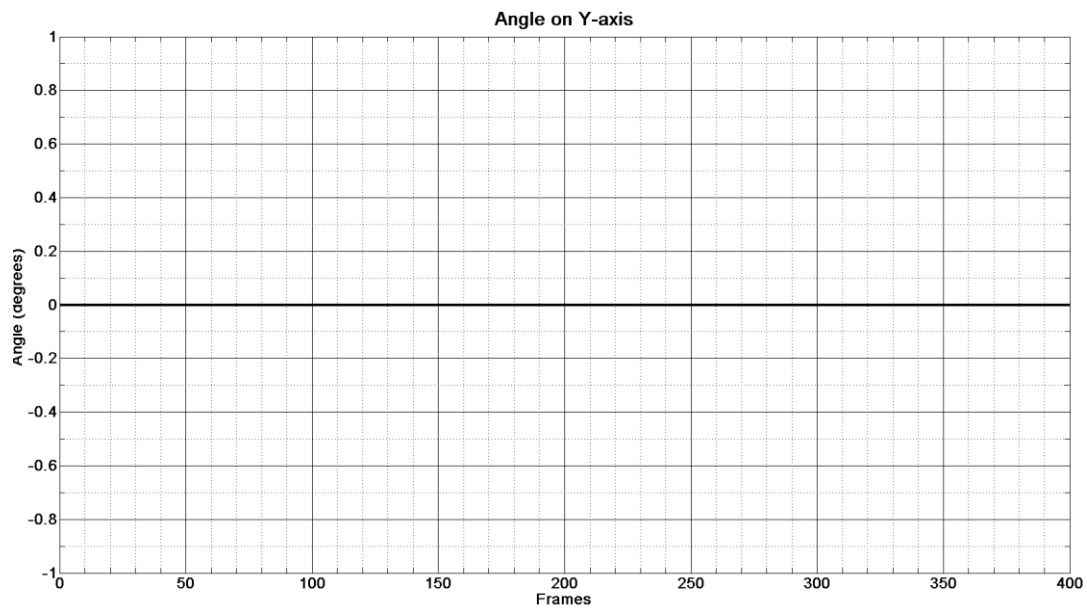
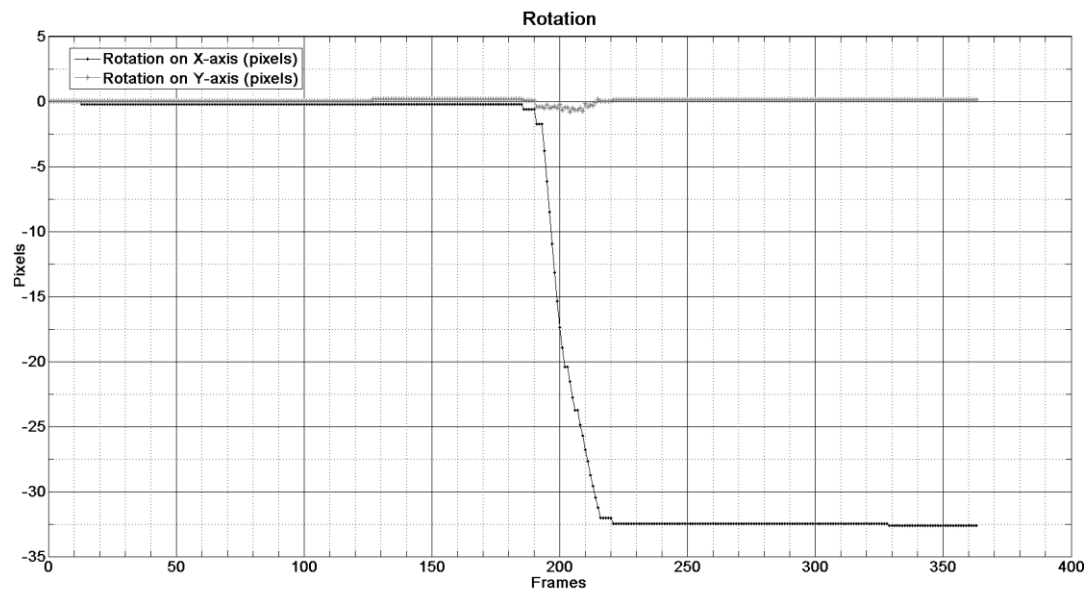


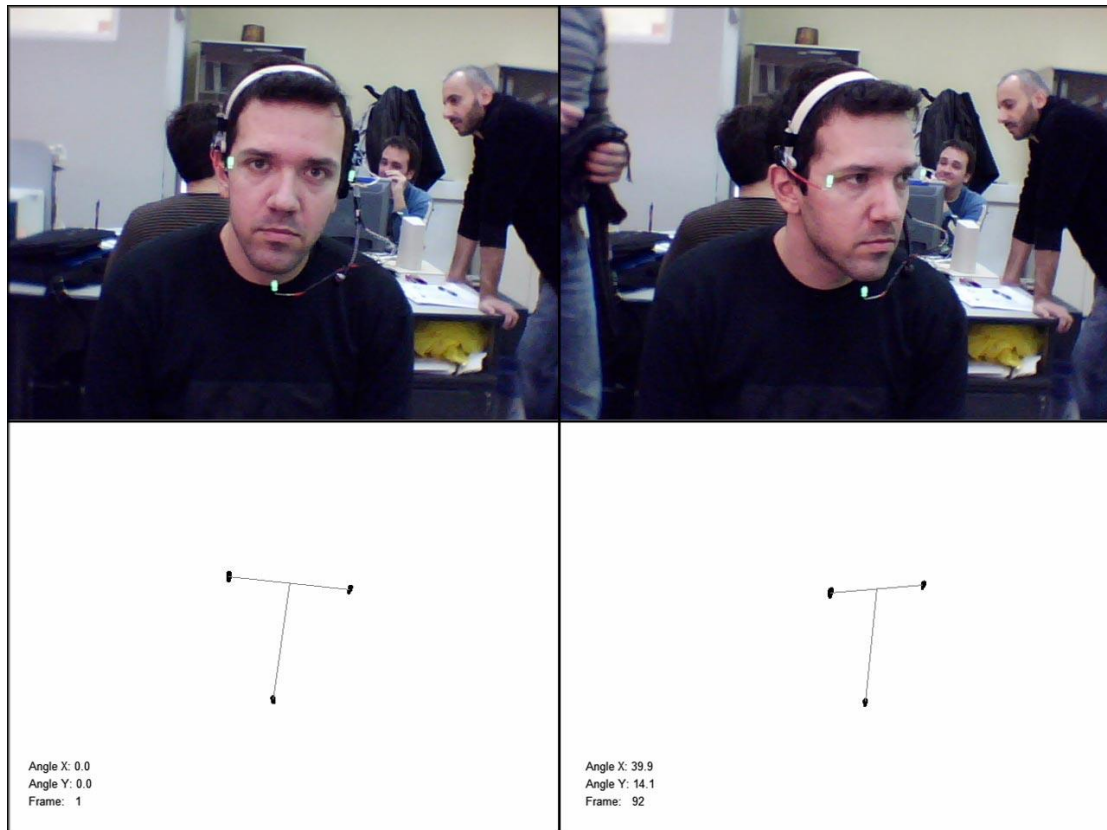


Ανάλυση	Χρωματική παλέτα	Συμπίεση	Αριθμός καρέ	Frames per sec	Ρύθμιση χρωμ. Επιπέδων (levels adjustment)	Καρε αποτυχίας ανίχνευσης (Bad Frames)
640x480	RGB24	όχι	364	30fps	×	-

Παρατηρήσεις:

Στο video9b δεν χρειάστηκε να γίνει καμία αλλαγή στις τιμές των παραμέτρων και η ρύθμιση χρωματικών επιπέδων ήταν απενεργοποιημένη. Το υποκείμενο περιστρέφει το κεφάλι του προς τα δεξιά στο 190 καρέ. Ο αλγόριθμος εκτιμά την κατεύθυνση χωρίς σφάλματα και στην μέτρηση σε εικονοστοιχεία και στην μέτρηση της γωνίας στους 2 άξονες, όπως φαίνεται και στα διαγράμματα που ακολουθούν. Δεν παρατηρούνται καθόλου μικρές γωνίες εντός της περιοχής ασάφειας καθώς και μεγάλη σταθερότητα στην στάση του κεφαλιού μετά την περιστροφή προς τα δεξιά, που σημαίνει ότι τα LEDs είχαν τοποθετηθεί με ακρίβεια στο κεφάλι του υποκειμένου, είχε πάρει ορθή αρχική θέση και η αλληλουχία των κινήσεων που έπρεπε να κάνει, πραγματοποιήθηκε με μεγάλη προσοχή και ακρίβεια από το υποκείμενο. Η μέγιστη γωνία περιστροφής είναι 26,8° δεξιά στον οριζόντιο άξονα και 0° στον κάθετο άξονα.



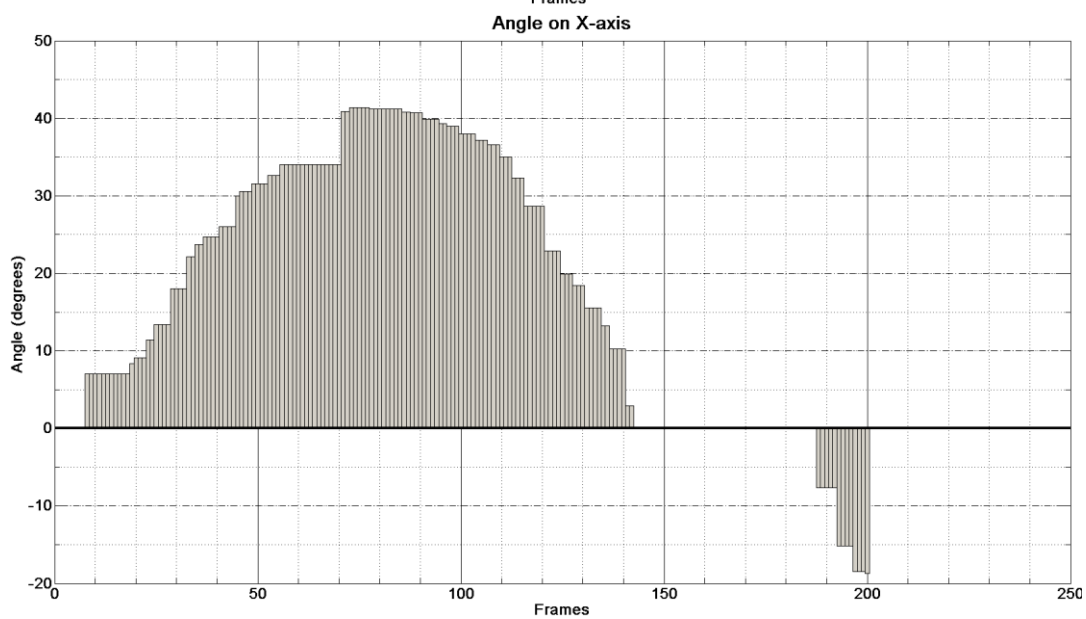
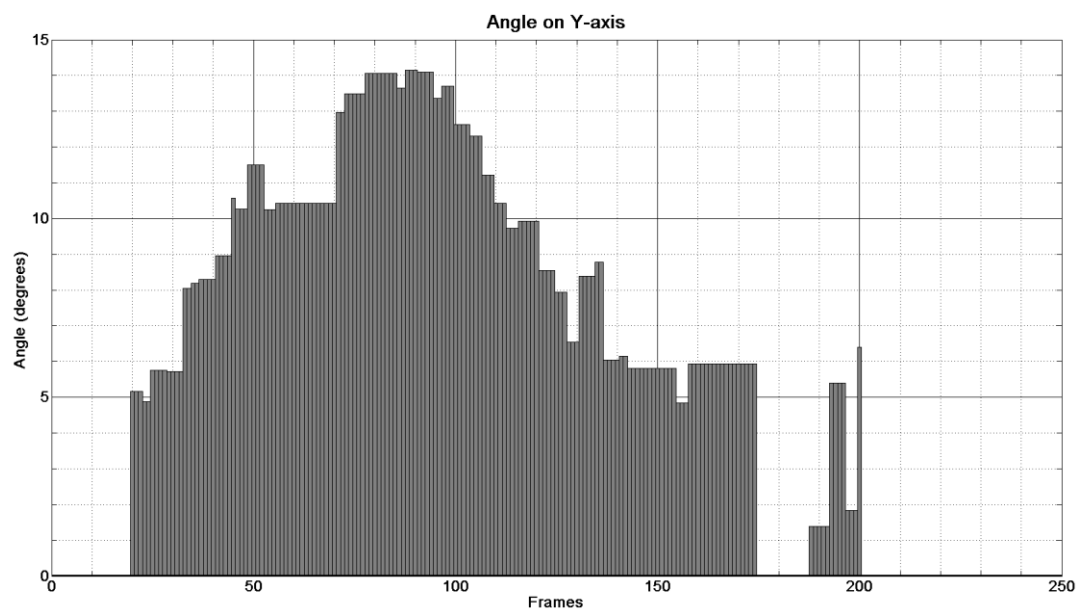
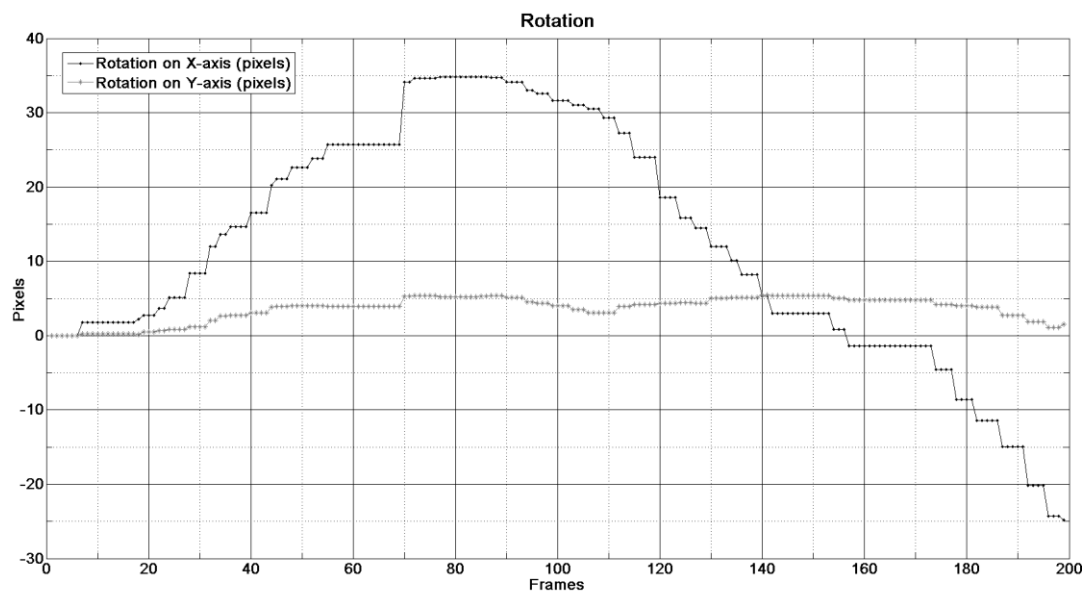


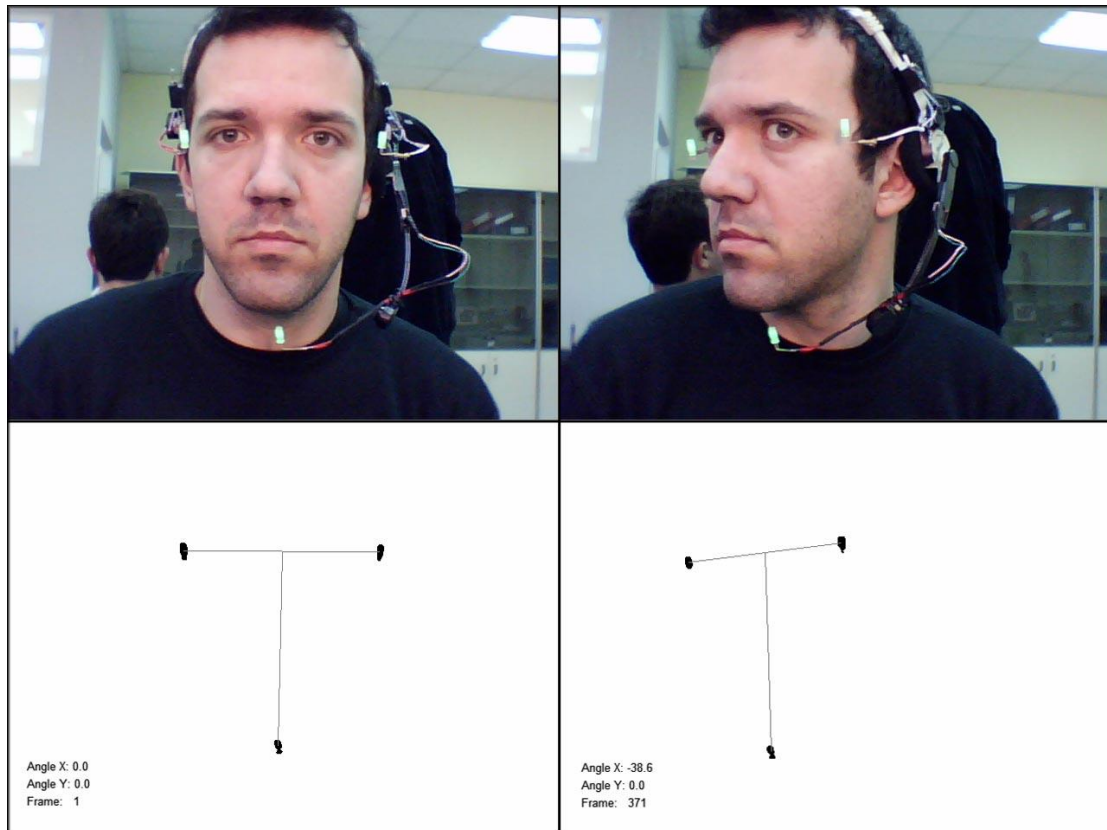
Ανάλυση	Χρωματική παλέτα	Συμπίεση	Αριθμός καρέ	Frames per sec	Ρύθμιση χρωμ. Επιπέδων (levels adjustment)	Καρε αποτυχίας ανίχνευσης (Bad Frames)
640x480	RGB24	όχι	200	30fps	*	-

Παρατηρήσεις:

Στο video10a χρησιμοποιήθηκαν οι προκαθορισμένες τιμές των παραμέτρων και δεν έγινε χρήση της ρύθμισης χρωματικών επιπέδων. Το υποκείμενο πραγματοποιεί ελεύθερη κίνηση κατά βούληση και επιλέγει να στρέψει το κεφάλι του αρχικά αριστερά και στην συνέχεια δεξιά με μία συνεχόμενη κίνηση και διαφορετικές κατευθύνσεις στο βλέμμα του. Δεν πραγματοποιεί μεταφορική κίνηση. Η μέθοδος εκτίμησης της πόζας σε εικονοστοιχεία ακολουθεί αρκετά καλά τις κινήσεις του υποκειμένου όπως φαίνεται και στο διάγραμμα περιστροφής. Οι γωνίες στον Y-άξονα εκτιμούνται χωρίς σφάλματα, με σωστή κλίση προς τα κάτω και μέγιστη τιμή τις 14°. Στον X-άξονα παρατηρούμε σωστή ανίχνευση προσήμου και ομαλή δια-

βάθμιση με εξαίρεση κάποιες απότομες αυξήσεις της τιμής της γωνίας σε κάποια καρέ, είτε λόγω γρήγορης κίνησης του κεφαλιού είτε επειδή η κάμερα πάγωνε και έχανε ένα μικρό τμήμα της κίνησης. Τέλος, οι μέγιστες γωνίες περιστροφής για τον Χ-άξονα είναι $41,4^\circ$ αριστερά και $18,5^\circ$ μοίρες δεξιά. Σημειώνουμε ότι το υποκείμενο αρχικά δεν έχει σταθεί σωστά με αποτέλεσμα η νοητή ευθεία μεταξύ των δύο πάνω LEDs να μην είναι παράλληλη με τον οριζόντιο άξονα Χ. Παρά την λάθος θέση, δεν παρατηρούνται αξιολογήσιμα σφάλματα στις εκτιμήσεις του αλγορίθμου.



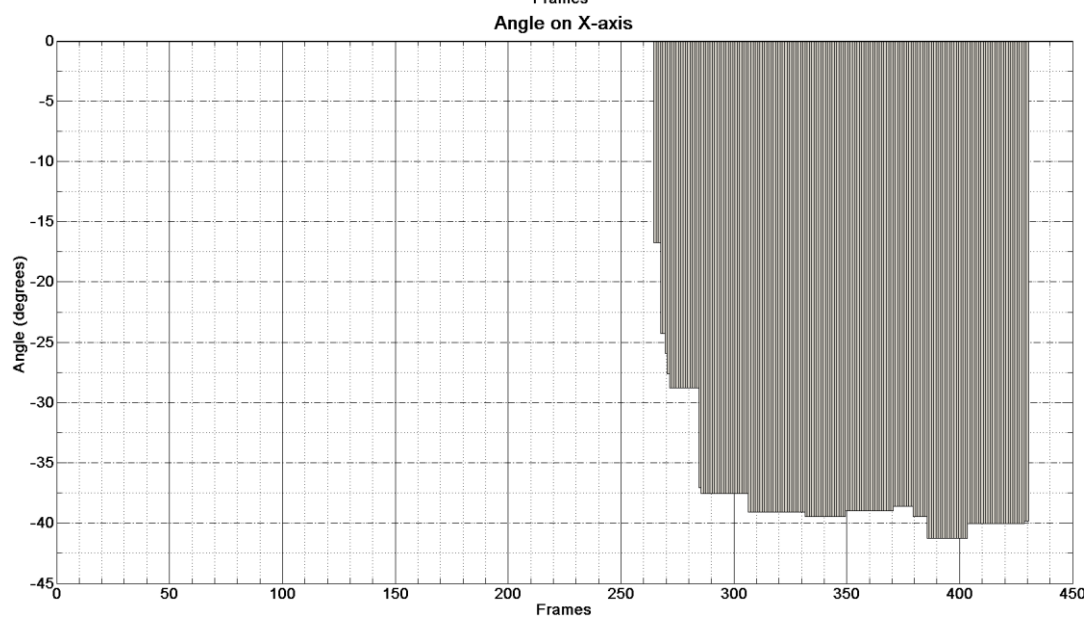
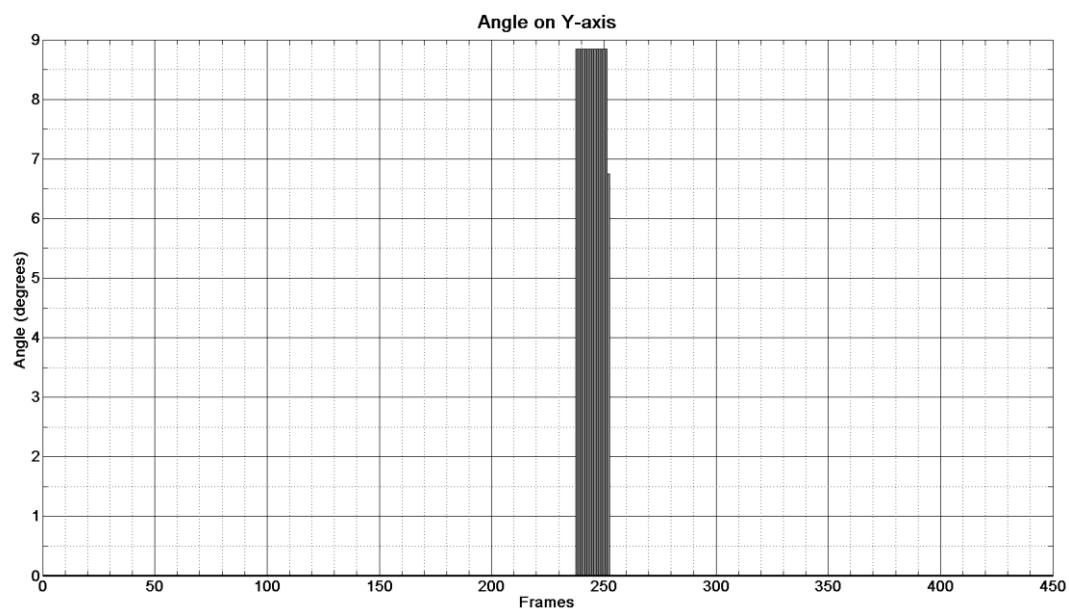
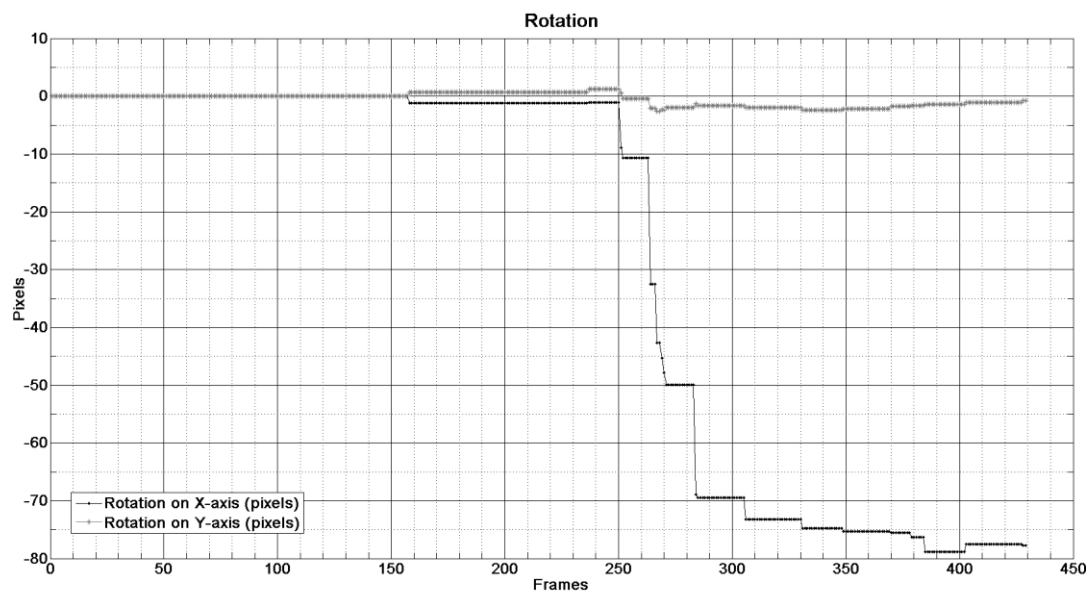


Ανάλυση	Χρωματική παλέτα	Συμπίεση	Αριθμός καρέ	Frames per sec	Ρύθμιση χρωμ. Επιπέδων (levels adjustment)	Καρε αποτυχίας ανίχνευσης (Bad Frames)
640x480	RGB24	όχι	430	30fps	×	-

Παρατηρήσεις:

Στο video10b δεν χρειάστηκε να γίνει καμία αλλαγή στις τιμές των παραμέτρων και η ρύθμιση χρωματικών επιπέδων ήταν απενεργοποιημένη. Το υποκείμενο περιστρέφει το κεφάλι του προς τα δεξιά στο 252 καρέ. Ο αλγόριθμος εκτιμά την κατεύθυνση χωρίς σφάλματα και στην μέτρηση της περιστροφής σε εικονοστοιχεία και στην μέτρηση της γωνίας στους 2 άξονες, όπως φαίνεται και στα διαγράμματα που ακολουθούν. Δεν παρατηρούνται παρά λίγες, μικρές γωνίες εντός της περιοχής ασάφειας που σημαίνει ότι τα LEDs είχαν τοποθετηθεί με ακρίβεια στο κεφάλι του υποκειμένου και είχε πάρει ορθή αρχική θέση. Ωστόσο δεν υπάρχει καλή διαβάθμιση στις εκτιμήσεις, γεγονός όμως που οφείλεται στην κακή λειτουργίας της

κάμερας στο συγκεκριμένο βίντεο, με πολλά διαδοχικά παγώματα της εικόνας και προβληματική καταγραφή της κίνησης του κεφαλιού, και όχι στην εσφαλμένη λειτουργία του προγράμματος. Η μέγιστη γωνία περιστροφής είναι $41,3^\circ$ δεξιά στον οριζόντιο άξονα και $8,8^\circ$ κάτω στον κάθετο άξονα.



Κεφάλαιο 5. Συμπεράσματα & μελλοντικές επεκτάσεις

5.1 Συμπεράσματα

Στην συγκεκριμένη διπλωματική εργασία ασχοληθήκαμε με την εκτίμηση της στάσης του κεφαλιού του ανθρώπου με την βοήθεια πρότυπης διάταξης τριών φωτοδιόδων τοποθετημένη στο κεφάλι του ανθρώπου. Η βασική ιδέα της εργασίας ήταν η δημιουργία μίας φυσικής βάσης προσώπων, με βίντεο από ανθρώπους σε έναν κοινό χώρο εργασίας που κινούνται αυτόβουλα μπροστά στον φακό μίας απλής usb web κάμερας του εμπορίου, καθώς και η ανάπτυξη ενός προγράμματος που εκμεταλλεύεται την ύπαρξη των φωτοδιόδων για την ανίχνευση της θέσης του κεφαλιού και την εξαγωγή μίας ενδεικτικής εκτίμησης της κίνησης του. Το σύστημα που χρησιμοποιήθηκε για την καταγραφή είναι αποκλειστικά μονοκάμερο χωρίς καμία πληροφορία για τις παραμέτρους της κάμερας ή του περιβάλλοντα χώρου, κάνοντας την βάση μας ιδανική για αλγόριθμους που δεν χρησιμοποιούν τέτοιου είδους πληροφορίες και δεν απαιτούν εξειδικευμένο εξοπλισμό από θέμα υλικού. Η ιδιαιτερότητα της συγκεκριμένης βάσης έγκειται στο γεγονός ότι δεν έχουν καταγραφεί δεδομένα που εκμεταλλεύονται άλλα ερευνητικά προγράμματα για την εκτίμηση της πόζας, όπως είναι η ακριβής απόσταση από την κάμερα και η εστιακή απόσταση του φακού. Ένα ακόμη στοιχείο που κάνει πιο απαιτητική της βάση μας, όσον αφορά το αλγοριθμικό κομμάτι, είναι το πολύπλοκο φόντο με ανθρώπινη δραστηριότητα σε πολλά από τα βίντεο, μιας και πολλοί ανιχνευτές/ιχνηλάτες μπορεί να «παραπλανηθούν» από την ανθρώπινη παρουσία. Επίσης είναι η πρώτη φορά που παρουσιάζεται μία βάση προσώπων που συνδυάζει παράλληλα κίνηση του κεφαλιού και αλλαγή της κατεύθυνσης του βλέμματος σε ένα κοινό πλαίσιο, όπως αυτή έχει καταγραφεί στην δεύτερη συνεδρία της.

Ο αλγόριθμος υπολογισμού της γωνίας περιστροφής που υλοποιείται με την συνάρτηση `angle.m`, δεν είναι ο βέλτιστος δυνατός και δεν ενδείκνυται η χρήση του σε γενικά προβλήματα. Παρόλα αυτά, για τις απαιτήσεις της συγκεκριμένης βάσης, είναι ιδιαίτερα γρήγορος και αρκετά αξιόπιστος για την εξαγωγή των δεδομένων αναφοράς, οπότε έχει προτιμηθεί σε σχέση με άλλους πιο πολύπλοκους αλγόριθμους.

Είναι σημαντικό να διευκρινίσουμε ότι η προσήμανση της γωνίας περιστροφής δεν δηλώνει την στιγμιαία φορά περιστροφής, αλλά την σχετική θέση του κεφαλιού σε σχέση με την αρχική μετωπική πόζα. Συνεπώς το πρόγραμμα έχει ως προαπαιτούμενο το υποκείμενο να ξεκινάει την καταγραφή του βίντεο πάντα από την μετωπική στάση.

Ο αλγόριθμος δεν έχει σχεδιαστεί ώστε να ανιχνεύει την κίνηση στην τρίτη διάσταση (άξονας Z). Δηλαδή, αν το υποκείμενο πλησιάσει ή απομακρυνθεί από τον φακό της κάμερας, το πρόγραμμα θα ιχνηλατήσει κανονικά τις φωτοδιόδους και θα εκτιμήσει σωστά την φορά

της κίνησης αλλά θα υπολογίσει με μειωμένη ακρίβεια την γωνία περιστροφής. Αυτό συμβαίνει επειδή το πρόγραμμα έχει σχεδιαστεί ώστε να εφαρμόζεται αποκλειστικά στην συγκεκριμένη βάση και τα υποκείμενα κατά την διαδικασία καταγραφής της έχουν ελευθερία κίνησης μόνο στον οριζόντιο Χ και στον κατακόρυφο Υ άξονα (2 βαθμοί ελευθερίας). Οπότε δεν χρειάστηκε να ασχοληθούμε με κίνηση που το υποκείμενο προσεγγίζει ή απομακρύνεται από την κάμερα, για την εξαγωγή των δεδομένων αναφοράς.

Στην βάση ο φωτισμός παραμένει αμετάβλητος, οπότε το πρόγραμμα δεν έχει δοκιμαστεί σε συνθήκες μεταβλητού φωτισμού. Επειδή γίνεται χρήση φωτοδιόδων για την ανίχνευση της στάσης του κεφαλιού, το πρόγραμμα λειτουργεί βέλτιστα σε συνθήκες χαμηλού φωτισμού. Παρόλα αυτά, λειτουργεί απροβλημάτιστα σε συνθήκες έντονου εσωτερικού φωτισμού γραφείου, όπως αυτός των βίντεο της βάσης, και χωρίς να επηρεάζεται από κινούμενα αντικείμενα στο φόντο.

Επίσης να επισημάνουμε ότι στην βάση δεν γίνεται χρήση αδιαφανών αντικειμένων που μπορεί να αλλοιώνουν τα χαρακτηριστικά του προσώπου, όπως γυαλιά ηλίου, κασκόλ ή σκούφος. Ωστόσο το πρόγραμμα δεν αντιμετωπίζει κανένα πρόβλημα στην ανίχνευση της πόζας του κεφαλιού ακόμα και στην ακραία περίπτωση που ο άνθρωπος έχει πλήρως καλυμμένο το πρόσωπο του. Το ίδιο συμβαίνει και στην περίπτωση που ο άνθρωπος κάνει μορφασμούς, μιλάει, γελάει και γενικά αλλάζει με κάποιον τρόπο τα χαρακτηριστικά του προσώπου του. Ο ιχνηλάτης του προγράμματος μένει ανεπηρέαστος από τέτοιου είδους μεταβολές επειδή δεν αναλύει κάποια χαρακτηριστικά του προσώπου για να ανιχνεύσει την στάση του κεφαλιού, αλλά επικεντρώνεται μόνο στις τρεις φωτοδιόδους της διάταξης. Συνεπώς, όσο οι φωτοδιόδοι είναι ευδιάκριτες από την κάμερα η πόζα του κεφαλιού μπορεί να εκτιμηθεί απροβλημάτιστα.

Γενικά η χρήση της διάταξης φωτοδιόδων και η ανίχνευσή τους από τον ιχνηλάτη του προγράμματος παρουσιάζει πλεονεκτήματα σε σχέση με άλλους ανιχνευτές που εκμεταλλεύονται διαφορετικά χαρακτηριστικά της εικόνας, τα κυριότερα εκ των οποίων είναι:

- Ο ιχνηλάτης δεν αποσπάται από την ανθρώπινη κίνηση και δραστηριότητα στον περιβάλλοντα χώρο και μένει επικεντρωμένος στο πρόσωπο του υποκειμένου.
- Δεν επηρεάζεται η λειτουργία του από αντικείμενα που αποκρύπτουν χαρακτηριστικά του προσώπου, όπως μάσκες, γυαλιά, καπέλα κασκόλ.
- Δεν επηρεάζεται από εκφράσεις του προσώπου, καθιστώντας το ιδανικό για εφαρμογές στις οποίες ο χρήστης θέλει ελευθερία κινήσεων και εκφράσεων.
- Ανιχνεύει την πόζα του κεφαλιού ακόμα και σε συνθήκες απόλυτου σκότους. Για χρήση σε τέτοιες συνθήκες η αντικατάσταση των φωτοδιόδων πράσινου χρώματος με υπέρυθρες εξαλείφει πλήρως οποιαδήποτε πηγή φωτός ορατή από το μάτι, επιτρέποντας την ιχνηλάτηση της διάταξης χωρίς να τυφλώνεται ο χρήστης.

- Πολλοί ιχνηλάτες αντιμετωπίζουν προβλήματα όταν ο φωτισμός του προσώπου δεν είναι επαρκής. Με την χρήση των φωτοδιόδων ο ιχνηλάτης μας λειτουργεί απρόσκοπα και μάλιστα βέλτιστα σε συνθήκες χαμηλού φωτισμού, αφού η αντίθεση των φωτοδιόδων με το φόντο γίνεται μέγιστη. Το ίδιο συμβαίνει και σε περιπτώσεις που η ποιότητα του βίντεο είναι χαμηλή εξαιτίας θορύβου, τεχνουργημάτων από την συμπίεση και χρωματικών αλλοιώσεων.

5.2 Μελλοντικές επεκτάσεις

Η συγκεκριμένη απλή διαρρύθμιση που επιλέχτηκε συναντάται σε εκατομμύρια χρήστες ηλεκτρονικών υπολογιστών ανά την υφήλιο που έχουν στην διάθεση τους μίας web cam συνδεδεμένη στον υπολογιστή τους για τις καθημερινές τους ανάγκες για επικοινωνία με οπτικοακουστικό υλικό μέσω του διαδικτύου. Με τον τρόπο αυτό οποιοσδήποτε αλγόριθμος αναπτυχθεί και λειτουργήσει αξιόπιστα με τα βίντεο της βάσης μπορεί κάλλιστα να ενσωματωθεί σε εφαρμογές που απευθύνονται σε απλούς χρήστες, χωρίς σημαντικές απαιτήσεις σε ειδικό εξοπλισμό και διαρρύθμιση χώρου.

Αξιοποιώντας την βάση HPEG είναι δυνατή η ανάπτυξη αλγορίθμων που συνδυάζουν την εκτίμηση της στάσης του κεφαλιού με την κατευθυντικότητα του βλέμματος σε μονοκάμερα συστήματα και μη βαθμονομημένες συνθήκες. Κατά την διάρκεια εκπόνησης της παρούσας διπλωματικής εργασίας η βιβλιογραφία που υπήρχε πάνω στο συγκεκριμένο θέμα ήταν ιδιαίτερα φτωχή και περιοριζόταν μόνο σε βάσεις προσώπων με στατικές εικόνες και όχι με ακολουθίες καρέ όπως η παρούσα.

Ο αλγόριθμος που αναπτύχθηκε για την συγκεκριμένη βάση περιορίζεται στην ανάλυση της κίνησης σε δύο βαθμούς ελευθερίας. Μια ενδιαφέρουσα επέκταση είναι η εξέλιξη του αλγορίθμου ώστε να γίνεται εκτίμηση της κίνησης σε τρεις βαθμούς ελευθερίας, επιτρέποντας το υποκείμενο να απομακρύνεται ή να πλησιάζει τον φακό της κάμερας, κάτι που είναι δυνατό με την παρακολούθηση των αλλαγών της απόστασης ανάμεσα στα μάτια ή των δύο πάνω φωτοδιόδων.

Τέλος, οι εφαρμογές των συμπερασμάτων και της έρευνας πάνω στην ταυτόχρονη παρακολούθηση της κατεύθυνσης του κεφαλιού και του βλέμματος είναι ποικίλες και με άμεση εφαρμογή στην καθημερινότητα του ανθρώπου. Η ενσωμάτωση τέτοιου είδους τεχνολογιών σε ηλεκτρονικούς υπολογιστές διευκολύνει τον χειρισμό τους από άτομα με αναπηρίες ή με ειδικές ανάγκες. Δίνει την δυνατότητα χειρισμού εφαρμογών και αξιοποίησης ειδικών χαρακτηριστικών τους χωρίς την χρήση ποντικιού, όπως σε εφαρμογές εικονικής πραγματικότητας και σε τρισδιάστατα παιχνίδια πρώτου προσώπου, δίνοντας τους ακόμα πιο ρεαλιστικό χαρακτήρα. Μπορεί να αξιοποιηθούν στον τομέα της ρομποτικής και της όρασης υπολογιστών

σε έξυπνες διαδραστικές διατάξεις που αλληλεπιδρούν με τον άνθρωπο, όπως και σε εφαρμογές e-learning και human-agents. Ήδη σε όλο και περισσότερα αυτοκίνητα γίνεται χρήση καμερών από τις αυτοκινητοβιομηχανίες για λόγους ασφαλείας και διευκόλυνσης του οδηγού. Συστήματα που παρακολουθούν σε πραγματικό χρόνο το που έχει στραμμένη την προσοχή του ο οδηγός, σε συνδυασμό με πληροφορίες από άλλα συστήματα ασφαλείας του οχήματος μπορεί να ενημερώσουν έγκαιρα τον οδηγό για επικίνδυνες καταστάσεις, συνδράμοντας σημαντικά στην πρόληψη και μείωση των τροχαίων ατυχημάτων.

Κεφάλαιο 6. Βιβλιογραφία

- [1] V.F.Ferrario, C.Sforza, G.Serrao, G.Grassi, and E.Mossi, “Active range of motion of the head and cervical spine: a three-dimensional investigation in healthy young adults,” *J.Orthopaedic Research*, vol.20, no.1, pp.122-129, 2002.
- [2] S.Langton, H.Honeyman, and E.Tessler, “The influence of head contour and nose angle on the perception of eye-gaze direction” *Perception and Psychophysics*, vol.66, no.5, pp.752-771, 2004.
- [3] W. H. Wollaston, “On the apparent direction of eyes in a portrait,” *Phil. Trans. Royal Society of London*, vol. 114, pp.247-256, 1824.
- [4] H. Kobayasi and S. Kohshima, “Unique morphology of the human eye,” *Nature*, vol. 387, no. 6635, pp. 767-768, 1997.
- [5] S. Langton and V. Bruce, “You must see the point: Automatic processing of cues to the direction of social attention,” *J. Experimental Psychology: Human Perception and Performance*, vol. 26, no. 2, pp. 747-757, 2000.
- [6] W. H. Wollaston, “On the apparent direction of eyes in a portrait,” *Phil. Trans. Royal Society of London*, vol. 114, pp.247-256, 1824.
- [7] E. Murphy-Chutorian and M.M. Trivedi, “Head pose estimation in computer vision: a survey,” in *IEEE transactions on pattern analysis and machine intelligence*, vol. 31, 2009, pp. 607-626.
- [8] N. Gourier, D. Hall, and J. Crowley, “Estimating face orientation from robust detection of salient facial structures,” in *Proc. Pointing 2004 Workshop: Visual Observation of Deictic Gestures*, 2004, pp. 17-25.
- [9] J. Tu, Y. Fu, Y. Hu, and T. Huang, “Evaluation of head pose estimation for studio data,” in *Multimodal Technologies for Perception of Humans, Int’l. Workshop Classification of Events Activities and Relationships, CLEAR 2006*, ser. Lecture Notes in Computer Science, R. Stiefel-hagen and J. Garofolo, Eds., vol. 4122, 2007, pp. 281-290.
- [10] R. Rae and H. Ritter, “Recognition of human head orientation based on artificial neural networks,” *IEEE Trans. Neural Networks*, vol. 9, no. 2, pp. 257-265, 1998.
- [11] —, “Multi-view face detection and pose estimation using a composite support vector machine across the view sphere,” in *Proc. IEEE Int’l. Workshop Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems*, 1999, pp. 14-21.
- [12] L.-P. Morency, A. Rahimi, and T. Darrell, “Adaptive view-based appearance models,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2003, pp. 803-810.
- [13] —, “Head pose estimation for driver assistance systems: A robust algorithm and ex-

- perimental evaluation,” in *Proc. IEEE Conf. Intelligent Transportation Systems*, 2007, pp. 709–714.
- [14] Y. Zhu and K. Fujimura, “3D head pose estimation with optical flow and depth constraints,” in *Proc. IEEE Int’l. Conf. 3-D Digital Imaging and Modeling*, 2003, pp. 211–216.
 - [15] U. Weidenbacher, G. Layher, P. Strauss, and H. Neumann, “A comprehensive head pose and gaze database,” in *3rd IET International Conference on Intelligent Environments (IE 07)*, 2007, pp. 455–458.
 - [16] K. Messer, J. Matas, J. Kittler, J. Luetten, and G. Maitre, “XM2VTSDB: The extended M2VTS database,” in *Second International Conference on Audio and Video-based Biometric Person Authentication*, Springer Verlag, New York: Citeseer, 1999, p. 965–966.
 - [17] A. Georgiades, P. Belhumeur, and D. Kriegman, “From few to many: generative models for recognition under variablepose and illumination,” in *Fourth IEEE International Conference on Automatic Face and Gesture Recognition, 2000. Proceedings*, 2000, p. 277–284.
 - [18] A.M. Martinez and R. Benavente, “The AR Face Database,” CVC Technical Report,no. 24, June 1998.
 - [19] P. Peer, “CVL Face Database,” 1999 [Online]. Available: <http://www.lrv.fri.uni-lj.si/facedb.html> [Accessed: Nov. 14, 2008]
 - [20] V. Jain and A. Mukherjee, “The Indian Face Database,” 2002 [Online]. Available: <http://vis-www.cs.umass.edu/~vidit/IndianFaceDatabase/> [Accessed: Nov. 15, 2008]
 - [21] N. Gourier, D. Hall, and J. Crowley, “Estimating face orientation from robust detection of salient facial structures,” in *FG Net Workshop on Visual Observation of Deictic Gestures (POINTING)*, 2004, pp. 17–25.
 - [22] S. Asteriadis, D. Soufleros, K. Karpouzis, S. Kollias, “A Natural Head Pose and Eye Gaze Dataset,” in *International Conference on Multimodal Interfaces (ICMI 2009)*, November 2–6, Boston, MA. Available: <http://manolito.image.ece.ntua.gr/papers/611.pdf>