



ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΛΟΠΟΝΝΗΣΟΥ
ΣΧΟΛΗ ΜΗΧΑΝΙΚΩΝ
ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

**"Ανίχνευση κίνησης σε ζώνες ή μέτωπα
περιορισμένης πρόσβασης με χρήση του Matlab και
χρήση video analytics σε στοιχεία καμερών"**

**"Motion detection in restricted access zones or fronts
using Matlab and using video analytics on camera
elements"**

ΦΟΙΤΗΤΗΣ: ΤΕΤΡΑΔΗΣ ΕΥΣΤΑΘΙΟΣ ΕΥΑΓΓΕΛΟΣ (1101823)

ΕΠΙΒΛΕΠΩΝ: ΑΣΗΜΑΚΟΠΟΥΛΟΣ ΓΕΩΡΓΙΟΣ

ΠΑΤΡΑ 2022

ΠΕΡΙΕΧΟΜΕΝΑ

ΠΕΡΙΛΗΨΗ.....	3
ABSTRACT.....	4
ΚΕΦΑΛΑΙΟ 1: ΕΙΣΑΓΩΓΗ.....	5
ΚΕΦΑΛΑΙΟ 2: ΚΑΜΕΡΕΣ.....	9
2.1 Είδη κάμερας.....	9
2.2 Κάμερες και νυχτερινή λήψη.....	11
2.3 Πλεονεκτήματα - Μειονεκτήματα.....	11
ΚΕΦΑΛΑΙΟ 3: ΤΕΧΝΙΚΕΣ ΑΝΙΧΝΕΥΣΗΣ ΑΝΘΡΩΠΟΥ.....	15
3.1 Εισαγωγή.....	15
3.2 Κατάτμηση κίνησης με τεχνικές αφαίρεσης φόντου.....	15
3.3 Ανίχνευση προσώπου.....	21
3.4 Ανίχνευση κίνησης.....	26
3.5 Ανίχνευση δέρματος.....	29
3.6 Εντοπισμός χαρακτηριστικών σημείων.....	40
ΚΕΦΑΛΑΙΟ 4: ΕΝΤΟΠΙΣΜΟΣ ΚΙΝΟΥΜΕΝΗΣ ΟΝΤΟΤΗΤΑΣ.....	49
4.1 Εισαγωγή.....	49
4.2 Point tracking – Deterministic method.....	51
4.3 Point tracking – Statistical method.....	53
4.4 Kernel tracking.....	54
4.5 Silhouette tracking.....	54
4.6 Φίλτρο Kalman.....	55
ΚΕΦΑΛΑΙΟ 5: ΥΛΟΠΟΙΗΣΗ.....	59
5.1 Εισαγωγή στο Matlab.....	59
5.2 Περιγραφή αλγορίθμου.....	67
5.3 Παρουσίαση της εφαρμογής.....	70
ΣΥΜΠΕΡΑΣΜΑΤΑ.....	95
ΒΙΒΛΙΟΓΡΑΦΙΑ.....	96

ΠΕΡΙΛΗΨΗ

Στην παρούσα πτυχιακή εργασία αρχικά μελετώνται οι κάμερες και οι τεχνικές ανίχνευσης ανθρώπινης κίνησης χρησιμοποιώντας κάμερες. Επίσης γίνεται αναφορά στις εφαρμογές αναγνώρισης ανθρώπου, οι οποίες θα υλοποιηθούν μέσω τεχνικών.

Στόχος της εργασίας είναι η κατασκευή, με τη βοήθεια του Matlab, μιας εφαρμογής για συστήματα ασφαλείας που θα ανιχνεύει την κίνηση που προέρχεται από ανθρώπινη παρουσία και θα επιστρέφει ένα μήνυμα στο χρήστη προκειμένου να του γνωστοποιεί αν εντόπισε άνθρωπο σε ένα βίντεο.

Ο χρήστης θα εισάγει ένα βίντεο και αυτό στη συνέχεια θα αναλύεται προκειμένου να εξακριβωθεί αν υπάρχει ανθρώπινη παρουσία στη σκηνή. Οι τεχνικές που χρησιμοποιούνται για τη δημιουργία της εφαρμογής είναι η αφαίρεση φόντου και η ανίχνευση δέρματος.

ABSTRACT

In this thesis, cameras and human motion detection techniques using cameras are initially studied. Reference is also made to human recognition applications, which will be implemented through techniques.

The aim of the work is to build, with the help of Matlab, an application for security systems that will detect the movement originating from human presence and will return a message to the user in order to notify him if he has detected a human in a video.

The user will input a video and it will then be analyzed to determine if there is a human presence in the scene. The techniques used to create the app are background subtraction and skin detection.

ΚΕΦΑΛΑΙΟ 1: ΕΙΣΑΓΩΓΗ

Η αναγνώριση και παρατήρηση ενός κινούμενου αντικειμένου αποτελεί ένα σημαντικό κεφάλαιο στο πεδίο της υπολογιστικής όρασης. Η αύξηση της υπολογιστικής ισχύος των μηχανών και η διάθεση ολοένα και υψηλότερης ποιότητας και χαμηλού κόστους συσκευών λήψης, σε συνδυασμό με την αυξανόμενη ανάγκη για αυτοματοποιημένη ανάλυση βίντεο έχει δημιουργήσει μεγάλο ενδιαφέρον στην επιστημονική κοινότητα για την ανάπτυξη αλγορίθμων που θα προσομοιώνουν την ανθρώπινη αντίληψη αναγνώρισης της κίνησης. Η προσομοίωση αυτή βέβαια δεν είναι ακόμα ώριμη να γίνει σε ποιοτικά χαρακτηριστικά και σε καμία περίπτωση δεν είναι σε θέση να προσεγγίσει την πολυπλοκότητα των ανθρώπινων νευρώνων στην αντιμετώπιση μιας υπόθεσης απλής αναγνώρισης και παρακολούθησης κίνησης. Παρόλα αυτά γίνεται μεγάλη προσπάθεια προσέγγισης του προβλήματος κάνοντας χρήση άλλων χαρακτηριστικών.

Παρόλο που οι μηχανές έχουν σημαντικά πλεονεκτήματα όσο αναφορά την ταχύτητα της επεξεργασίας πολύπλοκων δεδομένων και την αποθήκευση και γρήγορη ανάκτηση μεγάλου όγκου πληροφοριών αδυνατούν να προσεγγίσουν τα ποιοτικά χαρακτηριστικά που ακολουθεί ο ανθρώπινος εγκέφαλος προκειμένου να επιτελέσει την διαδικασία της αναγνώρισης και παρακολούθησης τροχιάς ενός αντικειμένου ακόμα και κάτω από περιπτώσεις κακής ορατότητας ή πολλών δευτερογενών αντικειμένων που σταδιακά κρύβουν το αντικείμενο που είναι προς παρακολούθηση. Η ανθρώπινη αντίληψη σε τέτοιες περιπτώσεις ενεργοποιείται αυτόματα, χωρίς κάποια προσπάθεια, και η περίπτωση λάθους είναι πολύ μικρή. Η αντίληψη αυτή προέρχεται μέσω ερεθισμάτων που λαμβάνουμε από τις αισθήσεις μας και στην συγκεκριμένη περίπτωση μέσω του αισθητήριου οργάνου της όρασης, δηλαδή του ματιού. Στην περίπτωση των υπολογιστικών συστημάτων η αίσθηση της όρασης εξομοιώνεται καταγράφοντας ακολουθίες εικόνων μέσω μιας ή περισσότερων καμερών. Στα πλαίσια αυτά λοιπόν, μια μηχανή συνδεδεμένη με μια απλή κάμερα είναι σε θέση να "βλέπει". Όπως όμως αναφέραμε και παραπάνω η μεγάλη διαφορά έγκειται στο γεγονός ότι ο ανθρώπινος οργανισμός είναι σε θέση να αντιλαμβάνεται άμεσα αυτά που βλέπει, ενώ το υπολογιστικό σύστημα θα πρέπει να εκτελέσει μια σειρά από σύνθετες λειτουργίες προκειμένου να φτάσει σε χειρότερα ή στην καλύτερη περίπτωση στα ίδια συμπεράσματα.

Για να πάρει, δηλαδή, μια απόφαση θα πρέπει να εξάγει κάποια χαρακτηριστικά από τις εικόνες που δέχεται, βάσει των οποίων θα μπορέσει να διακρίνει, να αναγνωρίσει και να χαρακτηρίσει αυτά τα οποία βλέπει. Στη προσπάθεια λοιπόν ένα σύστημα να κατανοήσει αυτά που βλέπει μέσω της μηχανικής όρασης, δέχεται σε πρώτη φάση τα δεδομένα εισόδου, τα οποία πριν από το στάδιο της επεξεργασίας τους δεν έχουν καμία πρακτική χρησιμότητα στην αντίληψη της μηχανής. Δεν είναι παρά οι συντεταγμένες των εικονοστοιχείων των εικόνων στο χώρο των χρωμάτων. Με βάση μόνο τη πληροφορία αυτή, το σύστημα καλείται να επιτελέσει υψηλού επιπέδου εργασίες που σε κάποιο βαθμό να φανερώνουν την αντίληψή του. Η μεθοδολογία που ακολουθείται στη λύση τέτοιων προβλημάτων (ώστε να επιτελεστούν οι εργασίες αυτές στη γενική περίπτωση) εξαρτάται από ένα σύνολο παραμέτρων που έχουν να κάνουν με τα επιθυμητά αποτελέσματα εξόδου ως προς τη φύση τους, την πολυπλοκότητά τους, και την ακρίβεια με την οποία θέλουμε να λαμβάνονται και με ποικίλους άλλους τρόπους.

Έτσι εντελώς διαφορετική προσέγγιση θα υιοθετήσουμε για παράδειγμα όταν θέλουμε να μετρήσουμε το ποσοστό ενός χρώματος σε μία εικόνα, διαφορετική όταν θέλουμε να ανιχνεύσουμε κινούμενες οντότητες σε μια ακολουθία εικόνων και διαφορετική όταν θέλουμε να υπολογίσουμε την σχετική απόσταση μεταξύ δύο στόχων. Στη πράξη, ένα σύστημα λέμε ότι αντιλαμβάνεται τα περιεχόμενα μιας τέτοιας πληροφορίας όταν είναι σε θέση να μετασχηματίσει τη χρωματική πληροφορία των ψηφιακών εικόνων σε κάποια άλλη μορφή πληροφορίας με περισσότερο σημασιολογικό περιεχόμενο. Για παράδειγμα, ο χωρισμός μιας εικόνας σε χρωματικές περιοχές, η εξαγωγή των ακμών της και η εκτίμηση της κίνησης των εικονοστοιχείων από καρέ σε καρέ, αποτελούν κάποια αρχικά στοιχεία που μπορεί κάποιος να εξάγει και που περιέχουν πληροφορία χρήσιμη για τα επόμενα στάδια επεξεργασίας. Τα στοιχεία αυτά ονομάζονται γνωρίσματα (features).

Γενικά, οι μέθοδοι εξαγωγής γνωρισμάτων χρησιμοποιούν παρεμφερείς μεθόδους. Αρχικά, ένα ή περισσότερα γνωρίσματα χαμηλού επιπέδου εξάγονται από τις εικόνες, όπως το χρώμα, η υφή, το σχήμα, η κίνηση κλπ. με χρήση κάποιας εκ των προτέρων γνώσης, γενικής ή ειδικότερης ανάλογα με την εφαρμογή και στη συνέχεια εξάγεται πληροφορία υψηλότερου επιπέδου υπολογίζοντας νέα χαρακτηριστικά με μεγαλύτερη σημασιολογική πληροφορία. Τα χαρακτηριστικά αυτά μπορούν με τη σειρά τους να χρησιμοποιηθούν για να χαρακτηρίσουν μια εικόνα, μια ακολουθία βίντεο ή ακόμη και αντικείμενα των εικόνων, όπως την απόσταση ενός στόχου από

τη κάμερα, το τρισδιάστατο μοντέλο ενός αντικειμένου ή τον προσδιορισμό της φύσης μιας κινούμενης οντότητας. Κάποια από τα πιο βασικά γνωρίσματα χαμηλού επιπέδου που επεξεργάζονται κατά κόρον σε εφαρμογές του ενδιαφέροντος μας είναι τα εξής:

- Χρώμα

Αποτελεί ίσως την πιο άμεσα εξαγόμενη πληροφορία μιας εικόνας. Όπως τα υπόλοιπα γνωρίσματα, έτσι και το χρώμα μπορεί να μοντελοποιηθεί με διάφορους τρόπους. Στη γενικότερη περίπτωση, προσπαθούμε η μοντελοποίηση που θα κάνουμε να περιέχει όσο το δυνατόν περισσότερη πληροφορία, ανάλογα με την εφαρμογή. Έτσι το χρώμα μοντελοποιείται συνήθως από έναν μικρό αριθμό βασικών χρωμάτων (RGB) ή μέσω του χρωματικού ιστογράμματος.

- Υφή

Όμοια με το χρώμα αντιμετωπίζεται και η υφή μιας εικόνας. Διάφορα μοντέλα έχουν προταθεί για τη μοντελοποίηση της καθώς είναι ένα μέγεθος που συνδέεται άμεσα με την εγκυρότητα των υπολογισμών μας, ειδικά σε περιπτώσεις εκτίμησης κίνησης.

- Σχήμα

Ενώ μπορεί γενικά να θεωρηθεί ως γνώρισμα υψηλότερου επιπέδου, μοντελοποιείται και αυτό σε χαμηλό επίπεδο με βάση τις ιδιότητες της εικόνας όπως για παράδειγμα τις ακμές, το εμβαδό, τη θέση κλπ. Το περίγραμμα αντικειμένων ενδιαφέροντος μιας εικόνας εξάγεται και μοντελοποιείται σε υψηλότερο επίπεδο.

- Κίνηση

Αποτελεί πολύ σημαντικό γνώρισμα για τις περισσότερες εφαρμογές του κλάδου και εξάγεται από δύο ή περισσότερα καρέ μιας ακολουθίας βίντεο. Μέχρι και σήμερα, έχει αναπτυχθεί ένας μεγάλος αριθμός μεθόδων για τον υπολογισμό της κίνησης καθώς η εκτίμηση της αποτελεί βάση πολλών εφαρμογών, όπως για παράδειγμα ο υπολογισμός της τρισδιάστατης κίνησης ενός αντικειμένου και η παρακολούθηση τροχιάς.

- Πρότυπα

Μια συνηθισμένη τακτική σε κάποιες από τις κατηγορίες εφαρμογών είναι η εξαγωγή προτύπων αντί χαρακτηριστικών από ακολουθίες εικόνων τα οποία στη συνέχεια χρησιμοποιούνται αυτούσια, μετασχηματισμένα ή και παραμορφωμένα σε διαδικασίες εντοπισμού και αναγνώρισης. Ως παράδειγμα αναφέρουμε την εξαγωγή προτύπων από εικόνες με ανθρώπινα πρόσωπα.

- Μετασχηματισμοί

Ενδιαφέροντα γνωρίσματα για εφαρμογές μηχανικής όρασης προκύπτουν και από το μετασχηματισμό των εικόνων ή συγκεκριμένων περιοχών τους. Για παράδειγμα ο μετασχηματισμός Fourier δίνει πληροφορίες για το συχνοτικό περιεχόμενο των εικόνων.

Σε πάρα πολλές εφαρμογές, όπως για παράδειγμα τον εντοπισμό εισβολέων σε κατοικία ή τη παρακολούθηση της τροχιάς κινούμενων αντικειμένων χρησιμοποιείται ως πρώτο στάδιο ο εντοπισμός κύριων κινούμενων αντικειμένων. Σε τέτοιες περιπτώσεις χρησιμοποιούνται συνήθως γνωρίσματα κίνησης, σχήματος ή και συνδυασμός των δυο (π.χ. κίνηση περιοχών ή ακμών). Όταν είναι επιθυμητή η λεπτομερής εξαγωγή της μορφής κινούμενων αντικειμένων, χαρακτηριστικά κίνησης συνδυάζονται συνήθως με γνωρίσματα χρώματος για πιο έγκυρα αποτελέσματα.

Σε εφαρμογές που στόχο έχουν την αναγνώριση και ταξινόμηση αντικειμένων τα γνωρίσματα σχήματος είναι περισσότερο κατάλληλα και ειδικότερα τα υψηλού επιπέδου χαρακτηριστικά όπως το περίγραμμα. Στη περίπτωση που μελετάται ο διαχωρισμός δύσκαμπτων και εύκαμπτων αντικειμένων, τα χαρακτηριστικά κίνησης είναι επίσης πολύ χρήσιμα και καταλήγουν σε σαφείς διαχωρισμούς με μικρά επίπεδα σφάλματος. Βεβαίως, όταν τα υπό αναγνώριση αντικείμενα έχουν ιδιότητες που τα διαφοροποιούν στο πεδίο του χρώματος, της συχνότητας ή της υφής δεν αποκλείεται και η χρήση των γνωρισμάτων αυτών, αν και η περιοχή εφαρμογών που καλύπτουν είναι στη πραγματικότητα αμελητέα.

ΚΕΦΑΛΑΙΟ 2: ΚΑΜΕΡΕΣ

Η βιντεοκάμερα είναι μια φωτογραφική μηχανή που χρησιμοποιείται για την απόκτηση εικόνας κίνησης.

Αρχικά, αναπτύχθηκε από την τηλεοπτική βιομηχανία αλλά τώρα είναι πολύ διαδεδομένη και σε άλλες εφαρμογές. Οι πρώτες βιντεοκάμερες είναι αυτές του John Logie Baird, βασισμένες στον ηλεκτρομηχανικό δίσκο Nipkow και χρησιμοποιήθηκαν από το BBC, σε πειραματικές εκπομπές από το 1930.

Όλα τα ηλεκτρονικά σχέδια που είχαν ως βάση τον καθοδικό σωλήνα, όπως το εικονοσκόπιο του Vladimir Zworykin και τον image dissector του Philo T. Farnsworth, αντικατέστησαν την τεχνολογία του Baird το 1940 και χρησιμοποιήθηκαν ευρέως μέχρι τα τέλη του 1980, όταν οι κάμερες με αισθητήρες εικόνας στερεάς κατάστασης όπως CCD και αργότερα οι CMOS με ενεργούς αισθητήρες εικόνας εξάλειψαν τα κοινά προβλήματα της τεχνολογίας σωλήνα, όπως το "κάψιμο" της εικόνας και έκανε την ψηφιακή ροή εργασίας πιο πρακτική.

Οι βιντεοκάμερες χρησιμοποιούνται κυρίως με δυο τρόπους. Ο πρώτος, είναι στη ζωντανή τηλεόραση, όπου η κάμερα τροφοδοτεί σε πραγματικό χρόνο εικόνες απευθείας σε μια οθόνη για άμεση παρατήρηση. Λίγες κάμερες συνεχίζουν να εξυπηρετούν προϊόντα ζωντανής τηλεόρασης αφού οι περισσότερες ζωντανές συνδέσεις είναι για λόγους ασφαλείας, για στρατιωτικές και εμπορικές επιχειρήσεις στις οποίες απαιτείται ζωντανή προβολή. Στο δεύτερο τρόπο, οι εικόνες αποθηκεύονται σε μια αποθηκευτική μονάδα για περαιτέρω επεξεργασία, για πολλά χρόνια η βιντεοκασέτα ήταν η κυρίαρχη μορφή για αυτό το σκοπό, αλλά τα οπτικά μέσα αποθήκευσης την αντικατέστησαν. Καταγεγραμμένα βίντεο χρησιμοποιούνται στην τηλεόραση και στην παρακολούθηση και επιτήρηση εργασιών όπου η χωρίς παρακολούθηση καταγραφή μιας κατάστασης απαιτείται για μετέπειτα ανάλυση.

2.1 Είδη κάμερας

Οι κάμερες ασφαλείας μπορούν να διαχωριστούν με βάση τον τύπο τους σε τέσσερις κατηγορίες.

Αρχικά έχουμε τις κάμερες dome (θόλου) οι οποίες πήραν το όνομα τους διότι βρίσκονται εντός προστατευτικού θόλου με αποτέλεσμα να είναι ανθεκτικότερες σε προσπάθεια βανδαλισμού τους. Επιπλέον ο θόλος βοηθά έτσι ώστε να μην είναι εμφανής η κατεύθυνση στην οποία είναι στραμμένη η κάμερα.

Στη δεύτερη κατηγορία έχουμε τις κάμερες box οι οποίες έχουν σχήμα κουτιού και είναι οι πλέον συνήθεις σε μορφή, κάμερες. Έχουν τη δυνατότητα να τοποθετηθούν σε οποιαδήποτε κατακόρυφη επιφάνεια με τη βοήθεια κατάλληλων υποστηριγμάτων. Επίσης, μπορούν να εξοπλιστούν με μεγάλους φακούς όταν είναι αναγκαία παρακολούθηση απομακρυσμένων σημείων.

Στη συνέχεια έχουμε τις bullet κάμερες, οι οποίες πήραν το όνομά τους από το κυλινδρικό τους σχήμα. Είναι λιγότερο ογκώδεις από τις box κάμερες, αλλά λόγω του μεγέθους τους έχουν περιορισμούς στους φακούς που μπορούν να λάβουν και γι αυτό είναι κατάλληλες για παρακολουθήσεις μικρών και μεσαίων αποστάσεων με την ευκρίνεια της εικόνας που παράγουν να είναι συνήθως υποδεέστερη από αυτήν των box καμερών.

Στην τελευταία κατηγορία συναντάμε τις κρυφές κάμερες οι οποίες είναι κατασκευασμένες με τέτοιο τρόπο, έτσι ώστε να μην προδίδουν ότι είναι κάμερες παρακολούθησης. Λόγω του μικρού τους μεγέθους οι κάμερες αυτές έχουν χαμηλή ευκρίνεια και είναι αποτελεσματικές μόνο σε παρακολουθήσεις μικρών αποστάσεων. Σε μεγάλο βαθμό οι κάμερες ασφαλείας λειτουργούν σε συνθήκες χωρίς φωτισμό. Αυτό καθίσταται δυνατό μέσω των εξής τεχνολογιών:

1. image intensification
2. active illumination
3. thermal



Εικόνα 2.1: Κάμερες Ασφαλείας.

2.2 Κάμερες και νυχτερινή λήψη

Η νυχτερινή λήψη είναι η τεχνολογία που μας παρέχει το θαύμα της παρακολούθησης σε απόλυτο σκοτάδι και της βελτίωσης της εικόνας σε περιβάλλον με χαμηλό φωτισμό. Η νυχτερινή λήψη μπορεί να χωριστεί σε τρεις μεγάλες κατηγορίες.

Αρχικά έχουμε την image intensification τεχνολογία η οποία λειτουργεί με βάση την αρχή της μεγιστοποίησης της ποσότητας των ληφθέντων φωτονίων από διάφορες φυσικές πηγές όπως το φώς των αστεριών και του φεγγαριού. Παραδείγματα αυτής της τεχνολογίας είναι τα γυαλιά νυχτερινής όρασης και διάφορα είδη από low light κάμερες.

Η δεύτερη κατηγορία είναι η active illumination (ενεργού φωτισμού) η οποία λειτουργεί με βάση την αρχή της εντατικοποίησης της ζεύξης, στην τεχνολογία απεικόνισης με μια ενεργή πηγή φωτισμού στο φάσμα της εγγύς υπέρυθρης ή της βραχέως υπέρυθρης ακτινοβολίας (NIR και SWIR αντίστοιχα). Παραδείγματα αυτής της τεχνολογίας περιλαμβάνουν τις low light κάμερες.

Τέλος έχουμε την thermal απεικόνιση. Σε αυτή την κατηγορία οι κάμερες λειτουργούν ανιχνεύοντας την διαφορά θερμοκρασίας των αντικειμένων που βρίσκονται στο παρασκήνιο σε σχέση με αυτά που βρίσκονται στο προσκήνιο.

2.3 Πλεονεκτήματα - Μειονεκτήματα

Η low light απεικόνιση είναι η πιο γνωστή μέθοδος για να πραγματοποιήσουμε νυχτερινή λήψη είναι με τη χρήση ενισχυτών εικόνας.

Αυτή η μέθοδος της νυχτερινής όρασης ενισχύει το διαθέσιμο φώς για να επιτευχθεί καλύτερη όραση εστιάζοντας το διαθέσιμο φώς για να αυξήσει την ποιότητα της εικόνας. Χρησιμοποιούμε πράσινο χρώμα σε αυτές τις εφαρμογές επειδή το ανθρώπινο μάτι μπορεί να διαφοροποιήσει περισσότερες σκιές πράσινου από κάθε άλλο χρώμα, επιτρέποντας μεγαλύτερη διαφοροποίηση των αντικειμένων στο πλαίσιο μας.



Εικόνα 2.2: Παράδειγμα low light απεικόνισης.

Πλεονεκτήματα:

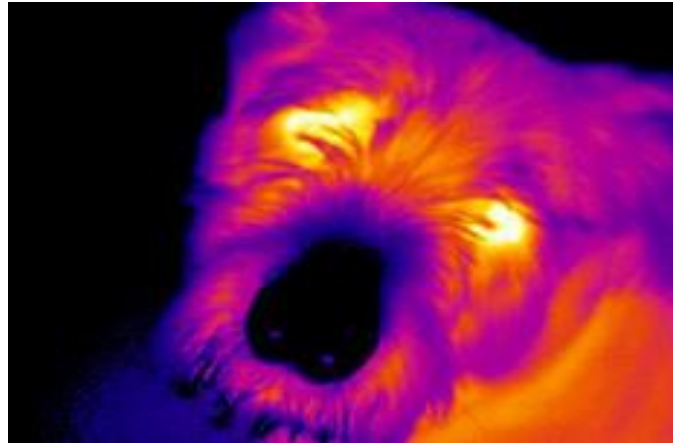
1. Εξαιρετική ευαισθησία σε λήψεις με χαμηλό φωτισμό.
2. Υψηλή ανάλυση.
3. Χαμηλό κόστος.
4. Δυνατότητα εντοπισμού ανθρώπων.
5. Ενισχυμένη οπτική απεικόνιση με τη καλύτερη δυνατή επίδοση σε αναγνώριση και ταυτοποίηση.

Μειονεκτήματα:

1. Επειδή βασίζονται σε μεθόδους ενίσχυσης, απαιτείται κάποια ποσότητα από φώς άρα η μέθοδος δεν είναι χρήσιμη σε χώρους με απόλυτο σκοτάδι.
2. Έχουν κατώτερη επίδοση σε φωτεινές λήψεις από κάμερες που χρησιμοποιούνται αποκλειστικά σε φωτεινές λήψεις.
3. Έχουν μεγάλη πιθανότητα βλάβης σε περίπτωση που υπάρξει φωτεινό αντικείμενο σε συνθήκες χαμηλού φωτισμού.

Οι θερμικές κάμερες διαφέρουν από τις lowlight καθώς δεν χρειάζονται καθόλου φώς για να λειτουργήσουν.

Οι θερμικές κάμερες λειτουργούν με βάση την αρχή ότι όλα τα αντικείμενα εκπέμπουν υπέρυθη ακτινοβολία ως συνάρτηση της θερμότητάς τους. Γενικότερα, όσο πιο ζεστό είναι ένα αντικείμενο τόσο μεγαλύτερη ακτινοβολία εκπέμπει.



Εικόνα 2.3: Παράδειγμα θερμικής κάμερας.

Πλεονεκτήματα:

1. Υψηλή αντίθεση σε σενάρια χωρίς φώς.
2. Ανιχνεύει πολύ εύκολα ανθρώπους και οχήματα.
3. Δεν επηρεάζεται από φωτεινές πηγές.

Μειονεκτήματα:

1. Είναι πολύ ακριβό στην αγορά και στο χειρισμό.
2. Δεν έχει καλή απόδοση σε περιπτώσεις που το αντικείμενο κινείται με μεγάλη ταχύτητα.

Οι near infrared απεικόνιση η οποία είναι μια δημοφιλής και αρκετές φορές οικονομική μέθοδος για να πραγματοποιήσουμε νυχτερινή λήψη. Σε αυτή τη μέθοδο, χρησιμοποιείται μια συσκευή η οποία είναι ευαίσθητη στην αόρατη υπέρυθρη ακτινοβολία, σε συνδυασμό με μια διάταξη φωτισμού υπέρυθρων. Με αυτό τον τρόπο, αυξάνουμε την ποιότητα εικόνας σε σκηνές με χαμηλό φωτισμό, παράγοντας εικόνες ικανοποιητικής ποιότητας.



Εικόνα 2.4: Παράδειγμα κάμερας near infrared illumination.

Πλεονεκτήματα:

1. Έχει το χαμηλότερο κόστος σε σύγκριση με τις άλλες μεθόδους νυχτερινής λήψης.
2. Εξαλείφει τις σκιές και εντοπίζει ανθρώπους και αντικείμενα.
3. Μπορεί να πραγματοποιήσει λήψεις όταν το αντικείμενο κινείται με μεγάλη ταχύτητα.
4. Είναι ικανό να διαπεράσει έντονη βροχή, ομίχλη, καπνούς.

Μειονεκτήματα:

1. Οι κάμερες μπορούν να ανιχνευτούν εύκολα από τρίτους με συσκευές εντοπισμού υπερέθρων.

ΚΕΦΑΛΑΙΟ 3: ΤΕΧΝΙΚΕΣ ΑΝΙΧΝΕΥΣΗΣ ΑΝΘΡΩΠΟΥ

3.1 Εισαγωγή

Σχεδόν κάθε σύστημα που έχει ως στόχο την ανίχνευση κίνησης ξεκινά με την ανίχνευση ή αλλιώς κατάτμηση της κίνησης. Αυτή η ανίχνευση αποσκοπεί στον διαχωρισμό των περιοχών που περιέχουν κινούμενες οντότητες από την υπόλοιπη εικόνα. Είναι ένα πολύ σημαντικό βήμα για να προχωρήσει κάποιος σε περαιτέρω ανάλυση της ανθρώπινης δραστηριότητας καθώς οι διαδικασίες που ακολουθούν εξαρτώνται σε μεγάλο βαθμό από τα αποτελέσματα του.

Πρόκειται λοιπόν για ένα πολύ σημαντικό και απαιτητικό πρόβλημα, με στόχο την ανίχνευση των περιοχών εκείνων που αντιστοιχούν σε κινούμενα αντικείμενα, όπως οχήματα, άνθρωποι κ.α.. Ο εντοπισμός κινούμενων οντοτήτων επικεντρώνει ουσιαστικά τη προσοχή μας στις περιοχές που μας ενδιαφέρουν για τα επόμενα στάδια επεξεργασίας, όπως η παρακολούθηση της πορείας των αντικειμένων και η ανάλυση των δραστηριοτήτων τους. Παρόλα αυτά, αλλαγές λόγω καιρού, φωτισμού, σκιών και επαναλαμβανόμενων κινήσεων από ανεπιθύμητα αντικείμενα δυσκολεύουν την γρήγορη και αξιόπιστη επεξεργασία των δεδομένων.

Μέχρι σήμερα, οι δημοφιλέστερες μέθοδοι κατάτμησης χρησιμοποιούν είτε χρονική είτε χωρική πληροφορία που εξάγουν από τα διαδοχικά καρέ. Οι μέθοδοι που μπορούν να χρησιμοποιηθούν είναι πολλές κάθε μία εκ των οποίων έχει διαφορετικά πλεονεκτήματα και μειονεκτήματα. Γενικά όσο πιο ελαφριά είναι μία μέθοδος από άποψη υπολογιστικού κόστους τόσο χειρότερα είναι τα εξαγόμενα αποτελέσματα ενώ αντίθετα όσο περισσότερες πράξεις γίνονται και όσο πιο πολλές παράμετροι εξετάζονται τόσο πιο λεπτομερή είναι τα εξαγόμενα αποτελέσματα.

3.2 Κατάτμηση κίνησης με τεχνικές αφαίρεσης φόντου

Η τεχνική αυτή είναι μία πολύ δημοφιλής μέθοδος για τον εντοπισμό των κινούμενων περιοχών, ειδικά κάτω από συνθήκες σχετικά σταθερού φόντου. Λειτουργεί υπολογίζοντας τις διαφορές ανάμεσα στη τρέχουσα εικόνα και το φόντο αναφοράς για κάθε εικονοστοιχείο ξεχωριστά. Το πρόβλημα της είναι ότι παρουσιάζει πολύ μεγάλη ευαισθησία σε δυναμικές αλλαγές του σκηνικού λόγω φωτισμού και άλλων εξωγενών παραγόντων.

Υπάρχουν πολλές διαφορετικές προσεγγίσεις αυτού του προβλήματος οι οποίες διαφοροποιούνται στο τύπο του μοντέλου που χρησιμοποιούν για το φόντο καθώς και στον τρόπο με τον οποίο το ανανεώνουν. Η τρέχουσα εικόνα συγκρίνεται με ένα μοντέλο του φόντου το οποίο δεν περιέχει τα κινούμενα αντικείμενα. Συνήθως, αυτό το μοντέλο του φόντου ανακτάται με την πάροδο του χρόνου. Το πλεονέκτημα τέτοιων προσαρμόσιμων μεθόδων είναι ότι αφαιρούν το φόντο από τις εικόνες. Συνήθως με τον όρο αφαίρεση φόντου εννοούμε την απομάκρυνση των στατικών αντικειμένων έτσι ώστε να παραμείνει μόνο η περιοχή ενδιαφέροντος μας που είναι τα κινούμενα άτομα σε μια ακολουθία εικόνων.

Η βασική ιδέα είναι ο υπολογισμός της απόλυτης διαφοράς των εικόνων δύο διαδοχικών στιγμών, Δηλαδή έστω ότι έχουμε N καρέ για να βρούμε τα άτομα που κινούνται μεταξύ των καρέ $N-1$ και N , δεν έχουμε παρά να υπολογίσουμε την απόλυτη διαφορά τους και να προκύψει σαν στατικό (μαύρο) οτιδήποτε είναι ακίνητο ενώ οτιδήποτε άλλο θα είναι κινούμενο αντικείμενο. Μην ξεχνάμε ότι οι υπολογισμοί αυτοί γίνονται σε εικόνες και στην ουσία αυτό που υπολογίζουμε είναι η διαφορά των τιμών των εικονοστοιχείων σε κάθε καρέ, οπότε με βάση αυτά που είπαμε παραπάνω οτιδήποτε ακίνητο θα μας επιστρέψει την τιμή μηδέν μιας και οι τιμές του εικονοστοιχείου αυτού δεν άλλαξαν κατά τις δύο χρονικές στιγμές. Αυτό σημαίνει ότι τα αντικείμενα που ήταν στην σκηνή κατά την εξαγωγή του φόντου βαθμιαία θα αντικατασταθούν από το φόντο. Αυτό εξαλείφει την ανάγκη της αρχικοποίησης ενός άδειου φόντου. Επίσης, παρέχει ένα μοντέλο φόντου που προσαρμόζεται αυτόματα στις αλλαγές. Για παράδειγμα, όταν αλλάζει ο καιρός ή όταν ένα παρκαρισμένο αυτοκίνητο φεύγει από τη σκηνή.

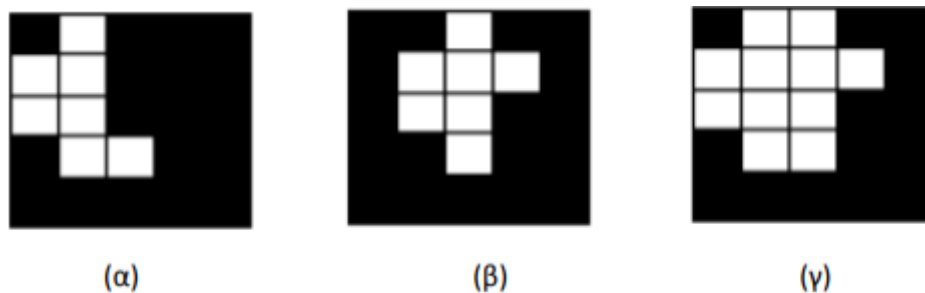
Ένα σημαντικό μειονέκτημα είναι η αλληλεξάρτηση (trade-off) μεταξύ δύο αντιτιθέμενων απαιτήσεων. Από τη μία, η ενημέρωση του φόντου θα πρέπει να γίνεται γρήγορα ώστε να μπορεί να αντιμετωπίσει τις αλλαγές φωτισμού και τις αλλαγές στο φόντο, όπως αλλαγές του φωτισμού λόγω αλλαγής της ώρας ή λόγω ανοίγματος του διακόπτη σε κλειστό χώρο, αλλαγές λόγω ανθρώπων που περπατούν ή λόγω αντικειμένων που φεύγουν από τη σκηνή. Από την άλλη, η ενημέρωση θα πρέπει να γίνεται αργά, ώστε τα αντικείμενα που κινούνται αργά να μην περιλαμβάνονται στο φόντο, όπως ένας άνθρωπος που κοιμάται. Η ταχύτητα λοιπόν, της ενημέρωσης εξαρτάται από την εφαρμογή.

Αφαίρεση φόντου μέσω τεχνικής διαδοχικών καρτέ

Αποτελεί τον πιο απλό τρόπο για τον διαχωρισμό του φόντου από το κινούμενο αντικείμενο σε ασπρόμαυρο βίντεο με στατική κάμερα είναι να συγκρίνουμε ανά δύο τα καρτέ του βίντεο παίρνοντας τη διαφορά τους. Αυτό γίνεται ως εξής:

1. Ορίζουμε έναν πίνακα με διαστάσεις ίδιες με αυτές των καρτέ τον οποίο αρχικοποιούμε.
2. Για κάθε δύο διαδοχικά καρτέ παίρνουμε κάθε εικονοστοιχείο και βρίσκουμε τη διαφορά της τιμής του μεταξύ των δύο καρτέ.
3. Ορίζουμε ένα κατώφλι (threshold) ($0 < \tau < 1$)
4. Αν η διαφορά των τιμών του εικονοστοιχείου είναι μικρότερη της τιμής του κατωφλίου, τότε βάζουμε στην αντίστοιχη θέση του πίνακα την τιμή 1, ενώ αν η διαφορά είναι μεγαλύτερη του κατωφλίου, τότε στην αντίστοιχη θέση βάζουμε την τιμή 0. Έτσι, από κάθε δύο διαδοχικά καρτέ παίρνουμε μία ασπρόμαυρη εικόνα όπου τα μαύρα εικονοστοιχεία αντιστοιχούν στα εικονοστοιχεία που δεν μετακινήθηκαν, δηλαδή στο φόντο, ενώ τα άσπρα εικονοστοιχεία αντιστοιχούν στα εικονοστοιχεία που άλλαξαν, δηλαδή στο κινούμενο αντικείμενο.

Στο παρακάτω σχήμα φαίνεται ένα παράδειγμα, όπως αυτό περιγράφηκε παραπάνω με κάποια υποτιθέμενα διαδοχικά frames, ώστε να οπτικοποιηθεί η περιγραφή.



Εικόνα 3.1: Τρία διαδοχικά frames α , β , γ όπου τα μαύρα εικονοστοιχεία αντιστοιχούν στα εικονοστοιχεία που δεν μετακινήθηκαν, δηλαδή στο φόντο, ενώ τα άσπρα εικονοστοιχεία αντιστοιχούν στα εικονοστοιχεία που άλλαξαν, δηλαδή στο κινούμενο αντικείμενο.

Με αυτόν τον τρόπο όμως, προκύπτουν διάφορα προβλήματα. Καταρχήν, χρειάζεται μεγάλη προσοχή στην τιμή που θα βάλουμε στο κατώφλι. Αν η τιμή είναι πολύ μικρή, τότε η παραμικρή κίνηση θα ανιχνευθεί, πράγμα που μπορεί να οδηγήσει σε εσφαλμένες εκτιμήσεις, διότι μπορεί να έχουμε μικρές αλλαγές στο φόντο, οι οποίες δε θέλουμε να θεωρηθούν ως κινούμενα αντικείμενα. Τέτοιες αλλαγές μπορεί να

οφείλονται σε θόρυβο λόγω της κάμερας, σε αλλαγές φωτισμού, σε σκιές, σε αέρα αν είμαστε σε εξωτερικό περιβάλλον, κ.ά. Απ' την άλλη, αν η τιμή του καταφλίου είναι πολύ μεγάλη, τότε κάποιες αλλαγές μπορεί να μην ανιχνευθούν καθόλου, όπως για παράδειγμα ένα αντικείμενο που κινείται με πολύ μικρή ταχύτητα.

Η απλότητα της μεθόδου έχει δύο πολύ μεγάλα μειονεκτήματα, πρώτον ότι για αντικείμενα με ομοιόμορφη κατανομή της τιμής της πυκνότητας τους τα εσωτερικά εικονοστοιχεία τους ερμηνεύονται ως μέρη του φόντου και δεύτερον ότι τα αντικείμενα πρέπει να κινούνται συνεχώς ειδήλλως στην περίπτωση που το αντικείμενο παραμένει ακίνητο για χρονικό διάστημα ενός καρέ ερμηνεύεται και πάλι σαν μέρος του φόντου.

Επίσης, μειονέκτημα της μεθόδου αποτελεί το γεγονός ότι το εσωτερικό των κινούμενων αντικειμένων αναγνωρίζεται από την μέθοδο ως φόντο και αφαιρείται και αυτό διότι μεταξύ δύο διαδοχικών καρέ δεν έχει προλάβει το αντικείμενο να απομακρυνθεί αρκετά έτσι ώστε στην ίδια περιοχή, τις δύο διαφορετικές χρονικές στιγμές, να υπάρχουν εικονοστοιχεία με ίδιο χρωματικό περιεχόμενο. Ένα καλό φυσικά που έχει η μέθοδος είναι ο περιορισμένος θόρυβος του φόντου.

Μορφολογικού μετασχηματισμού – Μετά επεξεργασία

Η τεχνική των μορφολογικών μετασχηματισμών δεν αποτελεί ουσιαστικά διαφορετική μέθοδο και μπορεί να χρησιμοποιηθεί για ενσωμάτωση σε κάθε μία από τις προηγούμενες μεθόδους βελτιώνοντας με αυτό τον τρόπο τα αποτελέσματα που παίρνουμε.

Μπορεί να αποτελέσει λοιπόν ένα πολύ σημαντικό κομμάτι στη κατάτμηση κίνησης που βελτιώνει θεαματικά τα αρχικά αποτελέσματα όλων των παραπάνω τεχνικών. Πιο συγκεκριμένα, σε κάθε καρέ, αφού υπολογίσουμε την δυαδική εικόνα πρώτου πλάνου, που έχει τιμή μονάδα για όλα τα κινούμενα εικονοστοιχεία σύμφωνα με την εκατοστέ μέθοδο, επιδρούμε πάνω της με μια σειρά μετασχηματισμών ώστε αυτή να γίνει πιο ομοιόμορφη. Οι μετασχηματισμοί αυτοί έχουν στόχο να σβήσουν από την εικόνα όσες ομάδες εικονοστοιχείων είναι πολύ μικρές για να παριστάνουν κάποιο αντικείμενο ενδιαφέροντος και παράλληλα γεμίζουν τυχόν τρύπες στις κινούμενες οντότητες που προέρχονται από κομμάτια ομοιόμορφης τιμής της έντασης. Επιπλέον εκτός από την εξάλειψη των μικρών περιοχών, οι οποίες αποτελούν θόρυβο, μπορούμε να χρησιμοποιήσουμε μορφολογικούς μετασχηματισμούς για εξάλειψη

και μεγάλων περιοχών που δεν είναι πιθανό να αντιπροσωπεύουν κάποια ανθρώπινη φιγούρα αλλά μπορεί να είναι κάποιο άλλο κινούμενο αντικείμενο όπως για παράδειγμα κάποιο αυτοκίνητο.

Η μείωση των περιοχών θορύβου στηρίζεται στην παρατήρηση ότι οι περιοχές θορύβου τείνουν να είναι μικρότερες από την περιοχή του αντικειμένου. Πρώτα, εφαρμόζεται ο κλασικός αλγόριθμος των συνδεδεμένων συστατικών (connected component algorithm) στην αρχική μάσκα αντικειμένου για να σημειωθούν οι απομονωμένες περιοχές. Στη συνέχεια, υπολογίζεται η έκταση κάθε περιοχής. Περιοχές με έκταση μικρότερη από κάποια τιμή κατωφλίου εξαλείφονται από την μάσκα αντικειμένου. Αντίστοιχα περιοχές μεγαλύτερες από μία προκαθορισμένη τιμή μπορούν να εξαλειφθούν επίσης. Έτσι, διατηρείται το σχήμα του αντικειμένου, ενώ μικρές περιοχές θορύβου ή μεγάλες περιοχές που αντιπροσωπεύουν κάποιο άλλο κινούμενο αντικείμενο αφαιρούνται. Εφόσον υπάρχουν δύο ήδη θορύβου, θόρυβος στο φόντο και θόρυβος στο προσκήνιο, χρειάζονται δύο περάσματα σε αυτό το στάδιο.

Το πρώτο πέρασμα αφαιρεί μικρές μαύρες περιοχές (περιοχές φόντου), οι οποίες είναι περιοχές θορύβου στο προσκήνιο ή 'τρύπες' στη μάσκα ανίχνευσης αλλαγών. Το δεύτερο πέρασμα αφαιρεί μικρές άσπρες περιοχές (περιοχές στο προσκήνιο), οι οποίες είναι περιοχές θορύβου στο φόντο ή λανθασμένες περιοχές αλλαγών στη μάσκα ανίχνευσης αλλαγών. Μετά την αφαίρεση των περιοχών θορύβου, εφαρμόζονται στη μάσκα αντικειμένου οι μορφολογικές πράξεις (morphological operations) ανοίγματος και κλεισίματος (open and close operations) με ένα 3x3 δομικό στοιχείο (structural element). Το μικρό δομικό στοιχείο επιλέγεται για τη λείανση του περιγράμματος του αντικειμένου, χωρίς να επηρεάζει τις λεπτομέρειες της πληροφορίας του σχήματος του αντικειμένου. Οι μορφολογικές πράξεις είναι η βάση για την απομάκρυνση του θορύβου μετά τον διαχωρισμό του φόντου και την κατωφλιοποίηση.

Πιο συγκεκριμένα, οι μορφολογικές τεχνικές που συνήθως χρησιμοποιούνται, αποτελούνται από δύο βασικές πράξεις: την διεύρυνση (dilation) και τη διάβρωση (erosion). Η διεύρυνση επεκτείνει το προσκήνιο της εικόνας, προσθέτοντας ένα εικονοστοιχείο στο προσκήνιο αν οποιοδήποτε από τα γειτονικά του εικονοστοιχείου εντός μιας συγκεκριμένης 'γειτονιάς' ακτίνας r (η οποία λέγεται δομικό στοιχείο - structuring element) ανήκει ήδη στο προσκήνιο. Η διάβρωση επεκτείνει το φόντο, μετακινώντας ένα εικονοστοιχείο από το προσκήνιο αν ένα από τα γειτονικά του

εικονοστοιχείου ανήκει ήδη στο φόντο. Αυτές οι δύο πράξεις μπορούν να συνδυαστούν. Μία διεύρυνση ακολουθούμενη από μία πανομοιότυπη διάβρωση λέγεται ‘κλείσιμο’ (closing), και γεμίζει τις τρύπες του προσκηνίου που είναι μικρότερες από την διάμετρο της γειτονιάς. Παρομοίως, μία διάβρωση ακολουθούμενη από μία πανομοιότυπη διεύρυνση λέγεται ‘άνοιγμα’ (opening), και χρησιμοποιείται για την εξάλειψη απομονωμένων εικονοστοιχείων του προσκηνίου. Ο θόρυβος στην εικόνα που προκύπτει από την αφαίρεση του φόντου, κάνει ορισμένα εικονοστοιχεία του προσκηνίου να μοιάζουν με αυτά του φόντου και αντίστροφα. Μία πράξη ‘άνοιγματος’ ακολουθούμενη από μία πράξη ‘κλεισίματος’ αντιμετωπίζει τις πηγές του λάθους: το ‘κλείσιμο’ γεμίζει τα εικονοστοιχεία του προσκηνίου που λείπουν (υποθέτοντας ότι αρκετά από τα γειτονικά εικονοστοιχεία έχουν προσδιοριστεί σωστά), και το ‘άνοιγμα’ αφαιρεί άσχετα εικονοστοιχεία του προσκηνίου που είναι περικυκλωμένα από εικονοστοιχεία του φόντου. Ιδιαίτερη προσοχή πρέπει να δοθεί στην επιλογή της ακτίνας r . Αν η ακτίνα είναι πολύ μικρή, τότε μεγαλύτερες ομάδες εικονοστοιχείων θορύβου θα μείνουν αδιόρθωτες. Αν απ’ την άλλη η ακτίνα είναι πολύ μικρή, τότε κάποια σωστά εικονοστοιχεία των αντικειμένων του προσκηνίου θα χαθούν.

Σε ιδιαίτερα θορυβώδεις εικόνες φόντου, τα λανθασμένα εικονοστοιχεία φόντου μπορεί να είναι τόσο πολλά και τόσο κοντά το ένα στο άλλο, ώστε η πράξη ‘κλεισίματος’ να γεμίσει τα κενά ανάμεσά τους. Η αύξηση του κατωφλίου για τον αρχικό διαχωρισμό φόντου-προσκηνίου οδηγεί στην αποφυγή αυτού του ανεπιθύμητου αποτελέσματος, διότι μεγαλώνοντας την τιμή του κατωφλίου τα εικονοστοιχεία του προσκηνίου που ταξινομούνται ως εικονοστοιχεία του φόντου είναι περισσότερα από τα εικονοστοιχεία του φόντου που τα ταξινομούνται ως εικονοστοιχεία του προσκηνίου. Έτσι, εφαρμόζοντας υπό αυτές τις συνθήκες την πράξη κλεισίματος θα κλείσουν τα κενά ανάμεσα σε σωστά εκτιμημένα εικονοστοιχεία του προσκηνίου.

Κατάτμηση κίνησης με γρήση στατιστικών μεθόδων

Στατιστικές μέθοδοι για να εξαχθούν οι περιοχές αλλαγών από το φόντο έχουν εμπνευστεί από τις βασικές μεθόδους αφαίρεσης φόντου που ήδη περιγράψαμε.

Οι στατιστικές αυτές προσεγγίσεις χρησιμοποιούν χαρακτηριστικά από ατομικά εικονοστοιχεία ή από ομάδες τους για να κατασκευάσουν πιο σύνθετα μοντέλα για το

φόντο. Οι στατιστικές αυτών των μοντέλων μάλιστα, αλλάζουν δυναμική κατά τη διάρκεια της επεξεργασίας. Κάθε εικονοστοιχείο στην τρέχουσα εικόνα χαρακτηρίζεται ως εικονοστοιχείο πρώτου πλάνου ή φόντου αφού συγκριθεί με τις στατιστικές του ενίοτε μοντέλου φόντου. Αυτό η προσέγγιση γίνεται όλο και πιο δημοφιλής λόγω της σθεναρότητας που παρουσιάζει στο θόρυβο, στις σκιές, στις αλλαγές των συνθηκών φωτός κλπ. Το πιο αντιπροσωπευτικό παράδειγμα της κατηγορίας ονομάζεται μίγμα γκαουσιανών (mixture of Gaussians - MOG).

Στη μέθοδο αυτή, το φόντο δεν είναι ένα καρέ με τιμές, αλλά παραμετρικό. Η κάθε θέση εικονοστοιχείο αντιπροσωπεύεται από έναν αριθμό γκαουσιανών συναρτήσεων που αθροίζονται όλες μαζί δημιουργώντας μία συσσωρευτική κατανομή πιθανότητας:

$$F(i_t = \mu) = \sum_{i=1}^k \omega_{i,t} \cdot \eta(\mu, \sigma)$$

Ο μέσος όρος της κάθε γκαουσιανής συνάρτησης μπορεί να γίνει αντιληπτός ως μια βάσιμη υπόθεση (educated guess) της τιμής του εικονοστοιχείου στην επόμενη εικόνα, θεωρώντας στο σημείο αυτό ότι τα εικονοστοιχεία ανήκουν γενικά στο φόντο. Το βάρος και η τυπική απόκλιση της κάθε συνάρτησης αποτελούν μέτρα της εμπιστοσύνης μας σε αυτή την υπόθεση όπου μεγαλύτερο βάρος και μικρότερο σ σημαίνουν μεγαλύτερη εμπιστοσύνη. Τυπικά χρησιμοποιούνται από τρεις έως πέντε γκαουσιανές κατανομές για κάθε εικονοστοιχείο, ανάλογα τους περιορισμούς μας στο μέγεθος της μνήμης του συστήματος.

Για να καθορίσουμε αν ένα εικονοστοιχείο είναι μέρος του σκηνικού το συγκρίνουμε με τις αντίστοιχες γκαουσιανές συναρτήσεις. Εάν η τιμή του βρίσκεται κοντά στη τυπική απόκλιση ενός στοιχείου του φόντου, σε επίπεδο μιας τάξης μεγέθους, θεωρείται ως μέρος του. Σε αντίθετη περίπτωση, καταχωρείται ως εικονοστοιχείο πρώτου πλάνου. Η μέθοδος αυτή είναι πολύ ισχυρή και περιέχει πολλές παραμέτρους με την αλλαγή των οποίων μπορούμε να τον προσαρμόσουμε σε οποιοδήποτε περιβάλλον. Η πολυπλοκότητα του όμως και σε σύγκριση με τα αποτελέσματα του μας κάνει να επιλέξουμε την άλλη μέθοδο, για μείωση υπολογιστικού κόστους.

3.3 Ανίχνευση προσώπου

Η αναγνώριση της παρουσίας ανθρώπου είναι μια από τις πιο σημαντικές λειτουργίες που μπορούν να εξαχθούν από βίντεο. Παρόλα αυτά, η εργασία αυτή γίνεται πιο

περίπλοκη με την παρουσία διαφόρων μεταβλητών στην εικόνα, όπως η φωτεινότητα, ο φωτισμός, η αντίθεση και το φόντο. Σκοπός του κεφαλαίου είναι η συζήτηση για την ανάπτυξη μιας μεθόδου, η οποία θα ανιχνεύει και θα προσδιορίζει την παρουσία ανθρώπου σε μια ακολουθία βίντεο, και θα την διαφοροποιεί από μη-ανθρώπινες κινήσεις.

Από την έρευνα που έγινε καταλήξαμε στο συμπέρασμα ότι ο πιο κοινός τρόπος για να επιτευχθεί ανίχνευση ανθρώπου είναι μέσω της ανίχνευσης προσώπου. Το πρόσωπο του ανθρώπου είναι το πιο ξεχωριστό σημείο στο ανθρώπινο σώμα, και αν ανιχνευθεί με ακρίβεια οδηγεί σε επιτυχή ανίχνευση ανθρώπινης σιλουέτας.

Από τη δεκαετία του 1970 έως και σήμερα έχουν γίνει αρκετές μελέτες στον τομέα της ανίχνευσης προσώπου, και έχουν δημοσιευθεί αρκετές έρευνες όσον αφορά τους αλγόριθμους που χρησιμοποιούνται σε αυτό τον τομέα. Γενικότερα οι αλγόριθμοι ανίχνευσης προσώπου μπορούν να κατηγοριοποιηθούν σε δυο κύρια είδη.

Στην πρώτη κατηγορία ανήκουν οι προσεγγίσεις με βάση τα χαρακτηριστικά(feature-based) παραδείγματος χάριν η «bottom-up» προσέγγιση. Στην αυτή την κατηγορία, εξάγονται τα χαρακτηριστικά του προσώπου από μία εικόνα και χειριζόμαστε τις παραμέτρους της, όπως γωνίες, μέγεθος και αποστάσεις. Στη δεύτερη κατηγορία ανήκουν οι προσεγγίσεις με βάση την εικόνα (image-based), για παράδειγμα προσεγγίσεις με βάση την εμφάνιση(appearancebased) αλλά και η προσέγγιση «template matching». Αυτό το είδος αλγόριθμου, βασίζεται σε τεχνικές εκπαίδευσης μηχανής για να αναγνωρίζει συγκεκριμένα σημεία ενδιαφέροντος. Μια μικρή περιγραφή των πιο κοινών προσεγγίσεων ακολουθεί στις επόμενες ενότητες.



Εικόνα 3.2: Ανίχνευση προσώπου σε εξωτερικό χώρο.

Information based (Top-Down)

Σε αυτή τη μέθοδο καταγράφονται οι σχέσεις μεταξύ των χαρακτηριστικών ενός προσώπου για να γίνει η αναπαράσταση του, και κωδικοποιούνται σαν ένα σύνολο κανόνων.

Παραδείγματος χάριν, οι άξιες έντασης για το κεντρικό μέρος του προσώπου είναι ομοιόμορφες από τον έναν άνθρωπο στον άλλο. Συνήθως, ένα πρόσωπο εμφανίζεται με συμμετρικά μεταξύ τους μάτια, μια μύτη και στόμα. Επιπροσθέτως, σε μεγάλο αριθμό αλγορίθμων αυτής της κατηγορίας χρησιμοποιείται η κλίμακα «coarse-to-fine». Σύμφωνα με την «coarse-to-fine» ξεκάνει μια αναζήτηση από την μεγαλύτερη κλίμακα (coarsest) και συνεχίζει στις επόμενες μέχρι ότου να εντοπιστεί η βέλτιστη.

Feature Invariant (Bottom-up)

Στην προσέγγιση αυτή, τα χαρακτηριστικά τα οποία μένουν αναλλοίωτα σε ένα πρόσωπο χρησιμοποιούνται για να βοηθήσουν στην ανίχνευσή του. Αυτά τα χαρακτηριστικά πρέπει να μην αλλάζουν σε ποικίλες οπτικές γωνίες, φωτισμούς και διαφορετικές συνθήκες. Οι συνήθεις αλγόριθμοι που ανήκουν σε αυτή την προσέγγιση είναι:

- Προσέγγιση με βάση το χρώμα ή μοντέλο δέρματος

Αυτή η μέθοδος αξιοποιεί το γεγονός ότι το χρώμα του δέρματος μπορεί να χρησιμοποιηθεί σαν στοιχείο παρουσίας ανθρώπου, δεδομένου του ότι μια μεγάλη ποικιλία αποχρώσεων του δέρματος από διαφορετικές φυλές μπορεί να καταταχτεί σε μια ενιαία περιοχή. Ο Vezhnevets V. παρουσίασε τέσσερις τεχνικές μοντελοποίησης με βάση τα pixel: την μη παραμετρική μοντελοποίηση διανομής του δέρματος (non-parametric skin distribution modelling), την παραμετρική (parametric skin distribution modelling), την δυναμική (dynamic skin distribution modelling) και τέλος την ρητώς οριζόμενη περιοχή του δέρματος (explicitly defined skin region).

- Προσέγγιση με βάση τα χαρακτηριστικά προσώπου

Αυτή η μέθοδος στην οποία, χρησιμοποιούνται τα παγκόσμια χαρακτηριστικά του ανθρώπινου σώματος, π.χ. δέρμα, μέγεθος, και σχήμα, αλλά και τα λεπτομερή, π.χ. μάτια, μύτη, χείλη, έγινε αρκετά δημοφιλής τα τελευταία χρόνια. Στις περισσότερες περιπτώσεις προκειμένου να επιτευχθεί η ανίχνευση προσώπου χρησιμοποιούνται τα παγκόσμια χαρακτηριστικά για να εντοπιστεί η υποψήφια περιοχή και στη συνέχεια γίνεται επαλήθευση με την χρήση των λεπτομερών χαρακτηριστικών.

- Υφής (Texture)

Το ανθρώπινο πρόσωπο διαφέρει από τα άλλα αντικείμενα σε υφή. Σε αυτή τη μέθοδο, εξετάζουμε την πιθανότητα η υποψήφια περιοχή να ανήκει σε υφή ανθρώπινου προσώπου, με τη χρήση ενός πίνακα Space Gray Level Dependence (SGLD matrix).

Template matching methods

Αυτές οι μέθοδοι βασίζονται στη μέτρηση του βαθμού ομοιότητας μεταξύ της υποψήφιας περιοχής και του αποθηκευμένου μοτίβου του προσώπου. Η προκαθορισμένη εικόνα μπορεί να είναι για όλο το μοτίβο του προσώπου ή για μοναδικά χαρακτηριστικά του, όπως είναι τα μάτια, τα χείλη και η μύτη.

Οι πιο συνήθεις αλγόριθμοι σε αυτή την προσέγγιση είναι:

- Προκαθορισμένα πρότυπα προσώπου

Σε αυτή την κατηγορία αλγορίθμου πολλά πρότυπα από ολόκληρο το πρόσωπο ή κομμάτια αυτού αποθηκεύονται με σκοπό να αναλυθούν τα μοτίβα του, για μετέπειτα χρήση τους.

- Παραμορφώσιμα πρότυπα

Σε αυτή την κατηγορία ένα ελαστικό μοντέλο των χαρακτηριστικών του προσώπου χρησιμοποιείται ως μοντέλο αναφοράς, στο οποίο το παραμορφώσιμο πρότυπο του αντικειμένου ενδιαφέροντος ταιριάζεται.

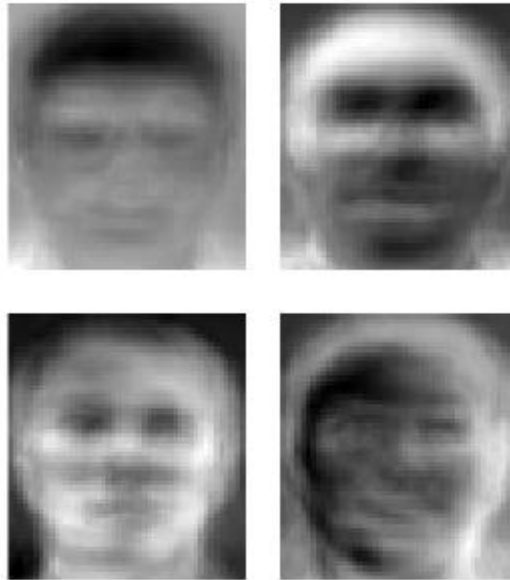
Appearance based

Σε αντίθεση με τις «template matching» μεθόδους στις οποίες τα πρότυπα έχουν προκαθοριστεί από ειδικούς, οι appearance-based μέθοδοι, μαθαίνουν τα πρότυπα από ένα σετ εικόνων, χρησιμοποιώντας στατιστική ανάλυση και εκμάθηση μηχανής. Μερικά παραδείγματα αλγορίθμων που χρησιμοποιούν αυτή τη μέθοδο είναι:

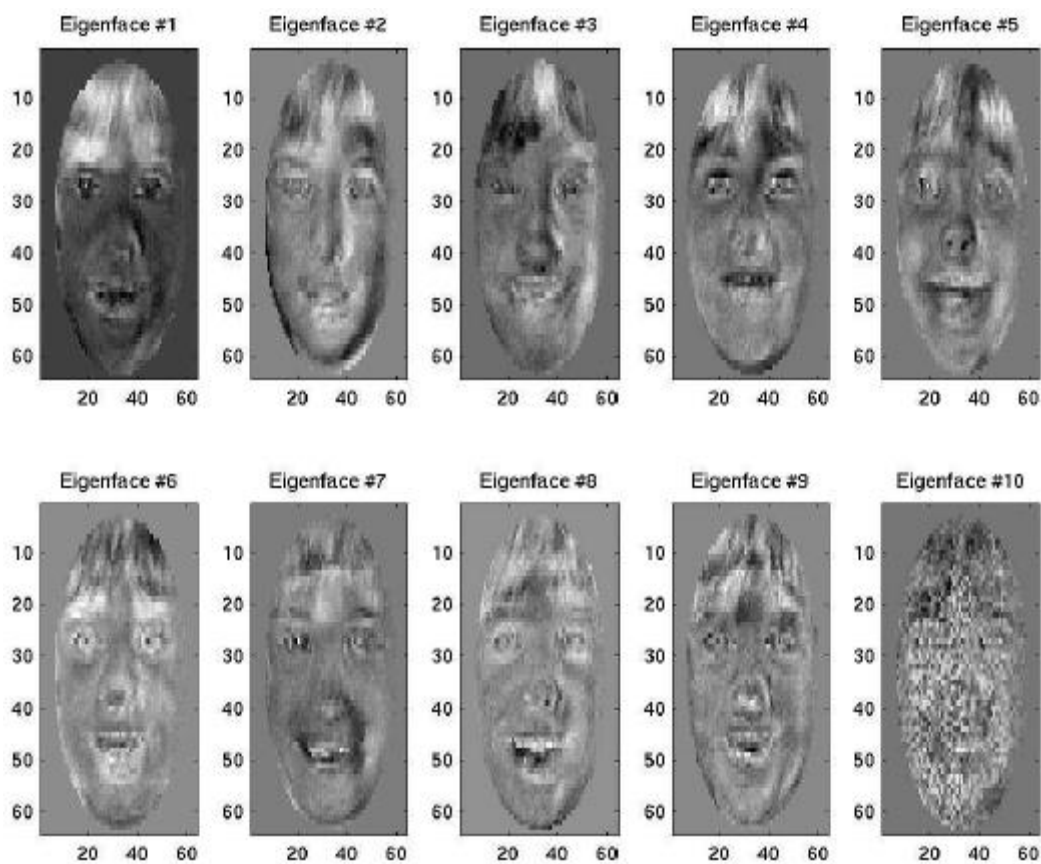
- Eigenfaces

Ο συγκεκριμένος αλγόριθμος είναι ένα σετ από ιδιοδιανύσματα που χρησιμοποιούνται στην τεχνητή όραση για να αντιμετωπίσουν το πρόβλημα της ανίχνευσης προσώπου. Αυτά τα ιδιοδιανύσματα προκύπτουν από τον πίνακα διακύμανσης της κατανομής των πιθανοτήτων για τα εκάστου ανθρώπινα πρόσωπα. Ουσιαστικά ένα σύνολο από eigenfaces μπορεί να παραχθεί από την εκτέλεση μιας διαδικασίας, που ονομάζεται μαθηματική ανάλυση κύριων συνιστωσών σε ένα μεγάλο σύνολο εικόνων που απεικονίζουν διαφορετικά ανθρώπινα πρόσωπα. Κάθε

ανθρώπινο πρόσωπο μπορεί να θεωρηθεί ότι είναι ένας συνδυασμός από τυποποιημένα πρόσωπα. Αξίζει να σημειωθεί ότι δεν χρειαζόμαστε πολλά eigenfaces για να επιτύχουμε μια σωστή προσέγγιση των περισσότερων προσώπων, επίσης, επειδή τα πρόσωπα αποθηκεύονται σαν μια απλή λίστα τιμών χρειαζόμαστε πολύ λιγότερο χώρο σε μια βάση δεδομένων για την αποθήκευσή τους.



Εικόνα 3.3: Παράδειγμα 1 από Eigenfaces.



Εικόνα 3.4: Παράδειγμα 2 από Eigenfaces.

- Distributed-based

Εδώ το καταναμημένο μοτίβο ενός αντικειμένου μαθαίνεται με τη χρήση ενός σετ θετικών και αρνητικών εικόνων του αντικειμένου.

- Neural Network

Τα νευρωνικά δίκτυα(κόμβοι) λειτουργούν για να παράγουν παράλληλες συναρτήσεις. Η ιδέα των νευρωνικών δικτύων ήρθε από το κεντρικό νευρικό σύστημα. Παρόλα αυτά, τα δίκτυα αυτά εκπαιδεύτηκαν για να ανιχνεύουν την παρουσία προσώπου αφού λαμβάνουν δείγματα από πρόσωπο.

- Support Vector Machine

Τα «Support Vector Machines» είναι εκπαιδευόμενες μηχανές οι οποίες κάνουν δυαδικούς συσχετισμούς. Η ιδέα είναι να μεγιστοποιηθεί το περιθώριο μεταξύ θετικών και αρνητικών σετ διανυσμάτων προκειμένου να επιτευχθεί ένα βέλτιστο όριο το οποίο θα διαχωρίζει τα δυο σετ διανυσμάτων.

3.4 Ανίχνευση κίνησης

Σε αντίθεση με τις ακίνητες εικόνες, οι ακολουθίες βίντεο έχουν περισσότερες πληροφορίες σχετικά με το ιστορικό των κινούμενων αντικειμένων, το οποίο βοηθά στο διαχωρισμό του φόντου από το προσκήνιο.

Η ανίχνευση κινούμενων αντικειμένων είναι πολύ σημαντική σε βίντεο παρακολούθησης. Σε πολλά περιβάλλοντα η κίνηση μπορεί να είναι είτε ενδιαφέρουσα, π.χ. άνθρωπος(salient- εμφανής), είτε μη ενδιαφέρουσα, π.χ. ταλαντευόμενα κλαδιά. Η ανίχνευση κίνησης σε βίντεο θεωρείται ένα σημαντικό και αρκετά δύσκολο πρόβλημα. Σε πολλά πραγματικά περιβάλλοντα η κίνηση προκαλείται και από ενδιαφέροντα αλλά και από μη ενδιαφέροντα ερεθίσματα.

Σαν εμφανή κίνηση θεωρούμε την κίνηση που προκύπτει από ένα τυπικό αντικείμενο παρακολούθησης, π.χ. άνθρωπο ή όχημα, σε αντίθεση με τις άλλες παραπλανητικές κινήσεις, όπως είναι οι ακτίνες του φωτός όταν πέφτουν στο νερό ή το ανέμισμα των φύλων ενός δέντρου. Οι παραπλανητικές αυτές κινήσεις, σε ένα πραγματικό περιβάλλον, κάνουν το πρόβλημα της ανίχνευσης κίνησης ακόμη πιο δύσκολο.

Γενικότερα, οι κινούμενες περιοχές ανιχνεύονται με τον εντοπισμό των αλλαγών που δημιουργούνται στην ακολουθία εικόνων. Οι περισσότερες εφαρμογές που έχουν δημιουργηθεί πάνω στην ανίχνευση κίνησης, εφαρμόζουν ορισμένα στάδια προεπεξεργασίας πριν εφαρμόσουν τον αλγόριθμο ανίχνευσης αλλαγών. Αυτά τα

στάδια περιέχουν γεωμετρικές προσαρμογές αλλά και προσαρμογές έντασης. Το πρόβλημα της ποικιλίας στην ένταση του φωτός, λύνεται με την προσαρμογή της. Σε αυτήν την προσαρμογή το αποτέλεσμα του φωτισμού μειώνεται ανάλογα με τη μέθοδο που θα χρησιμοποιήσουμε. Ο Elgammal σε έρευνα του δηλώνει ότι αν μετασχηματίσουμε τις τιμές RGB σε «chromatic colour space», το μοντέλο μας μένει ανεπηρέαστο από μικρές αλλαγές στο φως.

Υπάρχουν αρκετοί τρόποι για να εντοπίσεις μια αλλαγή σε ένα βίντεο. Πρόσφατες μελέτες έδειξαν ότι η μέθοδος διαφοροποίησης εικόνας (image differencing method) είναι πιο αποτελεσματική από τις υπόλοιπες στην ανίχνευση αλλαγών. Η αφαίρεση του φόντου είναι μια συμβατική και αποτελεσματική προσέγγιση για να ανιχνεύσεις κινούμενα αντικείμενα όταν το φόντο είναι ακίνητο. Για να ανιχνεύσεις όμως κινούμενα αντικείμενα σε μια δυναμική σκηνή, έχουν αναπτυχθεί προσαρμοστικές τεχνικές αφαίρεσης φόντου.

Ο Ren πρότεινε μια χωρική κατανομή από «Gaussian» μοντέλα, για να αντιμετωπίσει την ανίχνευση κινούμενων αντικειμένων, με αντιστάθμιση κίνησης, όπου υπάρχει. Τα αποτελέσματά του επέδειξαν την ικανότητα του συστήματος να ανιχνεύει πολύ μικρά αντικείμενα, τα οποία κινούνταν σε ένα φόντο με έντονη υφή. Ο Stuffer μοντελοποίησε κάθε pixel σαν μια μίξη από «Gaussians» και χρησιμοποίησε μια διαδικτυακή προσέγγιση για να ανανεώνει το μοντέλο. Το σύστημά του μπορούσε να αντιμετωπίσει αλλαγές στο φωτισμό, αντικείμενα που κινούνται με μικρή ταχύτητα και οποιαδήποτε αλλαγή στη σκηνή, που προέκυπτε από την πρόσθεση ή την αφαίρεση αντικειμένων.

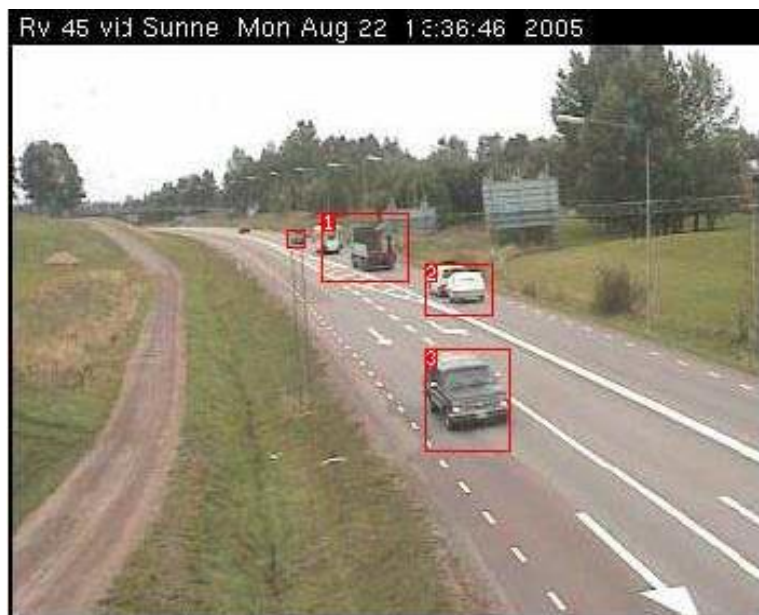
Ο Monnet πρότεινε μια διαδικτυακή μέθοδο με βάση την πρόβλεψη κίνησης, για την μοντελοποίηση δυναμικών σκηνών. Η μεθόδός του δοκιμάστηκε σε μια ακτή με κύματα και κινούμενα φύλλα δέντρων. Τα αποτελέσματα δεν ήταν ενθαρρυντικά αφού χρειάζονταν εκατοντάδες εικόνες της σκηνής χωρίς κινούμενα αντικείμενα, προκειμένου το σύστημα να καταφέρει να μάθει το φόντο. Επιπλέον, τα κινούμενα αντικείμενα δεν μπορούσαν να ανιχνευθούν αν κινούνταν στην ίδια κατεύθυνση με αυτή των κυμάτων της θάλασσας. Πρόσφατα, κάποιιοι υβριδικοί ανιχνευτές αναπτύχθηκαν, οι οποίοι συνδυάζουν μια απεικόνιση χρονικής διαφοράς με ένα προσαρμοστικό μοντέλο αφαίρεσης φόντου, για να αντιμετωπίσουν τις αλλαγές του φωτός.

Παρόλα αυτά, καμία από τις υπάρχουσες τεχνικές δεν μπορεί να προσαρμοστεί σε γρήγορες αλλαγές στην εικόνα, όπως είναι η αλλαγή ενός φωτός που αναβοσβήνει. Τα σημαντικότερα μειονεκτήματα των τεχνικών αυτών συνοψίζονται ως εξής:

- Χρειάζονται εκατοντάδες εικόνες για να μάθουν το φόντο.
- Δεν μπορούν να χειριστούν γρήγορες αλλαγές και παραπλανητικές κινήσεις.
- Δεν αφήνουν περιθώρια σε ακίνητα αντικείμενα να αρχίσουν να μετακινούνται.

Ποικίλες μέθοδοι για την ανίχνευση εμφανούς κίνησης έχουν αναπτυχθεί. Ο Wildes πρότεινε μια μέτρηση της ενδιαφέρουσας κίνησης με τη χρήση χωροχρονικού φιλτραρίσματος. Η μέτρηση αυτή υποθέτει ότι το αντικείμενο κινείται με συγκεκριμένη ταχύτητα, εξαιτίας της εξαρτώμενης από την ταχύτητα φύσεως των χωροχρονικών φίλτρων.

Αυτή η μέθοδος δεν λειτούργησε για αντικείμενα τα οποία κινούνται με πολύ μικρή ταχύτητα. Ο Wixson παρουσίασε μια μέθοδο για ανίχνευση κίνησης συσσωρεύοντας την ευθυγραμμισμένη ροή. Υπολόγισαν την οπτική ροή των υπό-pixel και ενσωμάτωσαν την οπτική ροή από frame σε frame του κάθε pixel, για να υπολογίσουν μια ακριβή προσέγγιση του πόσο έχει μετακινηθεί ολόκληρη η εικόνα. Σε κάθε frame ανανέωναν την τιμή η οποία ήταν άμεσα συνδεδεμένη με την απόσταση που έχει ταξιδέψει ένα σημείο. Η μέθοδός τους όμως ήταν πολύ χρονοβόρα και άφηνε ίχνη από τα κινούμενα αντικείμενα με το πέρασμα του χρόνου.



Εικόνα 3.5: Παράδειγμα αλγορίθμου ανίχνευσης κίνησης.

3.5 Ανίχνευση δέρματος

Το χρώμα και η υφή του δέρματος αποτελούν σημαντικές ενδείξεις που οι άνθρωποι χρησιμοποιούν συνειδητά ή ασυνείδητα για να συμπεράνουν μια ποικιλία από πολιτισμικές πτυχές των υπολοίπων. Το χρώμα του δέρματος και η υφή μπορεί να είναι μια ένδειξη της φυλής, της ηλικίας, και της ομορφιάς του καθενός από εμάς, αλλά οι ερμηνείες ποικίλουν ανά τους διαφορετικούς πολιτισμούς και τις χρονικές περιόδους.

Στις εικόνες και στα βίντεο όμως, το χρώμα του δέρματος είναι ένα σημάδι παρουσίας ανθρώπου. Γι αυτό το λόγο, τις τελευταίες δυο δεκαετίες αναλυτική έρευνα έχει εστιάσει στην ανίχνευση δέρματος. Η ανίχνευση δέρματος είναι η διαδικασία με την οποία εντοπίζουμε pixel και περιοχές που έχουν το χρώμα του δέρματος. Αυτή η διαδικασία χρησιμοποιείται συνήθως ως ένα βήμα προεπεξεργασίας για να βρούμε τις περιοχές που περιέχουν ενδεχομένως ανθρώπινα πρόσωπα και άκρα σε εικόνες.



Εικόνα 3.6: Ανίχνευση δέρματος.

Διαφορετικές προσεγγίσεις τεχνητής όρασης έχουν αναπτυχθεί για την ανίχνευση δέρματος. Ένας ανιχνευτής δέρματος συνήθως μετασχηματίζει ένα δεδομένο pixel σε ένα κατάλληλο χώρο χρωμάτων (color space) και στη συνέχεια χρησιμοποιεί ένα ταξινομητή δέρματος για να τα ονομάσει σε pixel δέρματος ή μη-δέρματος. Ένας

ταξινομητής καθορίζει ένα πεδίο στο χώρο χρωμάτων ως χρώματα δέρματος βασιζόμενος σε μια εκπαιδευόμενη βάση δεδομένων.

Όπως θα αναλύσουμε σε αυτό το κεφάλαιο, η ανίχνευση δέρματος είναι ένα υπολογιστικά εύκολο έργο και μπορεί να πραγματοποιηθεί πολύ αποτελεσματικά, ένα χαρακτηριστικό που ενθαρρύνει τη χρήση της ανίχνευσης δέρματος σε πολλές εφαρμογές που ασχολούνται με ανάλυση βίντεο. Παρόλα αυτά, η ανίχνευση pixel στον τόνο του δέρματος, αποδείχτηκε ότι είναι αρκετά περίπλοκη για πολλούς άλλους λόγους. Η εμφάνιση του δέρματος σε μια εικόνα εξαρτάται από τις συνθήκες του φωτισμού του μέρους που έγινε η λήψη. Εμείς σαν άνθρωποι, είμαστε πολύ καλοί στο να αναγνωρίζουμε τα χρώματα των αντικειμένων σε οποιαδήποτε φωτισμό.

Ως εκ τούτου, μια σημαντική πρόκληση στην ανίχνευση δέρματος είναι να αναπαραστήσουμε το χρώμα με ένα τρόπο με τον οποίο να μένει αναλλοίωτο ή τουλάχιστον να μην είναι πολύ ευαίσθητο στις αλλαγές του φωτισμού. Η επιλογή του χώρου χρωμάτων επηρεάζει πάρα πολύ την απόδοση οποιουδήποτε ανιχνευτή δέρματος, και την ευαισθησία του στις αλλαγές του φωτισμού. Άλλη μια πρόκληση αποτελούν τα αντικείμενα που έχουν τον ίδιο τόνο με το ανθρώπινο δέρμα, όπως είναι το ξύλο, το δέρμα, η άμμος κτλ. Αυτό προκαλεί, πολλές λάθος ανιχνεύσεις, σε οποιονδήποτε ανιχνευτή δέρματος, σε μη ελεγχόμενα περιβάλλοντα.

Η διαδικασία της ανίχνευσης δέρματος έχει δυο φάσεις. Τη φάση της εκμάθησης, και τη φάση της ανίχνευσης. Το στάδιο της εκπαίδευσης περιλαμβάνει τρία βασικά στάδια:

1. Συλλογή μιας βάσης δεδομένων με τόνους δέρματος από διαφορετικές εικόνες. Μια τέτοια βάση συνήθως περιέχει τόνους από μια μεγάλη ποικιλία ανθρώπων σε διαφορετικές καταστάσεις φωτός.
2. Να διαλέξουμε τον κατάλληλο χώρο χρώματος.
3. Εκμάθηση των παραμέτρων ενός ταξινομητή δέρματος.

Σε κάθε εκπαιδευόμενο ανιχνευτή δέρματος, ο προσδιορισμός των πιξελ του δέρματος γίνεται σε τρεις φάσεις:

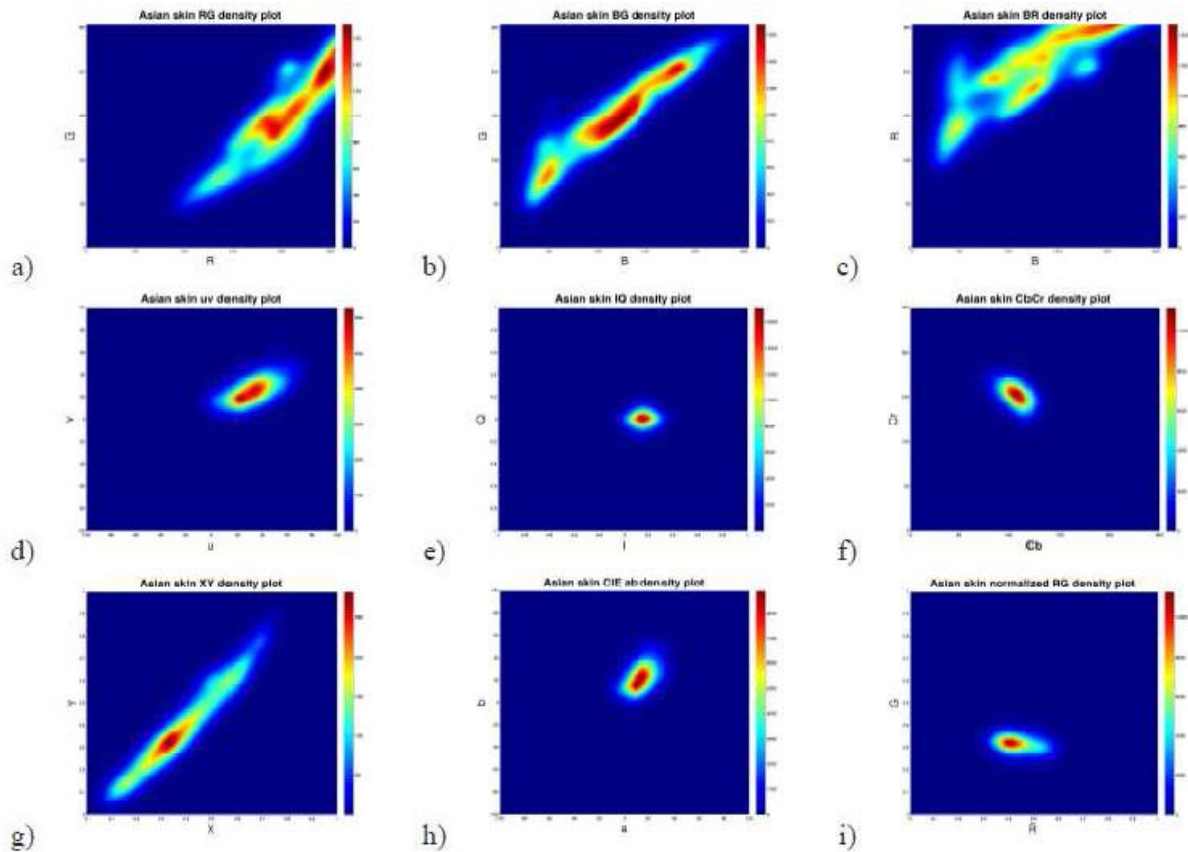
1. Τη μετατροπή της εικόνας στον ίδιο χώρο χρωμάτων που χρησιμοποιήσαμε στο στάδιο της εκμάθησης
2. Να κατατάξουμε τα pixel με τη χρήση του ταξινομητή δέρματος σε δέρμα ή μη-δέρμα.
3. Τέλος, τα pixel επεξεργάζονται μορφολογικά για να επιβληθεί χωρική ομοιογένεια στις περιοχές που εντοπίστηκαν.

Σε οποιοδήποτε χώρο χρωμάτων, το χρώμα του δέρματος καταλαμβάνει ένα συμπαγές και μεγάλο μέρος στο χώρο. Μια τέτοια περιοχή ονομάζεται σύμπλεγμα χρώματος δέρματος.

Ένας ταξινομητής δέρματος είναι ένα πρόβλημα ταξινόμησης πρώτης ή δεύτερης κατηγορίας. Ένα pixel κατηγοριοποιείται και επισημαίνεται είτε πρόκειται για δέρμα ή μη δέρμα, δεδομένου ενός μοντέλου δέρματος σε ένα συγκεκριμένο χώρο χρώματος. Στο πλαίσιο της ταξινόμησης του δέρματος τα αληθώς θετικά είναι pixel δέρματος που ο ταξινομητής ορθώς επισύναψε ως δέρμα. Τα αληθώς αρνητικά είναι περιοχές μη δέρματος που ο ταξινομητής ορθώς επισύναψε ως μη δέρμα. Όμως κάθε ταξινομητής κάνει λάθη.

Μπορεί να επισυνάψει μια περιοχή που δεν είναι δέρμα σαν δέρμα και το αντίστροφο. Αυτά τα λάθη ονομάζονται ψευδώς θετικά στην πρώτη περίπτωση και ψευδώς αρνητικά στη δεύτερη, αλλά και τα δυο οδηγούν σε λάθος ανίχνευση. Ένας καλός ταξινομητής πρέπει να έχει χαμηλά ποσοστά ψευδώς θετικών και αρνητικών επισυνάψεων.

Το ίδιο ισχύει και στην ανίχνευση δέρματος, γι αυτό το λόγο η επιλογή του χώρου χρωμάτων είναι μια πολύ σημαντική πτυχή της ανίχνευσης δέρματος. Το χρώμα χρειάζεται να αναπαρασταθεί σε ένα χώρο χρωμάτων στον οποίο η τάξη του δέρματος είναι πιο συμπαγής έτσι ώστε να μπορεί να ταξινομηθεί πιο εύκολα. Η επιλογή του χώρου χρωμάτων επηρεάζει άμεσα και τον ταξινομητή που θα χρησιμοποιήσουμε.



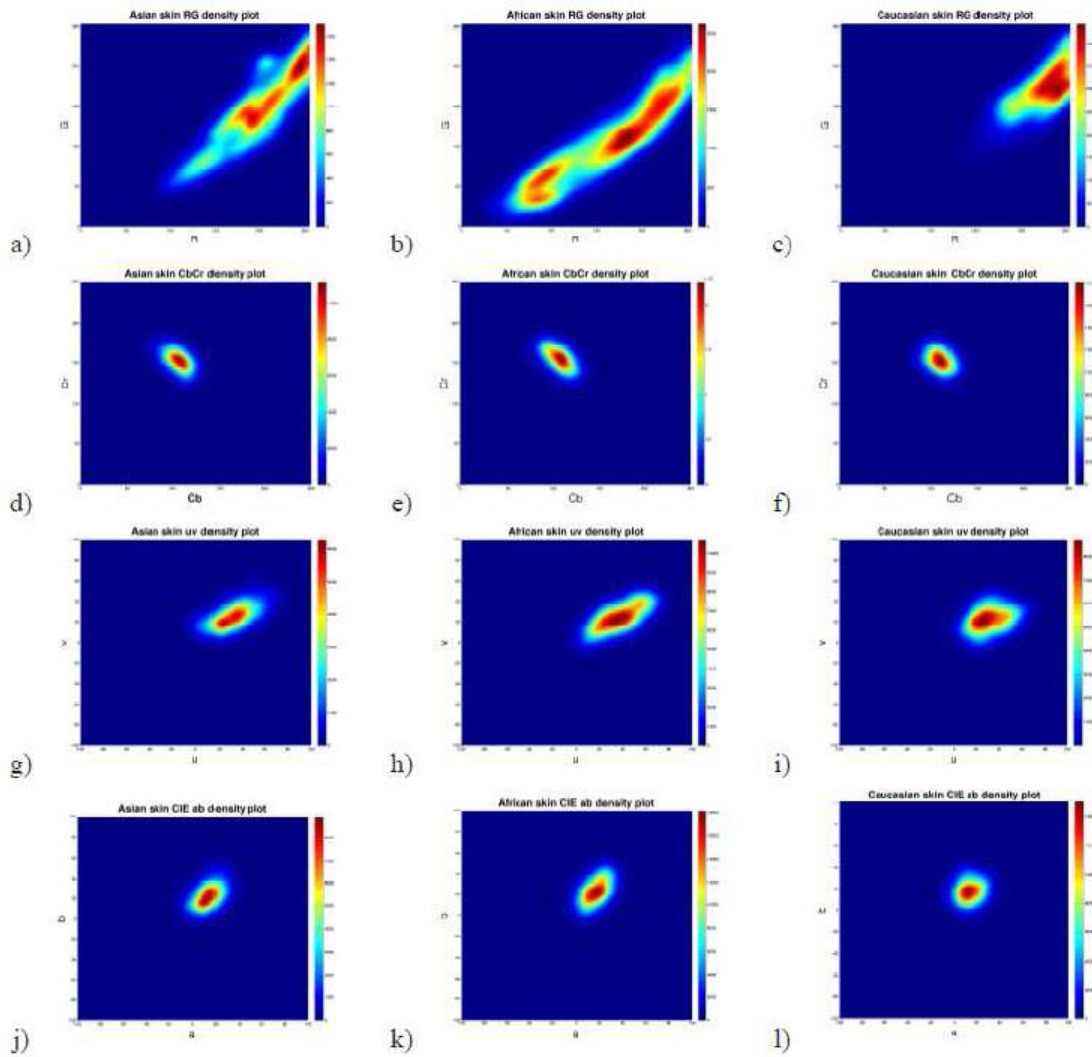
Εικόνα 3.7: Διάγραμμα πυκνότητας δέρματος Ασιατών σε διαφορετικούς χρωματικούς χώρους.

Ανίχνευση δέρματος και γόροι χρωμάτων

Οι Forsyth και Fleck επισήμαναν ότι το χρώμα του ανθρώπινου δέρματος έχει ένα περιορισμένο εύρος αποχρώσεων αφού η εμφάνιση του δέρματος προκύπτει από ένα συνδυασμό του κόκκινου του αίματος και της μελανίνης.

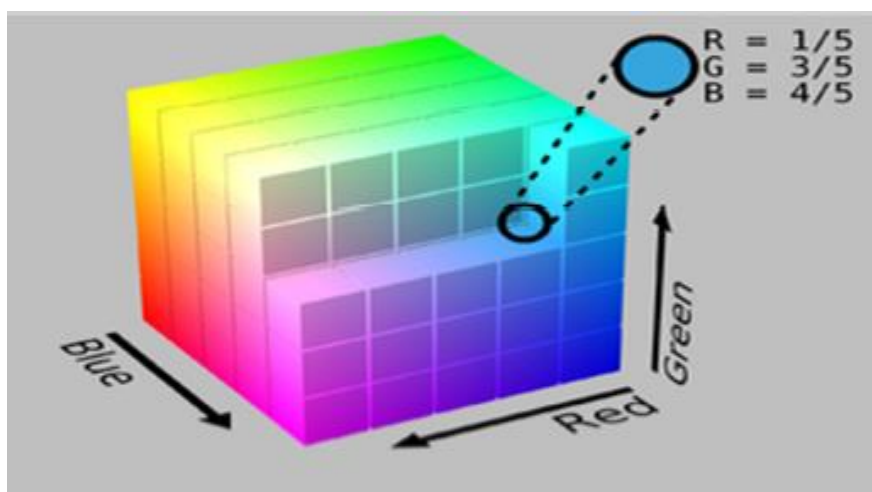
Γι αυτό το λόγο, το χρώμα του δέρματος δεν συμπίπτει τυχαία με ένα οποιοδήποτε χώρο χρωμάτων αλλά ομαδοποιείται σε μια μικρή περιοχή του χρωματικού χώρου. Μια μεγάλη ποικιλία από χρωματικούς χώρους χρησιμοποιείται προκειμένου να βρεθεί ο κατάλληλος στον οποίο το χρώμα του δέρματος κάτω από διαφορετικές συνθήκες φωτισμού θα παραμένει ανεπηρέαστο.

Σε αυτή την ενότητα θα περιγράψουμε ορισμένους χρωματικούς χώρους που χρησιμοποιούνται πιο συχνά στην ανίχνευση δέρματος και θα συζητήσουμε τον τρόπο με τον οποίο την επηρεάζουν.



Εικόνα 3.8: Διάγραμμα πυκνότητας δέρματος Ασιατών, Αφρικανών και Καυκάσιων σε διαφορετικούς χρωματικούς χώρους.

- RGB χρωματικός χώρος



Εικόνα 3.9: RGB χρωματικός χώρος.

Ο RGB είναι ο πιο ευρέως χρησιμοποιούμενος χρωματικός χώρος στις ψηφιακές εικόνες. Κωδικοποιεί τα χρώματα σαν ένα συνδυασμό των τριών βασικών χρωμάτων, του κόκκινου (Red), του πράσινου (Green) και του μπλε (Blue).

Ο RGB χρωματικός χώρος συνήθως απεικονίζεται σαν ένας τρισδιάστατος κύβος στον οποίο τα τρία αυτά χρώματα είναι οι 3 άξονες. Ένα κύριο πλεονέκτημα του RGB είναι η απλότητά του. Όμως, δεν είναι ομοιομόρφως αντιληπτά κάτι το οποίο σημαίνει ότι οι αποστάσεις του RGB δεν αντιστοιχούν γραμμικά με την ανθρώπινη αντίληψη. Επιπλέον, ο συγκεκριμένος χρωματικός χώρος δεν διαχωρίζει τη φωτεινότητα και τη χρωματικότητα και οι τρεις συνιστώσες συσχετίζονται σε πολύ μεγάλο βαθμό.

Μια οποιαδήποτε απόχρωση στο RGB αποτελείται από ένα συνδυασμό των βασικών αυτών χρωμάτων γι αυτό το λόγο αν αλλάξουμε το φωτισμό σε ένα κομμάτι δέρματος επηρεάζονται και οι τρεις συνιστώσες. Με άλλα λόγια, το μέρος στο οποίο βρίσκουμε το χρώμα του δέρματος στο κύβο RGB αλλάζει ανάλογα με τη φωτεινότητα στην οποία έγινε η λήψη. Αυτό έχει σαν αποτέλεσμα, ένα ευρύ φάσμα από αποχρώσεις του δέρματος στον RGB κύβο. Αυτό γίνεται αντιληπτό στην πρώτη γραμμή της εικόνας 3.7, στον οποίο μοντέλα δέρματος από εικόνες ασιατών, οι οποίες τραβήχτηκαν κάτω από τυχαίες συνθήκες φωτισμού, σχεδιάστηκαν στον RGB χρωματικό χώρο.

Η ομαδοποίηση του χρώματος του δέρματος επεκτείνεται σε ένα μεγάλο μέρος του χρωματικού χώρου προκειμένου να καλύψει τις αποχρώσεις που προκύπτουν από διαφορετικούς φωτισμούς. Ομοίως, οι ομάδες χρωμάτων από διαφορετικές φυλές εντοπίζονται σε διαφορετικά σημεία του RGB κύβου. Αυτό μπορούμε να το δούμε στην πρώτη σειρά της εικόνας 3.8.

Παρ' όλους τους περιορισμούς, το RGB χρησιμοποιείται στο θεωρητικό κομμάτι της ανίχνευσης δέρματος εξαιτίας της απλότητάς του.

- Χρωματικός χώρος TV

Μια διαφορετική κατηγορία χώρου χρωμάτων είναι οι ορθογώνιες, οι οποίες χρησιμοποιούνται στην τηλεοπτική μετάδοση. Σε αυτή περιλαμβάνονται οι YUV, YIQ και YCbCr.

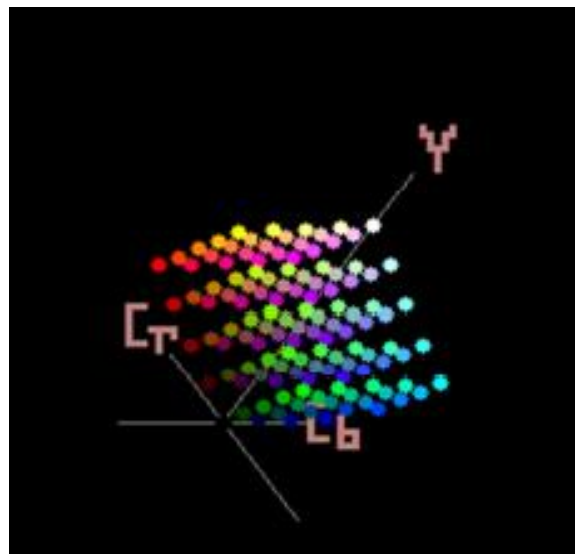
Η YIQ χρησιμοποιείται στην NTSC τηλεοπτική εκπομπή ενώ η YCbCr χρησιμοποιείται στο JPEG πρότυπο συμπίεσης εικόνας και στο MPEG πρότυπο συμπίεσης βίντεο. Ένα πλεονέκτημα που προκύπτει από τη χρήση αυτού του

χρωματικού χώρου είναι ότι τα περισσότερα βίντεο είναι ήδη κωδικοποιημένα σε αυτό το χρωματικό χώρο. Ο μετασχηματισμός από RGB σε οποιοδήποτε από αυτούς τους χρωματικούς χώρους είναι μια απλή γραμμική μετατροπή. Όλοι αυτοί οι χρωματικοί χώροι διαχωρίζουν το κανάλι του φωτισμού (Y) από τα δυο ορθογώνια χρωματικά κανάλια (UV , IQ , C_bC_r).

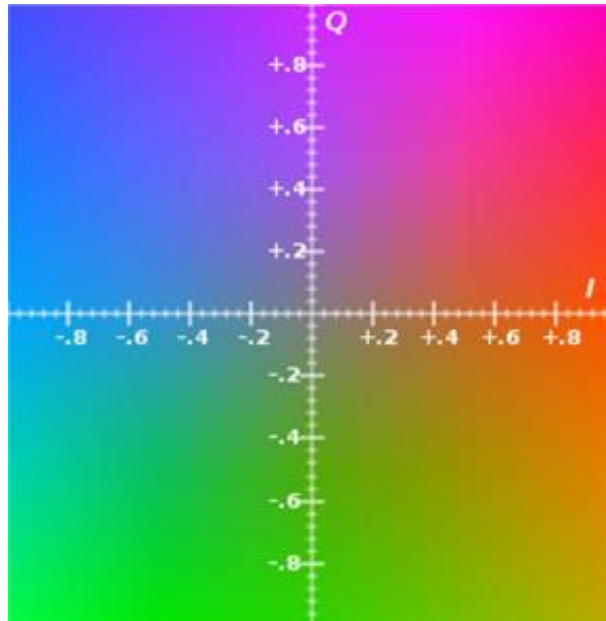
Συνεπώς, σε αντίθεση με το RGB, η τοποθεσία που βρίσκεται το χρώμα του δέρματος στο κανάλι χρώματος δεν επηρεάζεται από αλλαγή της έντασης της φωτεινότητας όπως βλέπουμε και στην εικόνα 12-d,e,f. Αυτό μας διευκολύνει και στην κατασκευή ενός ανιχνευτή δέρματος ο οποίος είναι αμετάβλητος από την ένταση της φωτεινότητας και χρησιμοποιεί πιο απλούς ταξινομητές. Η πυκνότητα του χρώματος του δέρματος πάνω στα χρωματικά κανάλια μπορεί εύκολα να προσεγγιστεί με τη χρήση μιας πολυμεταβλητής Gaussian κατανομής.

Επιπλέον, τα χρώματα του δέρματος διαφορετικών φυλών σχεδόν συμπίπτουν στα χρωματικά κανάλια, όπως φαίνεται στην δεύτερη και τρίτη σειρά της εικόνας 3.8. Ως εκ τούτου, η χρήση ενός τέτοιου χρωματικού χώρου έχει σαν αποτέλεσμα ανιχνευτές δέρματος, οι οποίοι είναι ανεξάρτητοι από τις πιθανές αποχρώσεις του ανθρώπινου δέρματος.

Η απλότητα αυτού του μετασχηματισμού καθώς και οι ιδιότητες που περιγράψαμε καθιστούν αυτούς τους χρωματικούς χώρους ευρέως χρησιμοποιούμενους σε εφαρμογές ανίχνευσης δέρματος.



Εικόνα 3.10: Αναπαράσταση του YCbCr χρωματικού χώρου.

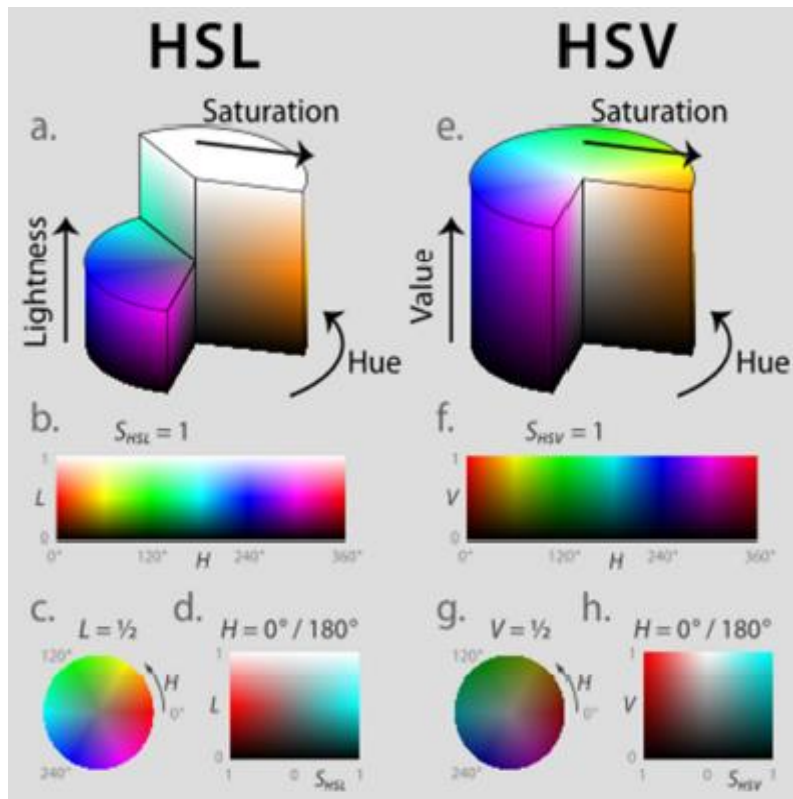


Εικόνα 3.11: Αναπαράσταση του YIQ χρωματικού χώρου.

- Αντιληπτικός χρωματικός χώρος

Αντιληπτικοί χρωματικοί χώροι, όπως είναι οι HIS, HSV/HSB και HSL (HLS) είναι εξίσου δημοφιλείς στην ανίχνευση δέρματος. Αυτοί οι χρωματικοί χώροι διαχωρίζουν τα τρία συστατικά: την απόχρωση (H), τον κορεσμό (S) και τη φωτεινότητα (I, V ή L). Ουσιαστικά, οι HSV χρωματικοί χώροι είναι μετασχηματισμοί του RGB χρωματικού κύβου, και μπορούν να απεικονιστούν από τον RGB χρωματικό χώρο μέσω ενός μη γραμμικού μετασχηματισμού.

Ένα από τα πλεονεκτήματα αυτού του χρωματικού χώρου στην ανίχνευση δέρματος είναι ότι επιτρέπουν στους χρήστες να προσδιορίσουν ενστικτωδώς το όριο της τάξης του χρώματος του δέρματος από άποψη κορεσμού και φωτεινότητας. Όσο για την τελευταία συνιστώσα (I, V, L) η οποία μας δίνει τη φωτεινότητα, συνήθως δεν την λαμβάνουμε τόσο υπόψη, προκειμένου να μειώσουμε την εξάρτηση του χρώματος του δέρματος από την φωτεινότητα.



Εικόνα 3.12: HSL και HSV.

- Χρωματομετρικός χρωματικός χώρος

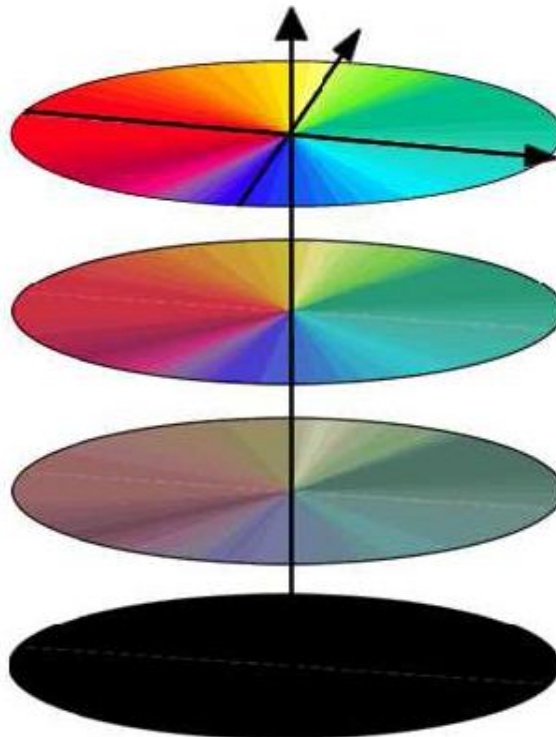
Ο διαχωρισμός της χρωματικότητας από την φωτεινότητα επιτυγχάνεται και στους χρωματομετρικούς χρωματικούς χώρους, όπως είναι οι CIE-XYZ, CIE-xy, CIE-lab. Ο CIE-XYZ χρωματικός χώρος είναι ένας από τους πρώτους μαθηματικώς ορισμένους χρωματικούς χώρους.

Είναι βασισμένος σε εκτεταμένες μετρήσεις της ανθρώπινης οπτικής αντίληψης και χρησιμεύει ως βάση για πολλούς άλλους χρωματομετρικούς χώρους. Ο CIE-XYZ μπορεί να αναπαρασταθεί μέσω ενός μετασχηματισμού γραμμικών συντεταγμένων του RGB χρωματικού χώρου. Η Y συνιστώσα αντιστοιχεί στη φωτεινότητα του χρώματος. Οι τιμές χρωματικότητας μπορούν να μετασχηματιστούν από το τρισδιάστατο επίπεδο και στη συνέχεια να προβληθούν στο XY.

Το αποτέλεσμα είναι το πολύ γνωστό CIE-xy χρωματικό διάγραμμα, το οποίο καθορίζει τον κορεσμό και την απόχρωση του κάθε χρώματος. Ένα από τα μειονεκτήματα των XYZ και xy χρωματικών χώρων, είναι ότι οι διαφορές χρώματος δεν γίνονται εύκολα αντιληπτές σε όλες τις περιοχές του χρωματικού χώρου. Σε αντίθεση ο CIElab διαχωρίζει την μεταβλητή της φωτεινότητας L, από τις δυο ενιαίες

χρωματικές μεταβλητές a,b. Στην εικόνα 3.7-h μπορούμε να δούμε την πυκνότητα του χρώματος του δέρματος των Ασιατών στον a,b χρωματικό χώρο. Στην τελευταία σειρά της εικόνας 3.8 βλέπουμε την πυκνότητα του χρώματος από διαφορετικές φυλές στον a,b χρωματικό χώρο.

Παρόλο που οι συγκεκριμένοι χρωματικοί χώροι έχουν πολλά πλεονεκτήματα, χρησιμοποιούνται σπανίως στην ανίχνευση δέρματος, κυρίως επειδή ο μετασχηματισμός από RGB είναι υπολογιστικά ακριβός.



Εικόνα 3.13: Αναπαράσταση του CIE-lab.

Ταξινομητές δέρματος

Μια ποικιλία από τεχνικές ταξινόμησης χρησιμοποιούνται στο θεωρητικό κομμάτι για την ταξινόμηση του δέρματος. Ένας ταξινομητής δέρματος είναι ένας πρώτης τάξης ταξινομητής, που καθορίζει ένα όριο απόφασης για την τάξη του χρώματος του δέρματος σε ένα χρωματικό χώρο. Κάθε pixel, του οποίου το χρώμα συμπίπτει με αυτά που υπάρχουν στον ταξινομητή του δέρματος, επισημαίνεται σαν δέρμα. Γι αυτό το λόγο, η επιλογή του ταξινομητή δέρματος είναι άμεσα συνδεδεμένη με το χρωματικό χώρο που θα χρησιμοποιήσουμε. Ο πιο απλός τρόπος για να

αποφασίζουμε αν ένα pixel είναι στο χρώμα του δέρματος ή όχι είναι να ορίσουμε ένα όριο.

Οι Brand και Mason δημιούργησαν έναν απλό μονοδιάστατο ταξινομητή, σύμφωνα με τον οποίο το pixel ονομαζόταν δέρμα αν η αναλογία μεταξύ των καναλιών R και G ήταν ανάμεσα σε ένα ανώτατο και έναν κατώτατο όριο. Πειραματίστηκαν επίσης με ένα μονοδιάστατο όριο στο IQ επίπεδο του YIQ χώρου όπου το I ήταν η τιμή που χρησιμοποιήθηκε για κατώφλι.

- Η προσέγγιση του Bayse για ανίχνευση δέρματος

Η ταξινόμηση του δέρματος μπορεί να οριστεί με τη χρήση πιθανοτήτων με τον ακόλουθο τρόπο.

Έστω ένα pixel με χρώμα c. Η πιθανότητα να είναι pixel δέρματος είναι $P(\text{skin}|c)$. Όταν υπολογιστεί αυτή η πιθανότητα, το pixel παίρνει την ετικέτα δέρμα αν η πιθανότητα αυτή είναι πάνω από το ορισμένο όριο, αλλιώς αν είναι κάτω από το όριο επισημαίνεται σαν μη-δέρμα. Προφανώς όμως, δεν μπορούμε να υπολογίζουμε την πιθανότητα για κάθε πιθανό χρώμα (πχ στο 24 bit RGB υπάρχουν περίπου 1.7 εκατομμύρια χρώματα). Χρησιμοποιώντας τον κανόνα του Bayse ο υπολογισμός της πιθανότητας μπορεί να ξαναγραφεί ως:

$$P(\text{skin}|c) = \frac{P(c|\text{skin})P(\text{skin})}{P(c|\text{skin})P(\text{skin}) + P(c|\text{notskin})P(\text{notskin})}$$

Σύμφωνα με τον κανόνα του Bayse αυτό έχει ως αποτέλεσμα τον υπολογισμό της πιθανής αναλογίας ενός δεδομένου χρώματος σε μια δεδομένη τάξη δέρματος σε σχέση με μια τάξη μη-δέρματος ($P(c|\text{skin})/P(c|\text{non skin})$).

Αυτή η αναλογία μπορεί στη συνέχεια να οριστεί σαν κατώφλι για να αποφασιστεί αν ένα pixel είναι δέρμα ή όχι.

- Εφαρμογές και παραδείγματα ανίχνευσης δέρματος

Ο εντοπισμός και η ανίχνευση του προσώπου είναι το πρώτο βήμα στην απόκτηση βιομετρικών στοιχείων του προσώπου. Το χρώμα του δέρματος είναι ένα χαρακτηριστικό στοιχείο στα ανθρώπινα πρόσωπα. Σε ένα ελεγχόμενο φόντο, η ανίχνευση δέρματος μπορεί να είναι ένα βολικό εργαλείο για τον εντοπισμό προσώπων στην εικόνα. Από τη στιγμή που η επεξεργασία του χρώματος είναι αρκετά πιο γρήγορη από την επεξεργασία οποιουδήποτε άλλου χαρακτηριστικού, η

ανίχνευση δέρματος μπορεί να χρησιμοποιηθεί σαν ένα προκαταρκτικό στάδιο για άλλες μεθόδους ανίχνευσης προσώπου. Η ανίχνευση δέρματος χρησιμοποιείται επίσης και για τον εντοπισμό άκρων του σώματος.

Οι Forsyth και Fleck απέδειξαν ότι η ανίχνευση δέρματος μπορεί να χρησιμοποιηθεί ως μέρος της διαδικασίας ανίχνευσης εικόνων με γυμνούς ή προκλητικά ντυμένους ανθρώπους. Η τεχνική τους είχε τρία στάδια. Αρχικά, ένα φίλτρο δέρματος, βασισμένο στο χρώμα και την υφή, χρησιμοποιούνταν για να επιλέξει τις εικόνες με μεγάλες περιοχές στο χρώμα του δέρματος. Έπειτα τα αποτελέσματα, περνούσαν από ένα γεωμετρικό φίλτρο, το οποίο προσδιόριζε τις περιοχές που ήταν στο χρώμα του δέρματος με κυλινδρικά σχήματα.

Αυτά τα κυλινδρικά σχήματα ομαδοποιούνταν σε πιθανά ανθρώπινα άκρα. Οι εικόνες που είχαν μεγάλες ομάδες από άκρα στο χρώμα του δέρματος θεωρούνταν ότι περιείχαν γυμνές φωτογραφίες.

- Σύνοψη

Ανεξαρτήτως από την επιλογή που θα κάνουμε για το χρωματικό χώρο και τη μέθοδο ταξινόμησης, οι περισσότερες έρευνες πάνω στην ανίχνευση δέρματος έχουν ποσοστό περίπου 60% σωστών ανιχνεύσεων, ενώ το ποσοστό λανθασμένων ανιχνεύσεων κυμαίνεται από 15-30%.

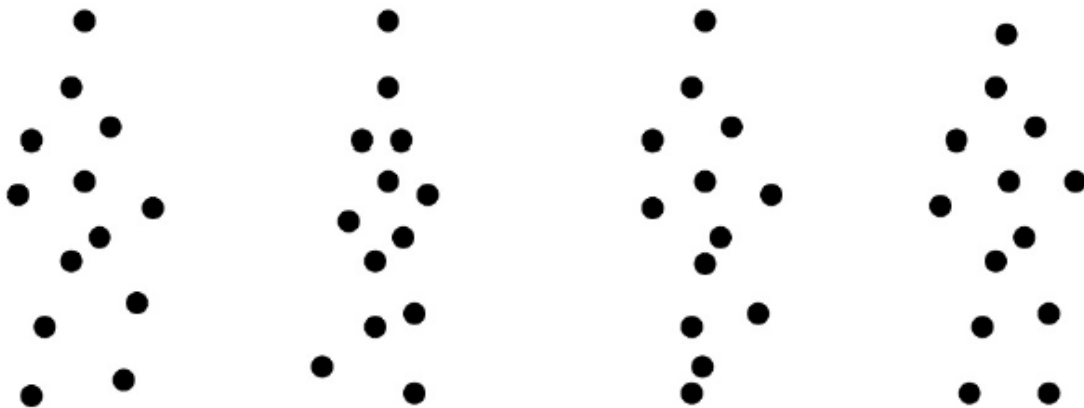
Η ανίχνευση δέρματος σε έγχρωμες εικόνες και βίντεο, είναι ένας πολύ εύκολος τρόπος για να ανιχνεύσουμε την παρουσία ανθρώπου. Παρόλα αυτά, αρκετά αντικείμενα στον πραγματικό χώρο έχουν χρώμα στον τόνο του ανθρώπινου δέρματος, τα οποία μπορούν πολύ εύκολα να ανιχνευθούν λανθασμένα ως δέρμα. Γι αυτό το λόγο η ανίχνευση δέρματος έχει εγγυημένα αποτελέσματα μόνο σε ελεγχόμενα περιβάλλοντα στα οποία το φόντο είναι ελεγχόμενο. Επίσης μιας και η ανίχνευση γίνεται με βάση το χρώμα του κάθε pixel η χρήση του περιορίζεται μόνο σε έγχρωμες εικόνες, οπότε δεν έχει καμία χρησιμότητα σε εικόνες gray-scale, υπέρυθρες και γενικότερα όσες δεν έχουν χρωματική πληροφορία.

3.6 Εντοπισμός χαρακτηριστικών σημείων

Η ανίχνευση ανθρώπου βασισμένη μόνο στην κίνηση, αρχικά, φαίνεται εξεζητημένη. Παρέχει η κίνηση των άκρων αρκετή πληροφορία για να εντοπίσει ανθρώπινη παρουσία;

Πειράματα που διεξήχθησαν από τον Johansson το 1970 απέδειξαν ότι αρκεί η κίνηση των άκρων προκειμένου να αναγνωριστεί ανθρώπινη παρουσία. Ο Johansson κινηματογράφησε ανθρώπους που κινιόντουσαν σε ένα κατάμαυρο δωμάτιο, με μόνη φωτεινή ένδειξη ένα λευκό σημείο φωτός προσκολλημένο στα άκρα τους. Απέδειξε ότι κάποιος που παρακολουθούσε το φιλμ μπορεί εύκολα να αναγνωρίσει την ανθρώπινη κίνηση, παρόλη την απουσία οπτικών ενδείξεων όπως σχήμα, υφή φωτεινότητα και χρώμα. Ένα παράδειγμα των σημείων του Johansson φαίνεται στην εικόνα 3.13.

Δεδομένου ότι ο άνθρωπος μπορεί χωρίς προσπάθεια να αναγνωρίσει την κίνηση, είναι πιθανό και ένας αλγόριθμος να κάνει το ίδιο. Επιπρόσθετα, τα μεμονωμένα σημεία κίνησης όπως χρησιμοποιούνται στο πείραμα του Johansson, μπορούν να αναπαρασταθούν αποτελεσματικά σε υπολογιστή. Αυτό μας δίνει την ελπίδα ότι ένας αποδοτικός αλγόριθμος είναι πιθανό να επιτευχθεί .



Εικόνα 3.14: Ακολουθία από σημεία του Johansson που δείχνει την πλαϊνή λήψη ενός ανθρώπου που περπατάει.

Το πείραμα του Johansson μας δείχνει ότι η ανίχνευση ανθρώπου από την κίνηση των χαρακτηριστικών αυτών σημείων είναι ένας ρεαλιστικός στόχος. Με αυτό το σκεπτικό στο μυαλό μας, μπορούμε να χωρίσουμε ένα τέτοιο σύστημα σε 3 διαφορετικές ρουτίνες.

1. Εξαγωγή των χαρακτηριστικών σημείων από βίντεο.
2. Εύρεση ενός μοντέλου που αντιπροσωπεύει με ακρίβεια την ανθρώπινη κίνηση.
3. Εφαρμογή αυτού του μοντέλου σε έναν ανιχνευτή, ο οποίος θα αποφασίζει αν μια ομάδα από κινούμενα σημεία είναι αντιπροσωπευτική ανθρώπινης κίνησης.

Κατανόηση του προβλήματος

Η προσέγγιση της Song στην ανίχνευση ανθρώπινης κίνησης είναι ένα πιθανολογικό μοντέλο, στο οποίο τα κοινά χαρακτηριστικά και μεταβλητές της ανθρώπινης κίνησης εισάγονται σε ένα μοναδικό στατιστικό μοντέλο. Αυτό το μοντέλο αποθηκεύεται σαν ένα γράφημα το οποίο περιέχει ακμές και γωνίες.

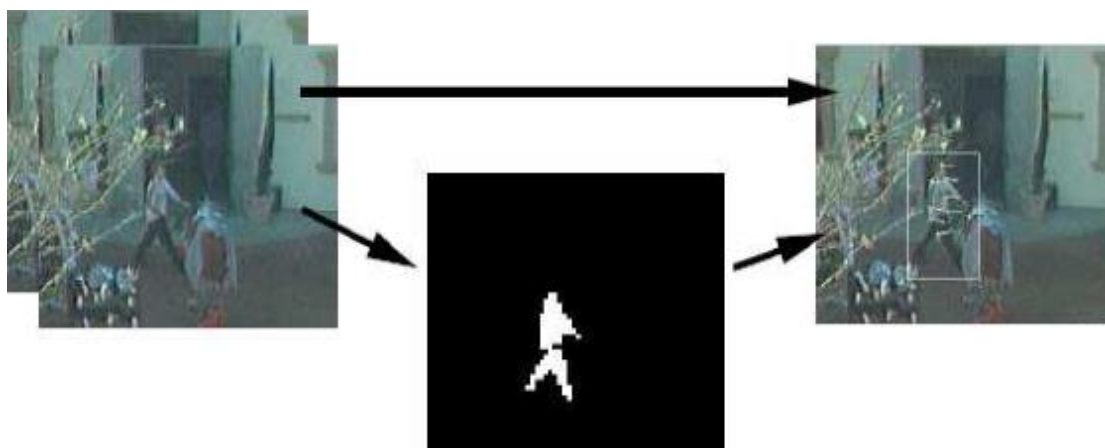
Κάθε κορυφή αναπαριστά ένα χαρακτηριστικό κίνησης σε κάποιο σημείο από ανθρώπινη σιλουέτα. Η κορυφή αναπαριστάται από μια ευκλείδεια θέση (x,y) η οποία προσδιορίζει το μέσο όρο της θέσης των χαρακτηριστικών αυτών, στον ανθρώπινο σώμα αλλά και το μέσο όρο της ταχύτητας με την οποία κινούνται (U_x, U_y) . Επιπρόσθετα, υπάρχει άλλος ένας πίνακας ο οποίος συσχετίζει αυτές τις τέσσερις παραμέτρους με αυτές σε κάθε άλλη κορυφή. Κάθε 2 κορυφές οι οποίες συνδέονται με μια ακμή, στατιστικώς θεωρούνται εξαρτημένες η μια από την άλλη, και αυτές που δεν ενώνονται θεωρούνται ανεξάρτητες.

Σε αυτό το σημείο πρέπει να αναφέρουμε ότι η αποδοτικότητα ενός τέτοιου συστήματος είναι $O(MN^3)$, όπου M είναι ο αριθμός των τριγώνων σε ένα γράφημα και N ο αριθμός των σημείων που αξιολογούμε. Γι αυτό το λόγο, πρέπει να φροντίσουμε να κρατήσουμε αυτά τα σημεία στο ελάχιστο. Η μείωση των σημείων αυτών απαιτεί πιο έξυπνη επιλογή τους. Θα ήταν μη πρακτικό να βρούμε κάθε σημείο κίνησης σε όλη την εικόνα και να τα εισάγουμε στο σύστημα για αξιολόγηση. Αρχικά τα σημεία αυτά θα ήταν παρά πολλά με αποτέλεσμα να έχουμε πολύ αργή απόδοση. Δεύτερον, θα δίναμε άχρηστα σημεία όπως είναι αυτά που ανήκουν στο background. Για να αποφύγουμε τη σπατάλη χρόνου εξετάζοντας το background, χρησιμοποιούμε image segmentation για να διαχωρίσουμε το πίσω από το μπροστά μέρος μιας σκηνής.

Η κατάτμηση εικόνας είναι η ενέργεια που κάνει οποιοσδήποτε αλγόριθμος ο οποίος διαχωρίζει τις περιοχές μιας εικόνας με τέτοιο τρόπο, όπως θα τις διαχώριζε και ένας άνθρωπος. Μιας και μας ενδιαφέρει η κίνηση, η φυσική προσέγγιση είναι να απομονώσουμε τις περιοχές της εικόνας που κινούνται, από το background. Αφού διαχωρίσουμε την εικόνα με βάση την κίνηση, τα σημεία που μας ενδιαφέρουν μπορούν να εντοπιστούν ξεχωριστά σε κάθε περιοχή. Αυτό μας επιτρέπει να τρέξουμε την αξιολόγηση ξεχωριστά για κάθε περιοχή.

Αφού η αξιολόγηση του μοντέλου έχει τύπο $O(MN^3)$ τρέχει πιο γρήγορα σε 2 περιοχές με 10 σημεία ενδιαφέροντος παρά σε μια περιοχή 40 σημείων. Αυτά τα δυο

βασικά στάδια, η κατάτμηση εικόνας και ο εντοπισμός σημείων συνθέτουν το πρώτο μισό από ένα σύστημα ανίχνευσης κίνησης.



Εικόνα 3.15: Κατάτμηση εικόνας.

Κατάτμηση εικόνας

Ο στόχος μας στην κατάτμηση εικόνας, είναι να διαχωρίσουμε την περιοχή του background με αυτή του foreground που παρουσιάζει ενδιαφέρον, ως προς την ανίχνευση κίνησης. Σε ιδανικές συνθήκες, δηλαδή αν τα στοιχεία του background δεν μετακινούνται και δεν υπάρχει απότομη αλλαγή στην φωτεινότητα του, ο αλγόριθμος, ο οποίος κάνει την αφαίρεση του background, δημιουργεί ένα μοντέλο της εικόνας του φόντου με την πάροδο του χρόνου.

Για κάθε frame του video αφαιρούμε το background του frame από αυτό του μοντέλου. Τα pixel με αποτέλεσμα κοντά στο 0 αντιμετωπίζονται σαν φόντο και αυτά με μεγαλύτερη τιμή σαν προσκήνιο. Έτσι λοιπόν, όταν έχουμε το μοντέλο του background ο αλγόριθμος είναι απλός και εύκολος στην εισαγωγή του. Το να αποκτήσουμε όμως το μοντέλο του φόντου είναι αρκετά περίπλοκο. Η πιο εύκολη προσέγγιση είναι να στήσουμε την κάμερα, να αδειάσουμε τη σκηνή από κινούμενα αντικείμενα και να πάρουμε έτσι ένα στιγμιότυπο. Κάτι τέτοιο όμως δεν είναι εφαρμόσιμο, σε εξωτερικές και πραγματικές σκηνές καθώς το φόντο μπορεί να αλλάξει με την πάροδο του χρόνου ή να είναι δύσκολο να αδειάσουμε τη σκηνή. Ο φωτισμός μπορεί να αλλάξει απότομα και η κάμερα να μετακινηθεί.

Μια πιο πρακτική προσέγγιση είναι μια που μπορεί να προσαρμοστεί σε ένα ήπια μεταβαλλόμενο φόντο, το οποίο θα περιγράψουμε παρακάτω. Ας υποθέσουμε ένα

χρονικώς μεταβαλλόμενο pixel στη θέση (x,y) σε μια grayscale ακολουθία video. Θα αναφερόμαστε σε αυτή την τιμή ως $V_{x,y}(t)$. Μπορούμε να συμπεριφερθούμε στην τιμή σαν μια τυχαία συνάρτηση του X_t .

$$X_t = V_{x,y}(t)$$

Τώρα υποθέτουμε ότι μπορούμε να μοντελοποιήσουμε την πιθανότητα παρατήρησης της τρέχουσας τιμής του Pixel ως μια μίξη από Gaussian κατανομές. Αυτή η πιθανότητα είναι:

$$P(X_t) = \sum_{i=1}^K \omega_{i,t} \eta(X_t, \mu_{i,t}, \Sigma_{f,t})$$

Εφόσον το φόντο υποθέτουμε ότι είναι στατικό, η αξία των pixel τα οποία αποτελούν το φόντο μπορούν να αναπαρασταθούν με ένα η περισσότερα Gaussians με μικρή μεταβολή εξαιτίας του θορύβου της εικόνας. Περισσότερα από ένα Gaussian είναι πιθανά σε σκηνές όπως δέντρων που κουνιούνται από τον αέρα ή φωτός που αναβοσβήνει. Επιπλέον, στις περισσότερες σκηνές, το φόντο μπορεί να είναι ορατό, συχνότερα απ' ό,τι το προσκείμενο σε συγκεκριμένα pixel, οπότε το Gaussian με το μεγαλύτερο βάρος ω είναι πιθανά το background. Αυτές οι ιδέες μας επιτρέπουν να προσεγγίσουμε το πρόβλημα της αφαίρεσης background με την ακόλουθη μεθοδολογία.

Για κάθε pixel σε ένα καρέ βίντεο:

- Λαμβάνουμε υπόψη μας τις τελευταίες N τιμές που έχει πάρει το συγκεκριμένο pixel.
- Βρίσκουμε τα K Gaussians και τα βάρη που ταιριάζουν καλύτερα σε αυτό το δείγμα από N τιμές χρησιμοποιώντας συγκεκριμένους αλγόριθμους όπως οι K-Means ή Expectation Maximization (EM)
- Επιλέγουμε το Gaussian με το μεγαλύτερο βάρος ω και αποθηκεύουμε το μέσο όρο, σαν την τιμή της εικόνας του φόντου για το συγκεκριμένο pixel.

Στην συνέχεια:

- Αφαιρούμε το φόντο από το καρέ.
- Στην προκύπτουσα εικόνα, οποιαδήποτε τιμή μεγαλύτερη από τρεις τυπικές αποκλίσεις από το μέσο όρο, θεωρείται foreground, και οποιαδήποτε άλλη τιμή θεωρείται φόντο.

Ο αλγόριθμος που προηγήθηκε είναι υπερβολικά απαιτητικός σε υπολογιστική ισχύ και αυτό καθιστά τη χρήση του σε πραγματικό χρόνο αρκετά δύσκολη, ιδίως το βήμα της τοποθέτησης των K Gaussians στον πίνακα δεδομένων του κάθε pixel για κάθε καρέ.

Για την απλούστευση του συστήματος, η εικόνα του φόντου πρέπει να επαναυπολογίζεται μόνο κάθε N frame. Έτσι, για τα περισσότερα χρονικά βήματα, οι τιμές του κάθε pixel απλά συλλέγονται και αποθηκεύονται για μετέπειτα επεξεργασία κάτι που γίνεται μία φορά ανά N frame. Το μειονέκτημα αυτής της προσέγγισης είναι ότι υπάρχει μια χρονική καθυστέρηση, για να μπορέσει ο αλγόριθμος να προσαρμόσει στο φόντο καινούργια ακίνητα αντικείμενα.

Εντοπισμός χαρακτηριστικών σημείων

Το βήμα της κατάτμησης εικόνας μας επιτρέπει να διαχωρίζουμε τα αντικείμενα στο προσκήνιο από το φόντο. Παρόλα αυτά, συνεχίζουμε να δουλεύουμε πάνω σε ολόκληρη την εικόνα και όχι στα επιμέρους σημεία κίνησης τα οποία είναι επιθυμητά για την ανίχνευση ανθρώπινης κίνησης. Το πρόβλημα του υπολογισμού της κίνησης σε μια εικόνα είναι γνωστό ως εντοπισμός της οπτικής ροής της εικόνας.

Υπάρχει μια μεγάλη ποικιλία από τεχνικές για τον υπολογισμό της οπτικής ροής, αλλά η μέθοδος Kanade-Lucas-Tomasi ξεχωρίζει για την απλότητά της, αλλά και την απουσία παραδοχών σχετικά με την υποκείμενη εικόνα. Ένα απλό παράδειγμα του αλγόριθμου αυτού φαίνεται στο σχήμα X. Ο πιο απλοϊκός αλγόριθμος για να επιτύχουμε point feature tracking σε 2 καρέ βίντεο παρουσιάζεται παρακάτω:

1. Διαλέγουμε ένα μικρό παράθυρο, για παράδειγμα 7 pixel σε κάθε κατεύθυνση, γύρω από ένα pixel που παρουσιάζει ενδιαφέρον στο πρώτο καρέ. Αυτό το pixel το ονομάζουμε pixel A.
2. Κάθε pixel κοντά στο A, στο δεύτερο καρέ, το ονομάζουμε pixel B.
3. Αφαιρούμε την τιμή του κάθε pixel στην 7×7 περιοχή γύρω από το pixel A από κάθε pixel της 7×7 περιοχής γύρω από το pixel B. Υψώνουμε στο τετράγωνο το αποτέλεσμα της διαφοράς, και αθροίζουμε αυτές τις 49 τιμές για να παράγουμε την ανομοιότητα για την επιλογή αυτού του pixel B.
4. Το pixel B στο δεύτερο καρέ με τη μικρότερη τιμή ανομοιότητας θεωρείται ότι είναι η νέα θέση του pixel A.

Παρά το γεγονός ότι ο αλγόριθμος αυτός θα μας δώσει μια νέα θέση και ταχύτητα για το pixel A, έχει κάποιες σημαντικές αδυναμίες. Αρχικά, θα είναι αρκετά αργός, αφού θα απαιτεί εκατοντάδες υπολογισμούς σε κάθε επανάληψη και εκατοντάδες επαναλήψεις. Δεύτερον, ο αλγόριθμος θα μας δίνει την θέση και την ταχύτητα του σημείου μόνο στο κοντινότερο ακέραιο pixel. Αν το σημείο έχει κινηθεί ενάμιση pixel θα πρέπει να το υπολογίσουμε, είτε στη θέση 1, είτε στη θέση 2.

Ο αλγόριθμος Kanade-Lucas-Tomasi μειώνει αυτά τα προβλήματα με την επαναλαμβανόμενη χρήση των βαθμίδων χρώματος της εικόνας, για να προβλέψει τη νέα θέση του σημείου, μέχρι αυτή να επιτευχθεί. Επειδή αυτή η προσέγγιση είναι βασισμένη στην σειρά επέκτασης Taylor (Taylor series expansion) δεν κάνει υποθέσεις σχετικά με την εικόνα που βρίσκεται από κάτω. Ένα τελευταίο χαρακτηριστικό του αλγόριθμου Kanade-Lucas-Tomasi, είναι η επιλογή των αρχικών σημείων. Είναι χάσιμο χρόνου να προσπαθήσουμε να ανιχνεύσουμε τα ίχνη όλων των pixel από την αρχική εικόνα μέχρι την εικόνα προορισμού.

Μια πιο χρήσιμη προσέγγιση είναι να ανιχνεύσουμε μόνο τα pixel τα οποία αναπαριστούν έντονα, σαφώς διακρινόμενα σημεία. Γι αυτό το λόγο, όταν διαλέγουμε τα σημεία που θα ανιχνεύσουμε, ταξινομούμε τα pixel έτσι ώστε οι ιδιοτιμές τους να είναι σε φθίνουσα σειρά και διαλέγουμε τις N πρώτες, όπου N ο αριθμός των pixel που θέλουμε να ανιχνεύσουμε.

Ανίχνευση μοντέλου κίνησης

Έστω ένα σετ από κινούμενα σημεία, κάθε ένα με μια θέση και ταχύτητα, ο σκοπός μας στη ανίχνευση μοντέλου κίνησης είναι να αποφασίσουμε αν μια ομάδα από σημεία είναι αντιπροσωπευτική ενός προκαθορισμένου μοντέλου.

Η προσέγγιση της Y. Song για την αποθήκευση του μοντέλου σαν μια συνάρτηση πιθανοτήτων, και η εύρεση μιας επισήμανσης για τα δεδομένα που μεγιστοποιούν την πιθανότητα, είναι μια από της πιο γνωστές και θα την περιγράψουμε στη συνέχεια του κεφαλαίου.

Αρχικά ορίζουμε μια ομάδα από ανθρώπινα μέλη, το καθένα από τα οποία αντιστοιχεί σε ένα χαρακτηριστικό σημείο, το οποίο θα μπορούσαμε να παρακολουθήσουμε. Έστω λοιπόν $S_{body} = \{AK, AA, A\Omega, K, \dots, \Delta\Pi\}$ το σετ των μελών του σώματος, όπου AK ο αριστερός Καρπός Αγκώνας Ωμος, αντίστοιχα, K το

κεφάλι, Π το πόδι κ.ο.κ. Θα αναφερόμαστε σε αυτά σαν το σετ των πιθανών ετικετών.

Από τη στιγμή που το μοντέλο για τον άνθρωπο μαθαίνεται χωρίς επίβλεψη, η πραγματική χαρτογράφηση των μελών του σώματος σε χαρακτηριστικά σημεία είναι απροσδιόριστη και δεν αντιστοιχεί επακριβώς σε μεμονωμένα άκρα. Κάθε μέλος του σώματος έχει ένα διάνυσμα από μετρήσεις, το οποίο αποτελείται από τη θέση και την ταχύτητα, και στο οποίο θα αναφερόμαστε σαν X_{AK} , X_{AA} κτλ. Το μοντέλο κίνησης αποθηκεύεται σαν μια συνάρτηση πιθανοτήτων P , που μπορεί να αξιολογηθεί για ένα δεδομένο σύνολο δεδομένων.

$$P_{S_{body}}(X_{AK}, X_{AA}, X_{AQ}, X_K, \dots, X_{\Delta\Pi})$$

Τώρα, θεωρούμε ένα διάνυσμα από N παρατηρούμενα σημεία:

$$\overline{X} = [X_1, X_2, \dots, X_N]$$

στο οποίο η αντιστοιχία μεταξύ σημείων και ετικετών δεν είναι γνωστή εκ των προτέρων. Θέλουμε να βρούμε την μετάθεση των σημείων που μεγιστοποιούν τη συνάρτηση της πιθανότητας. Με άλλα λόγια, θέλουμε να βρούμε μια επισήμανση $L = [L_1, L_2, \dots, L_N]$ όπου $L_i \in S_{body}$ είναι η ετικέτα του σημείου X_i , που μεγιστοποιεί τη συνάρτηση της πιθανότητας. Θα καθορίσουμε αυτή την βέλτιστη επισήμανση σαν:

$$\overline{L} = \operatorname{argmax} P(\overline{L} | \overline{X})$$

Όπου:

$$P(\overline{L} | \overline{X}) \text{ η πιθανότητα παρατήρησης του } \overline{X} \text{ δεδομένης της επισήμανσης } \overline{L} .$$

Ας υποθέσουμε ότι N ο αριθμός των σημείων που ισούνται με τα μέλη του σώματος, η προσπάθεια για τη λύση της εξίσωσης αυτής με brute force θα είχε ως αποτέλεσμα χρόνο εκτέλεσης $O(e^N)$. Μία τέτοια κακή απόδοση όμως είναι απαγορευτική για ένα πρακτικό αλγόριθμο. Το βασικό επίτευγμα, της προσέγγισης της Song είναι η υπόθεση ότι ορισμένες παράμετροι της συνάρτησης πιθανότητας είναι ανεξάρτητες μεταξύ τους. Με τον καθορισμό μιας τέτοιας δομής κατά την εκμάθηση του μοντέλου, η αποτελεσματικότητα του ανιχνευτή μπορεί να βελτιωθεί πάρα πολύ χρησιμοποιώντας ένα δυναμικό αλγόριθμο προγραμματισμού.

Πριν κατανοήσουμε γιατί μια τέτοια βελτίωση είναι εφικτή, πρέπει πρώτα να καταλάβουμε πως αναπαριστάται το μοντέλο. Όπως περιγράψαμε, το μοντέλο

κίνησης αποθηκεύεται σαν ένα γράφημα στο οποίο κάθε κορυφή αναπαριστάται από ένα χαρακτηριστικό σημείο, και κάθε ακμή αναπαριστά μια στατιστική εξάρτηση ενός χαρακτηριστικού από ένα άλλο. Μια βοηθητική υπόθεση είναι να αναγκάσουμε το γράφημα να έχει μια ειδική μορφή ενός αποσυντιθέμενου τριγωνικού γραφήματος. Εξαιτίας αυτής της δομής του, το γράφημα θα έχει μια κορυφή η οποία εξαρτάται από μόνο 2 κορυφές. Όταν εξαλειφθεί αυτή η κορυφή, θα υπάρξει κάποια άλλη κορυφή η οποία εξαρτάται από δυο άλλες και ούτω καθεξής. Η αλληλουχία της αφαίρεσης κορυφών είναι γνωστή σαν σειρά εξάλειψης του γραφήματος.

Για παράδειγμα, μια συνάρτηση πιθανοτήτων η οποία είναι αποθηκευμένη σε έναν αποσυντιθέμενο τριγωνικό γράφημα, μπορεί να προσεγγιστεί σαν ένα προϊόν ανεξάρτητων συναρτήσεων. Για παράδειγμα, αν η σειρά εξάλειψης των κορυφών ενός γραφήματος είναι A,B,C,D,E,F, η συνάρτηση πιθανότητας μπορεί να μετασχηματιστεί σε:

$$P(A,B,C,D,E,F)=P(A|B,E)P(B|E,F)P(C|E,F)P(D,E,F)$$

Επειδή, η συνάρτηση της πιθανότητας με πολλές παραμέτρους, μπορεί πλέον να αντιπροσωπευτεί από ένα προϊόν μικρότερων εξαρτημένων συναρτήσεων, ένας αλγόριθμος δυναμικού προγραμματισμού είναι εφικτός, και μπορεί να ψάξει για όλες τις πιθανές επισημάνσεις πολύ πιο βολικά. Κάθε εξαρτημένη συνάρτηση, μπορεί να ανιχνευθεί ατομικά στο $O(N^3)$ προκειμένου να εντοπίσει την ετικέτα που την μεγιστοποιεί για κάθε δυνατό ζεύγος ετικετών.

Επομένως, για ένα μοντέλο με M κορυφές, ο συνολικός χρόνος εκτέλεσης προκειμένου να μεγιστοποιηθεί η συνάρτηση της πιθανότητας είναι $O(M N^3)$, το οποίο αποτελεί μια πολύ σημαντική βελτίωση από το $O(e^N)$ που προκύπτει από την τεχνική brute force. Με αυτό τον τρόπο, μπορούμε να βρούμε την ετικέτα L που μεγιστοποιεί την συνάρτηση πιθανότητας P . Για ένα δοσμένο X , αυτή η διαδικασία μας δίνει δυο χρήσιμες πληροφορίες.

Αρχικά την αριθμητική τιμή της συνάρτησης πιθανοτήτων και δεύτερον την βέλτιστη επισήμανση των χαρακτηριστικών σημείων. Προκειμένου να ολοκληρώσουμε την διαδικασία της ανίχνευσης, πρέπει να διαλέξουμε ένα κατώφλι για την τιμή της συνάρτησης πιθανοτήτων. Αν η τιμή είναι μεγαλύτερη από το κατώφλι, θεωρούμε ότι τα σημεία που δώσαμε σαν είσοδο ταιριάζουν στο μοντέλο, ενώ αν είναι μικρότερη λέμε ότι δεν υπήρξε ταύτιση. Η τιμή του κατωφλιού διαλέγεται εμπειρικά σαν συνάρτηση των λάθος ανιχνεύσεων και των ανιχνεύσεων που δεν βρέθηκαν.

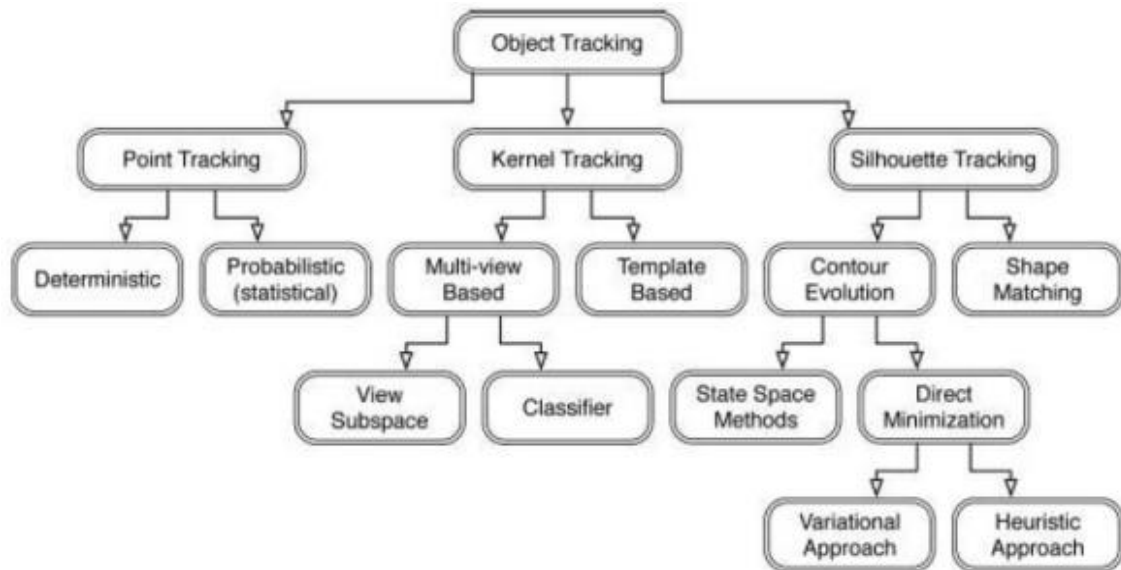
ΚΕΦΑΛΑΙΟ 4: ΕΝΤΟΠΙΣΜΟΣ ΚΙΝΟΥΜΕΝΗΣ ΟΝΤΟΤΗΤΑΣ

4.1 Εισαγωγή

Μετά την διαδικασία της κατάτμησης η οποία είχε σαν αποτέλεσμα τον εντοπισμό και την αναγνώριση του αντικειμένου, το οποίο βρίσκεται υπό μελέτη, το επόμενο βήμα είναι η παρακολούθηση του αντικειμένου καθ' όλη την διάρκεια της κίνησης του. Δεν θα είχε νόημα απλώς μια αναγνώριση του αντικειμένου ενδιαφέροντος αν αυτή δεν συνδυαζόταν και με μία παρακολούθηση της τροχιάς της κίνησης του, αφού αυτό που πραγματικά μας ενδιαφέρει δεν είναι απλώς ο εντοπισμός του αντικειμένου αλλά περισσότερο η συμπεριφορά του στο χώρο δραστηριοποίησης του και η οποία προφανώς είναι απόρροια της κίνησής του.

Στόχος της διαδικασίας της παρακολούθησης αντικειμένου είναι η καταγραφή της τροχιάς του καθώς αυτό αλλάζει θέση σε κάθε καρέ της ακολουθίας βίντεο. Τυπικά, η παρακολούθηση αυτή περιλαμβάνει ταίριασμα των σημείων του αντικειμένου σε διαδοχικά καρέ χρησιμοποιώντας γνωρίσματα όπως σημεία, γραμμές, ή οντότητες. Πιο συγκεκριμένα, έχει να κάνει με την εύρεση συναφών σχέσεων μεταξύ χαρακτηριστικών της εικόνας σε διαδοχικά καρέ, όπως για παράδειγμα το χρώμα, η υφή, η ταχύτητα και η θέση. Η παρακολούθηση τροχιάς των αντικειμένων είναι σαφώς ένα ιδιαίτερα σημαντικό και βασικό κομμάτι στην ανάλυση ανθρώπινης κίνησης.

Οι τεχνικές που χρησιμοποιούνται για την παρακολούθηση της κίνησης ποικίλουν ανάλογα την εφαρμογή και η χρήση τους εξαρτάται από πολλούς παράγοντες όπως το μέγεθος του αντικειμένου, το περιβάλλον στο οποίο δραστηριοποιείτε, τις απαιτήσεις στην ποιότητα της παρακολούθησης κ.α..

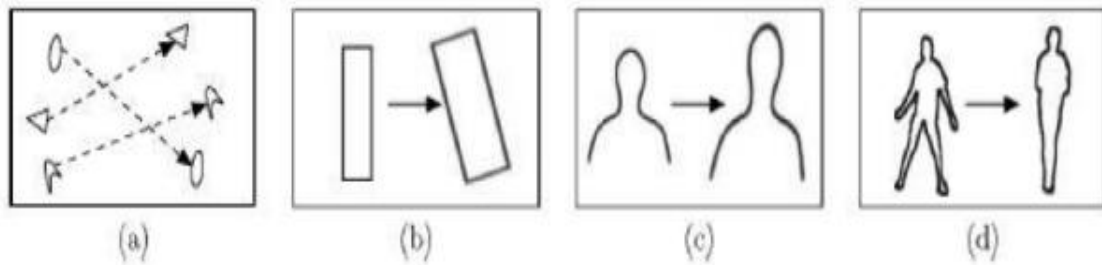


Εικόνα 4.1: Τεχνικές παρατήρησης κίνησης αντικειμένου.

Από το παραπάνω σχήμα τρεις είναι οι βασικές κατηγορίες επίλυσης του προβλήματος, κάθε μία εκ των οποίων, μπορεί να προσεγγιστεί από διαφορετική μέθοδο. Οι κατηγορίες αυτές είναι:

1. Point tracking. Σε αυτή την κατηγορία το αντικείμενο που είναι προς παρακολούθηση αντιπροσωπεύεται και αναπαριστάτε από ένα σημείο ή ένα πλήθος σημείων. Η σύνδεση των σημείων αυτών βασίζεται στην προηγούμενη κατάσταση του αντικειμένου η οποία περιέχει την θέση του και την κίνησή του. Ένα παράδειγμα εφαρμογής αυτής της μεθόδου φαίνεται στην παρακάτω εικόνα (a).
2. Kernel tracking. Οι τεχνικές Kernel βασίζονται κυρίως στην εμφάνιση και στο σχήμα του αντικειμένου. Το αντικείμενο που εντοπίζεται εισάγεται εντός περιθωρίου το σχήμα του οποίου είναι κάποιο ορθογώνιο ή έλλειψη. Το αντικείμενο παρακολουθείται υπολογίζοντας την κίνηση του περιθωρίου σε κάθε ένα από τα συνεχόμενα καρέ. Αυτή η κίνηση είναι συνήθως σε μορφή κάποιου παραμετρικού μετασχηματισμού όπως π.χ. περιστροφή. Στην παρακάτω εικόνα (b) φαίνεται ένα τέτοιο παράδειγμα.
3. Silhouette tracking. Στην κατηγορία αυτή η παρακολούθηση γίνεται υπολογίζοντας την περιοχή που καταλαμβάνει το αντικείμενο σε κάθε ένα καρέ της ακολουθίας βίντεο. Η πληροφορία που χρησιμοποιείται βρίσκεται εντός του αντικειμένου και αντιπροσωπεύει την πυκνότητα ή το σχήμα του αντικειμένου και τα οποία είναι χαρακτηριστικά που συνήθως χρειάζονται προκειμένου να

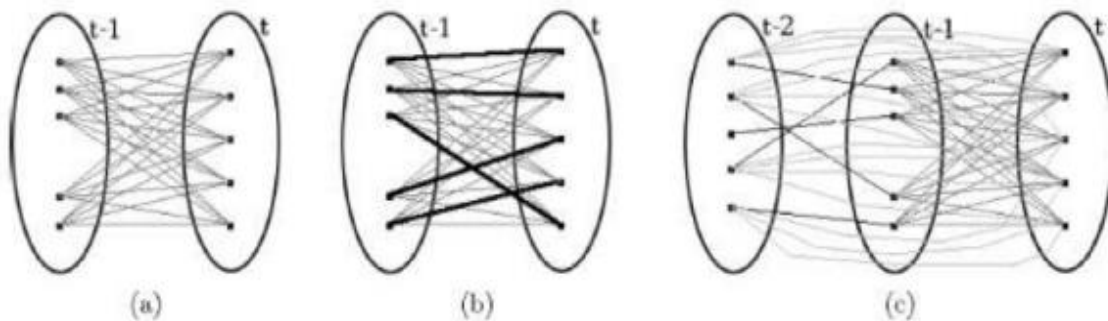
εξαχθεί ο χάρτης ακμών. Παραδείγματα φαίνονται στην παρακάτω εικόνα (c)-(d).



Εικόνα 4.2: Παραδείγματα εφαρμογής διαφορετικών μεθόδων για την παρακολούθηση κινούμενου αντικειμένου.

4.2 Point tracking – Deterministic method

Η Ντετερμινιστική μέθοδος για την παρακολούθηση αντικειμένου κάνει χρήση εξισώσεων κόστους στις οποίες γίνεται προσπάθεια ελαχιστοποίησής τους καθώς γίνεται σύγκριση του αντικειμένου που παρατηρείται στο καρέ $t-1$ με το αντικείμενο που παρατηρείται στο καρέ t . Λύση σε αυτό το πρόβλημα πετυχαίνεται με την ακριβή αντιστοίχιση ένα προς ένα σημείων όπως αυτή φαίνεται στην παρακάτω εικόνα (b).



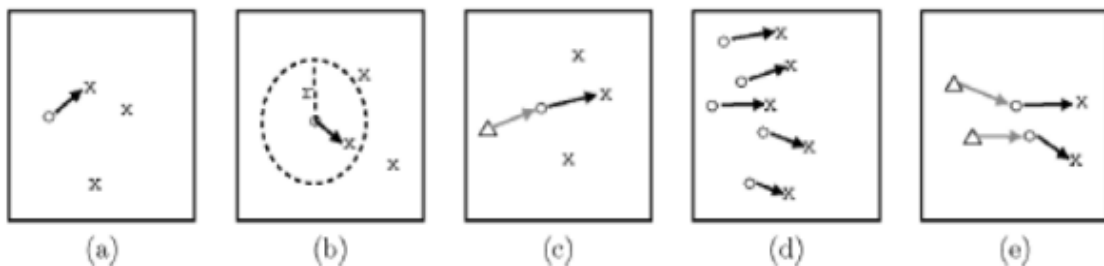
Εικόνα 4.3: Αντιστοίχιση σημείων δύο διαφορετικών καρέ.

Η συνάρτηση του κόστους διασύνδεσης συνήθως ορίζεται κάνοντας χρήση συνδυασμού των παρακάτω παραδοχών.

1. Εγγύτητα. Ένα αντικείμενο μπορούμε να δεχτούμε ότι σε διαφορά ενός καρέ δεν μπορεί να έχει παραμορφωθεί σε τέτοιο βαθμό ώστε η απόσταση μεταξύ των σημείων του να έχει μεταβληθεί αρκετά. Αυτό σημαίνει ότι γειτονικά σημεία

θα πρέπει να παραμένουν γειτονικά με την ίδια απόσταση μεταξύ τους. Αυτή η παραδοχή φαίνεται στην παρακάτω εικόνα (a).

2. Μέγιστη ταχύτητα. Η ταχύτητα ενός αντικειμένου είναι περιορισμένη επομένως και η θέση στην οποία θα βρίσκονται τα σημεία του σε επόμενο καρέ θα πρέπει να είναι περιορισμένη και σίγουρα εντός μιας συγκεκριμένης ακτίνας. Αυτό μπορεί να φανεί στην παρακάτω εικόνα (b).
3. Μεταβολή κίνησης. Αυτό δηλώνει ότι δεν μπορεί να υπάρξει δραματική μεταβολή στα χαρακτηριστικά της κίνησης, δηλαδή τόσο στην τιμή της ταχύτητάς του αντικειμένου όσο και στην κατεύθυνση της κίνησης του. Χαρακτηριστικό αυτής της παραδοχής φαίνεται στην παρακάτω εικόνα (c).
4. Ομοιότητα κίνησης. Σύμφωνα με αυτή την παραδοχή δεχόμαστε ότι γειτονικά σημεία δεν μπορεί να κινούνται διαφορετικά, η κίνηση τους πρέπει να είναι όμοια όπως φαίνεται στην παρακάτω εικόνα(d).
5. Ακαμψία. Δεχόμαστε ότι τα αντικείμενα στον τρισδιάστατο κόσμο είναι άκαμπτα επομένως η απόσταση σημείων του αντικειμένου σε διαφορετικά καρέ θα πρέπει να παραμένει η ίδια, κάτι το οποίο φαίνεται στην παρακάτω εικόνα (e).

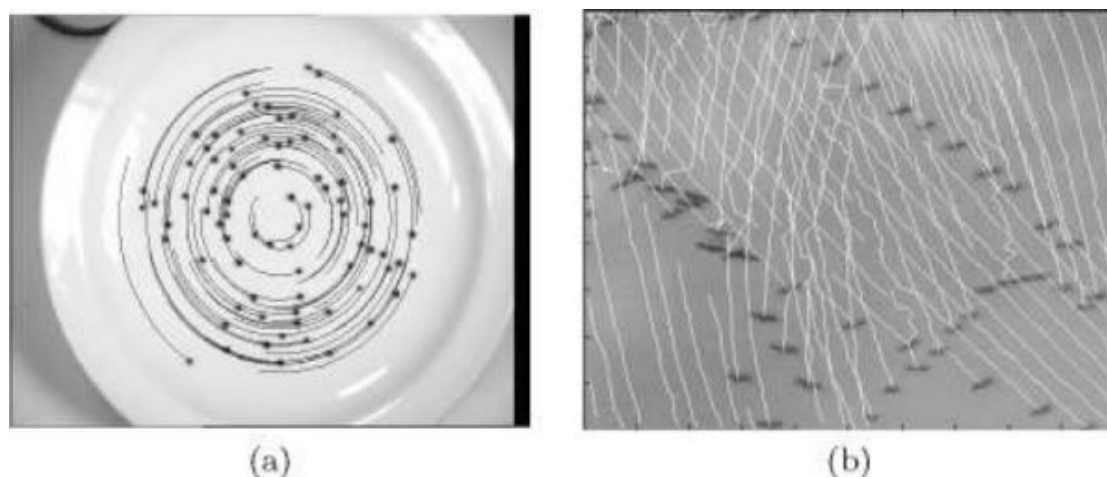


Εικόνα 4.4: Επεξηγηματική αναπαράσταση παραδοχών.

Από τα παραπάνω αντιλαμβανόμαστε ότι κάνοντας χρήση αυτής της μεθόδου ουσιαστικά εγκαταλείπουμε την ιδέα της παρακολούθησης ολόκληρου του αντικειμένου και επικεντρώνουμε την προσοχή μας σε ορισμένα σημεία αυτού. Τα πλεονεκτήματα της είναι ότι μειώνει σημαντικά την υπολογιστική πολυπλοκότητα, ενώ παράλληλα είναι σθεναρή σε μερική επικάλυψη των αντικειμένων με τη προϋπόθεση ότι κάποια από τα γνωρίσματα παραμένουν ορατά.

Η μέθοδος αυτή χωρίζεται σε δύο μέρη, στην εξαγωγή των γνωρισμάτων και στην αντιστοίχιση τους. Χαμηλότερου επιπέδου χαρακτηριστικά, όπως σημεία, είναι ευκολότερο να εξαχθούν αλλά πολύ πιο δύσκολο να παρακολουθηθούν αφού είναι πιο δύσκολο να προσδιοριστεί μια ένα-προς-ένα αντιστοιχία μεταξύ τους.

Αντίστροφα, υψηλότερου επιπέδου χαρακτηριστικά, όπως γραμμές ή οντότητες, είναι ευκολότερο να εντοπιστούν αλλά πιο δύσκολο να εξαχθούν. Φαίνεται λοιπόν πως υπάρχει μια αλληλεξάρτηση ανάμεσα στη πολυπλοκότητα των χαρακτηριστικών και την αποδοτικότητα της παρακολούθησης. Στα παρακάτω παραδείγματα φαίνονται τα αποτελέσματα κάνοντας χρήση αυτής της μεθόδου.



Εικόνα 4.5: Παραδείγματα εφαρμογών με χρήση του deterministic method.

4.3 Point tracking – Statistical method

Χαρακτηριστικό παράδειγμα αυτής της κατηγορίας είναι τα φίλτρα Kalman τα οποία συνήθως χρησιμοποιούνται ώστε να υπολογιστεί η κατάσταση του γραμμικού συστήματος υποθέτοντας ότι η κατάσταση αυτή έχει κανονική κατανομή κατά Gauss. Τα φίλτρα Kalman είναι μια τεχνική στοχαστικής πρόγνωσης και εκτίμησης των σωστών παραμέτρων ενός μοντέλου σε ένα δυναμικό περιβάλλον.

Σε αντίθεση με την ντετερμινιστική τεχνική, όπου οι παράμετροι του συστήματος θεωρούνται άγνωστοι μεν αλλά σταθεροί, στα φίλτρα Kalman οι παράμετροι θεωρείται πως εξαρτώνται κάθε στιγμή από την κατάσταση του συστήματος και επομένως αλλάζουν δυναμικά με κάθε μεταβολή της κατάστασής του. Έτσι, η τεχνική αυτή επανεκτιμά με κάθε νέα μέτρηση την κατάσταση του δυναμικού συστήματος, βασιζόμενη σε παρατηρήσεις που έχουν γίνει στις προηγούμενες χρονικές στιγμές και επαναπροσδιορίζει τις παραμέτρους του συστήματος.

Τα φίλτρα Kalman χρησιμοποιούνται σε εφαρμογές όπου απαιτείται η χρήση ενός δυναμικού μοντέλου του οποίου οι βέλτιστοι παράμετροι πρέπει να προσδιορίζονται σε διακριτά χρονικά διαστήματα. Η υλοποίησή τους πραγματοποιείται σε δύο στάδια,

της πρόβλεψης και της διόρθωσης. Το βήμα της πρόβλεψης χρησιμοποιεί μία αρχική κατάσταση ώστε να γίνει πρόβλεψη μιας νέας κατάστασης των μεταβλητών.

4.4 Kernel tracking

Η υπολογιστική πολυπλοκότητα ενός παρακολουθητή είναι ένα πολύ σημαντικό μέγεθος για τις περισσότερες εφαρμογές μηχανικής όρασης καθώς μόνο ένα μικρό ποσοστό των πόρων του συστήματος αφιερώνεται στην παρακολούθηση. Το υπόλοιπο, προορίζεται συνήθως για τα στάδια προεπεξεργασίας και για εργασίες υψηλότερου επιπέδου, όπως αναγνώριση και ερμηνεία τροχιάς.

Στη συνέχεια θα παρουσιάσουμε μια προσέγγιση για παρακολούθηση αντικειμένων σε πραγματικό χρόνο που βασίζεται σε οπτικά χαρακτηριστικά, όπως το χρώμα και η υφή, των οποίων η στατιστική κατανομή χαρακτηρίζει το αντικείμενο ενδιαφέροντος. Η προτεινόμενη παρακολούθηση είναι κατάλληλη για πολλές κατηγορίες αντικειμένων με διαφορετικά μοτίβα χρώματος ή/και υφής ενώ παράλληλα είναι σθεναρό σε μερικές επικαλύψεις και κινήσεις της κάμερας.

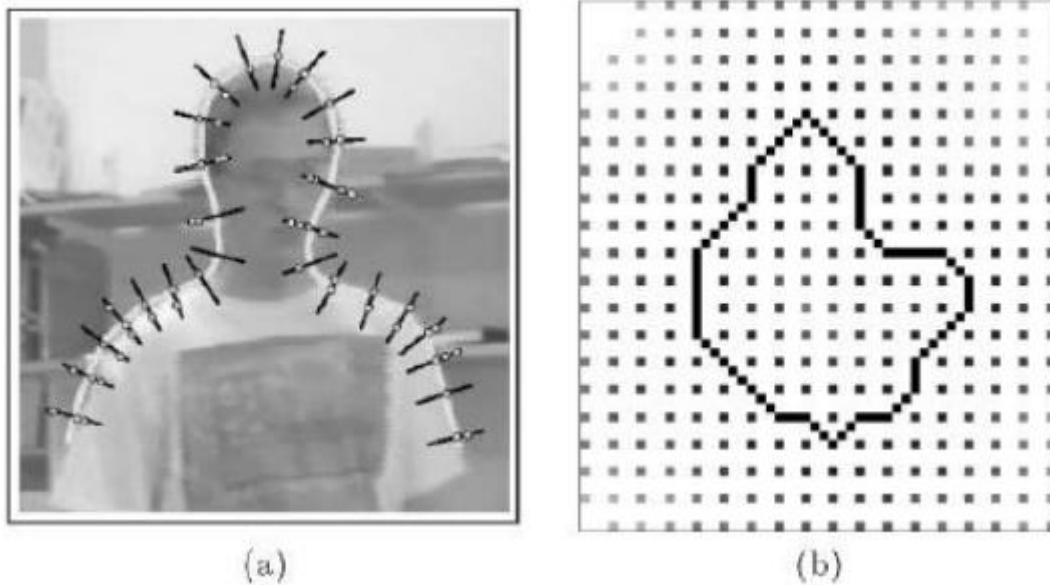
Οι επαναλήψεις ολίσθησης μέσου όρου χρησιμοποιούνται ώστε να βρεθεί η υποψήφια περιοχή που μοιάζει περισσότερο με το δοσμένο μοντέλο. Η ομοιότητα αυτό εκφράζεται από ένα μετρικό που βασίζεται στο συντελεστή Bhattacharyya.

4.5 Silhouette tracking

Τα αντικείμενα μπορεί να έχουν πολύπλοκα σχήματα, για παράδειγμα τα χέρια, το κεφάλι και οι ώμοι δεν μπορούν να περιγραφούν καλά από απλά γεωμετρικά σχήματα, κάτι το οποίο φαίνεται από στην εικόνα 4.6.

Γενικά το σχήμα είναι από τους πλέον σημαντικούς παραμέτρους για την αναγνώριση των αντικειμένων. Ο όρος σχήμα μπορεί να αναφέρεται είτε στο εξωτερικό δισδιάστατο (ή τρισδιάστατο) περίγραμμα του αντικειμένου, ή στη γενική του όψη και διάταξη. Γεωμετρικά, δύο αντικείμενα έχουν το ίδιο σχήμα αν το ένα μπορεί να μετατραπεί στο άλλο με χρήση κάποιου συνδυασμού από μετατοπίσεις, περιστροφές και ομοιόμορφες αλλαγές κλίμακας, με άλλα λόγια το σχήμα είναι όλη εκείνη η γεωμετρική πληροφορία για το αντικείμενο που παραμένει αμετάβλητη από την θέση, την κλίμακα και στην περιστροφή του αντικειμένου. Επιπλέον οι άνθρωποι αναγνωρίζουν δύο ίδια σχήματα ακόμα και αν το ένα αποτελεί κατοπτρισμού του άλλου.

Όσο αναφορά την όραση των υπολογιστών το πρόβλημα μπορεί να λυθεί είτε υπολογίζοντας το περίγραμμα των αντικειμένων (boundary based) είτε μελετώντας την περιοχή που περικλείεται εντός του περιγράμματος (region based).



Εικόνα 4.6: Απολογισμός μεθόδου silhouette tracking.

4.6 Φίλτρο Kalman

Γενικά

Το φίλτρο Kalman είναι ένα ιδιαίτερα διαδεδομένο και εύχρηστο φίλτρο που χρησιμοποιείται για απαλλαγή από τις θορυβώδεις μετρήσεις. Ο αλγόριθμος αυτός, πραγματοποιεί πρόβλεψη της θέσης του στόχου και χρησιμοποιεί τις μετρήσεις για να διορθώσει την πρόβλεψη αυτή. Επομένως έχει την δυνατότητα να παρέχει εκτιμήσεις με σφάλμα μικρότερο του σφάλματος μέτρησης. Αποτελεί μια επέκταση της μεθόδου ελαχίστων τετραγώνων και πρόκειται ουσιαστικά για ένα σύνολο μαθηματικών εξισώσεων, οι οποίες εισάγουν έναν εκτιμητή πρόβλεψης – διόρθωσης, ο οποίος είναι βέλτιστος από την άποψη, ότι ελαχιστοποιεί την συνδιακύμανση σφάλματος εκτίμησης, όταν η δυναμική κατάσταση του στόχου και ο θόρυβος μέτρησης έχουν μοντελοποιηθεί με ακρίβεια. Σε γραμμικά περιβάλλοντα ο αλγόριθμος του Kalman αποτελεί τη βέλτιστη λύση.

Η εκτεταμένη χρήση του φίλτρου Kalman οφείλεται στο ότι συγκεντρώνει πολλά πλεονεκτήματα, τα οποία βοηθούν στην καλύτερη εκτίμηση της δυναμικής κατάστασης του συστήματος. Κατά πρώτο λόγο, το κέρδος του φίλτρου ανανεώνεται

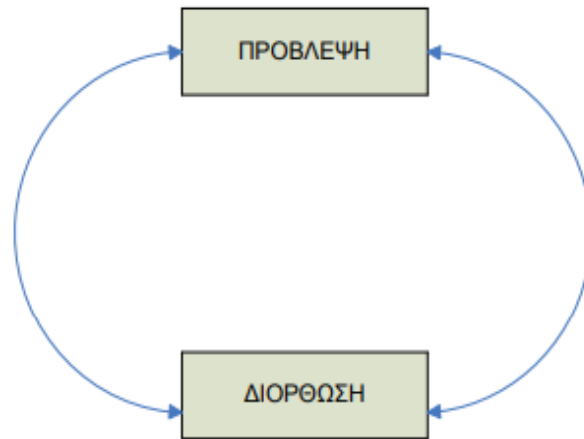
με κάθε νέα μέτρηση. Το γεγονός αυτό διευκολύνει την ανίχνευση σε συστήματα με χρονομεταβλητές δυναμικές καταστάσεις ή μετρήσεις, οι διαδικασίες των οποίων μεταβάλλονται με τον χρόνο. Επιπλέον, η συνδιακύμανση του θορύβου μέτρησης μπορεί να αυξομειωθεί, έτσι ώστε η κάθε μέτρηση να αποκτήσει διαφορετικό βάρος ανάλογα με την απόσταση από τον στόχο, την ισχύ του σήματος ή οποιαδήποτε άλλη ενδεχόμενη ένδειξη της ποιότητας της μέτρησης. Αυτό είναι ιδιαίτερα σημαντικό για περιπτώσεις όπου το διάλυμα μετρήσεων περιέχει μετρήσεις από διαφορετικού τύπου αισθητήρες γιατί δίνεται η δυνατότητα στον χρήστη να κάνει μία αρχικοποίηση ανάλογα με την κατά περίπτωση αναμενόμενη αξιοπιστία του εκάστοτε αισθητήρα. Άλλο πλεονέκτημα του φίλτρου είναι η αποδοτικότητά του σε υπολογιστικά συστήματα περιορισμένης μνήμης. Όλη η πληροφορία για τις αρχικές καταστάσεις, τις μετρήσεις και την κίνηση περιέχεται στον πίνακα συνδιακύμανσης σφάλματος P_k . Σημαντικό είναι επίσης το ότι δεν απαιτεί περιοδικές ενημερώσεις χρόνου ή μετρήσεων. Στην γενική περίπτωση, το φίλτρο ενημερώνει τον χρόνο με υψηλό ρυθμό και όταν κάποια μέτρηση παρέχεται από κάποιον αισθητήρα, η ενημέρωση χρόνου διακόπτεται προσωρινά, προκειμένου να πραγματοποιηθεί η ενημέρωση της μέτρησης. Επομένως έχει την δυνατότητα να διαχειρίζεται δεδομένα από ανόμοιους αισθητήρες, οι οποίοι δεν είναι καν συγχρονισμένοι.

Αλγόριθμος

Το φίλτρο Kalman αποτελεί έναν αναδρομικό αλγόριθμο εκτίμησης της κατάστασης ενός συστήματος (recursive state-estimation). Ο αλγόριθμος του φίλτρου πραγματοποιεί την εκτίμηση, χρησιμοποιώντας μια μορφή ελέγχου ανάδρασης (feedback control). Αρχικά το φίλτρο εκτιμά την κατάσταση του συστήματος για κάποια στιγμή και στην συνέχεια αποκτά ανάδραση από τις μετρήσεις.

Επομένως οι εξισώσεις του φίλτρου Kalman διακρίνονται σε δύο ομάδες, στις εξισώσεις ενημέρωσης χρόνου (time update) και στις εξισώσεις ενημέρωσης μέτρησης (measurement update). Οι εξισώσεις της πρώτης ομάδας είναι υπεύθυνες για την χρονική προώθηση της εκτίμησης της παρούσας κατάστασης και την a priori εκτίμηση της συνδιακύμανσης σφάλματος για το επόμενο βήμα. Οι εξισώσεις της άλλης ομάδας είναι υπεύθυνες για την ανάδραση, όπως για παράδειγμα την εισαγωγή νέων μετρήσεων στην a priori εκτίμηση, ώστε να παραχθεί βελτιωμένη a posteriori εκτίμηση.

Συνοπτικά μπορεί να ειπωθεί ότι οι εξισώσεις ενημέρωσης χρόνου είναι οι εξισώσεις πρόβλεψης (predictor equations), ενώ οι εξισώσεις ενημέρωσης μέτρησης είναι οι εξισώσεις διόρθωσης (corrector equations). Επομένως το φίλτρο Kalman μπορεί να θεωρηθεί ως ένας αλγόριθμος πρόβλεψης – διόρθωσης.



Εικόνα 4.7: Αλγόριθμος Kalman.

Οι εξισώσεις ενημέρωσης χρόνου του φίλτρου είναι οι εξής:

$$\hat{x}_k^- = A\hat{x}_{k-1} + Du_k$$

$$P_k^- = AP_{k-1}A^T + Q$$

και οι εξισώσεις ενημέρωσης μέτρησης είναι:

$$K_k = P_k^- H^T (HP_k^- H^T + R)^{-1}$$

$$\hat{x}_k = \hat{x}_k^- + K_k (z_k - H\hat{x}_k^-)$$

$$P_k = (I - K_k H)P_k^-$$

Αρχικά υπολογίζεται η a priori εκτίμηση της κατάστασης, με βάση την a posteriori εκτίμηση της προηγούμενης σάρωσης.

Αντίστοιχα υπολογίζεται και η a priori εκτίμηση του πίνακα συνδιακύμανσης σφάλματος από την a posteriori εκτίμηση του προηγούμενου βήματος. Αυτό είναι το στάδιο της πρόβλεψης.

Στη συνέχεια ακολουθεί η φάση της διόρθωσης των μεγεθών αυτών, με χρήση των νέων μετρήσεων. Υπολογίζεται η νέα τιμή του κέρδους K και πραγματοποιείται η a posteriori εκτίμηση της κατάστασης με βάση την προβλεπόμενη τιμή, το κέρδος και το υπόλοιπο μέτρησης.

Τέλος λαμβάνει χώρα η εκτίμηση της a posteriori τιμής του πίνακα συνδιακύμανσης σφάλματος. Οι a posteriori τιμές χρησιμοποιούνται στο επόμενο βήμα εκ νέου για ενημέρωση των προβλέψεων.

ΚΕΦΑΛΑΙΟ 5: ΥΛΟΠΟΙΗΣΗ

Σε αυτό το σημείο θα περάσουμε στο πρακτικό μέρος της πτυχιακής παρουσιάζοντας το εργαλείο με το οποίο θα υλοποιήσουμε την πτυχιακή μας. Στη συνέχεια θα δούμε ένα διάγραμμα, σύμφωνα με το οποίο θα πραγματοποιήσουμε το πρόγραμμα, το οποίο θα ανιχνεύει ανθρώπινη κίνηση στο χώρο.

5.1 Εισαγωγή στο Matlab

Το Matlab είναι ένα εργαλείο το οποίο προσφέρει ένα διαδραστικό προγραμματιστικό περιβάλλον στον χρήστη και χρησιμοποιείται σε ένα μεγάλο αριθμό εφαρμογών. Στους χρήστες του Matlab προσφέρεται μια υκαι επίλυση σύνθετων μαθηματικών, και όχι μόνο, προβλημάτων.

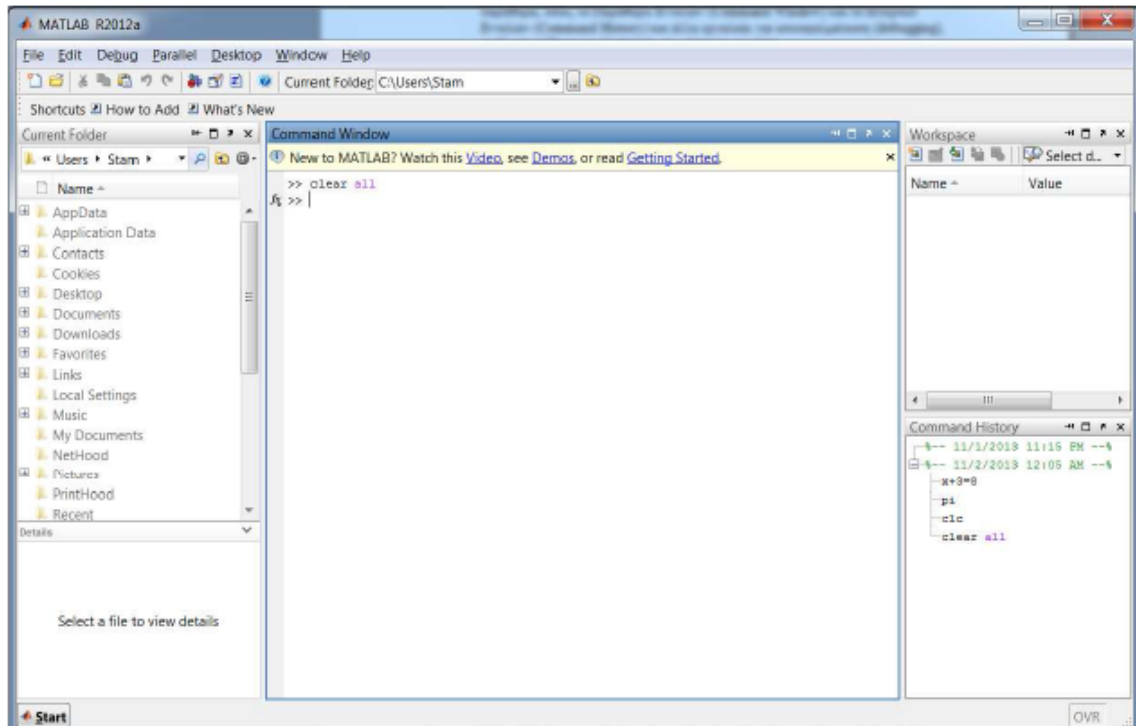
Το όνομά του προέρχεται από τις λέξεις Matrix Laboratory (εργαστήριο πινάκων) επειδή για τη λειτουργία του χρησιμοποιεί εξολοκλήρου πίνακες. Το Matlab εκτός των άλλων, επιτρέπει τον εύκολο χειρισμό πινάκων, τη γραφική απεικόνιση (plotting) συναρτήσεων και δεδομένων, την υλοποίηση αλγορίθμων και τη συνεργασία και διαλειτουργικότητα με προγράμματα γραμμένα σε άλλες γλώσσες προγραμματισμού. Επιπροσθέτως επειδή το Matlab μπορεί να χρησιμοποιηθεί σε πολλά επιστημονικά πεδία(επεξεργασία σήματος, νευρωνικά δίκτυα, συστήματα ελέγχου κ.τ.λ.), είναι δυνατή η ενσωμάτωση πρόσθετων πακέτων, που ονομάζονται toolboxes, και δίνουν τη δυνατότητα χρήσης εξεζητημένων συναρτήσεων.

Τα βασικά συστατικά του Matlab είναι:

- το Περιβάλλον Ανάπτυξης και τα αντίστοιχα εργαλεία του: περιλαμβάνει διάφορα παράθυρα, όπως το Παράθυρο Εντολών (Command Window) και το Ιστορικό Εντολών (Command History) και άλλα εργαλεία για αποσφαλμάτωση (debugging), ανάλυση κώδικα και πλοήγηση στο σύστημα αρχείων.
- η βιβλιοθήκη μαθηματικών συναρτήσεων: ένα από τα πιο σημαντικά συστατικά του Matlab, το οποίο προσφέρει μεγάλο εύρος αριθμητικών συναρτήσεων, από τις πιο απλές μέχρι τις πιο περίπλοκες.
- η γλώσσα προγραμματισμού: μια υψηλού επιπέδου προγραμματιστική γλώσσα με δομές δεδομένων, συναρτήσεις, εντολές ελέγχου ροής, εντολές εισόδου/εξόδου και στοιχεία από αντικειμενοστραφείς γλώσσες προγραμματισμού.

- τα γραφικά συστατικά: το Matlab παρέχει μια πληθώρα δυνατοτήτων απεικόνισης διανυσμάτων, πινάκων και γραφημάτων στις 2 και 3 διαστάσεις.

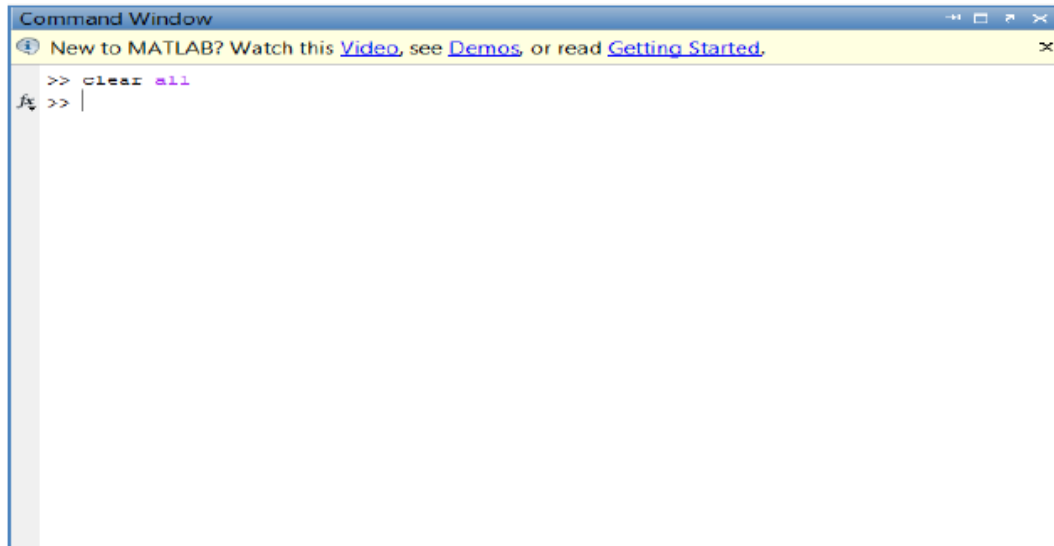
Στο σημείο αυτό θα περιγράψουμε το περιβάλλον ανάπτυξης (Development Environment) του Matlab.



Εικόνα 5.1: Matlab Desktop (Επιφάνεια Εργασίας Matlab).

Το MATLAB Desktop (Επιφάνεια Εργασίας MATLAB) αποτελεί τη βάση του Περιβάλλοντος Ανάπτυξης του συγκεκριμένου εργαλείου. Παρέχει στο χρήστη ένα προγραμματιστικό περιβάλλον στο οποίο του δίνεται η δυνατότητα να εκτελέσει εντολές, να αστικοποιήσει δεδομένα και αποτελέσματα, να δημιουργήσει και να τροποποιήσει αρχεία mfiles (απλά αρχεία κειμένου με κατάληξη .m που περιέχουν εντολές MATLAB) και να λάβει στοχευόμενη βοήθεια.

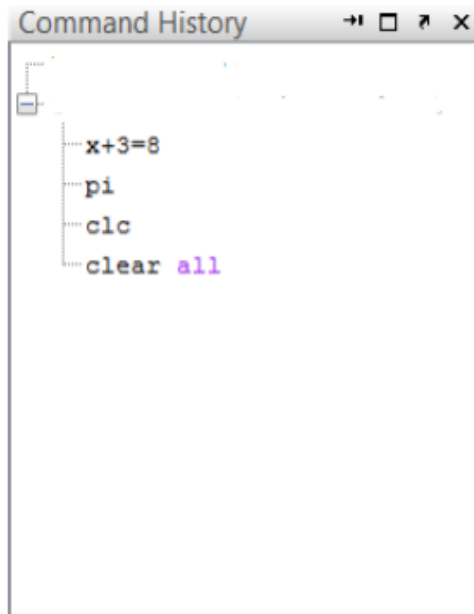
Αποτελείται από έναν αριθμό άλλων παραθύρων των οποίων η χρησιμότητα θα αναλυθεί στην συνέχεια.



Εικόνα 5.2: Παράθυρο Εντολών (Command Window).

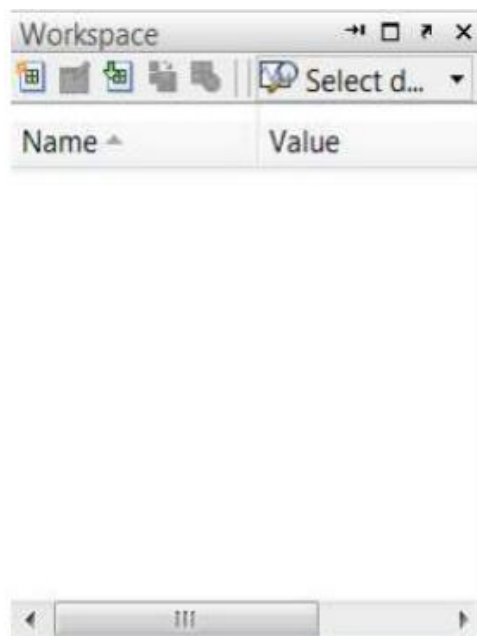
Στο παράθυρο εντολών μπορούμε να εκτελέσουμε διάφορες εντολές σε γλώσσα Matlab, προκειμένου να ολοκληρώσουμε έναν αριθμό εργασιών, όπως για παράδειγμα την δημιουργία και επεξεργασία μεταβλητών (variables), τη δημιουργία, τροποποίηση και κλήση αρχείων m-files, τη διαχείριση δεδομένων και πολλά άλλα. Στο παράθυρο αυτό η εκτέλεση της εντολής πραγματοποιείται με τη χρήση του πλήκτρου «Enter» και τα αποτελέσματα της εντολής γίνονται άμεσα εμφανή κάτω από την εντολή μας.

Να επισημάνουμε σε αυτό το σημείο ότι με τη χρήση του πλήκτρου «Tab» εμφανίζεται μια λίστα βοήθειας με όλες τις δυνατές εντολές, συναρτήσεις και μεταβλητές που αρχίζουν από τους χαρακτήρες που έχουμε εισάγει. Επίσης, είναι δυνατή η επανάκληση προηγούμενα εκτελεσμένων εντολών απλά χρησιμοποιώντας τα Up & Down Arrow Keys (πάνω και κάτω βελάκι).



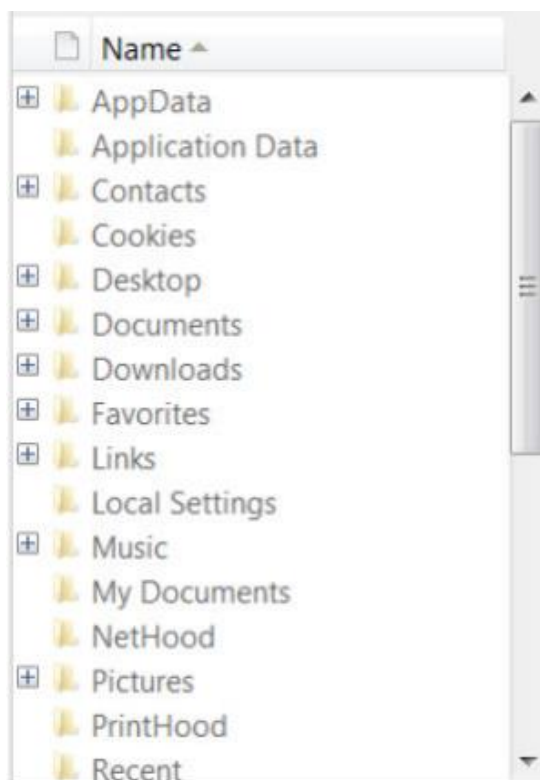
Εικόνα 5.3: Ιστορικό Εντολών (Command History).

Όλες οι εντολές που έχουμε εκτελέσει στο παράθυρο εντολών (Command Window) αποθηκεύονται στο παράθυρο αυτό. Οι εγγραφές αυτές, ταξινομούνται σύμφωνα με την ώρα εκτέλεσής τους. Μπορούμε με τη βοήθεια αυτού του παραθύρου να εκτελέσουμε ξανά εντολές που έχουμε ήδη εισάγει, ή να τις επεξεργαστούμε περισσότερο σέρνοντας τις στο παράθυρο εντολών. Οι εγγραφές του Ιστορικού Εντολών, τέλος, μπορούν να διαγραφούν, να ομαδοποιηθούν, να τροποποιηθούν ή να σωθούν σε ένα m-file.



Εικόνα 5.4: Χώρος εργασίας(Workspace).

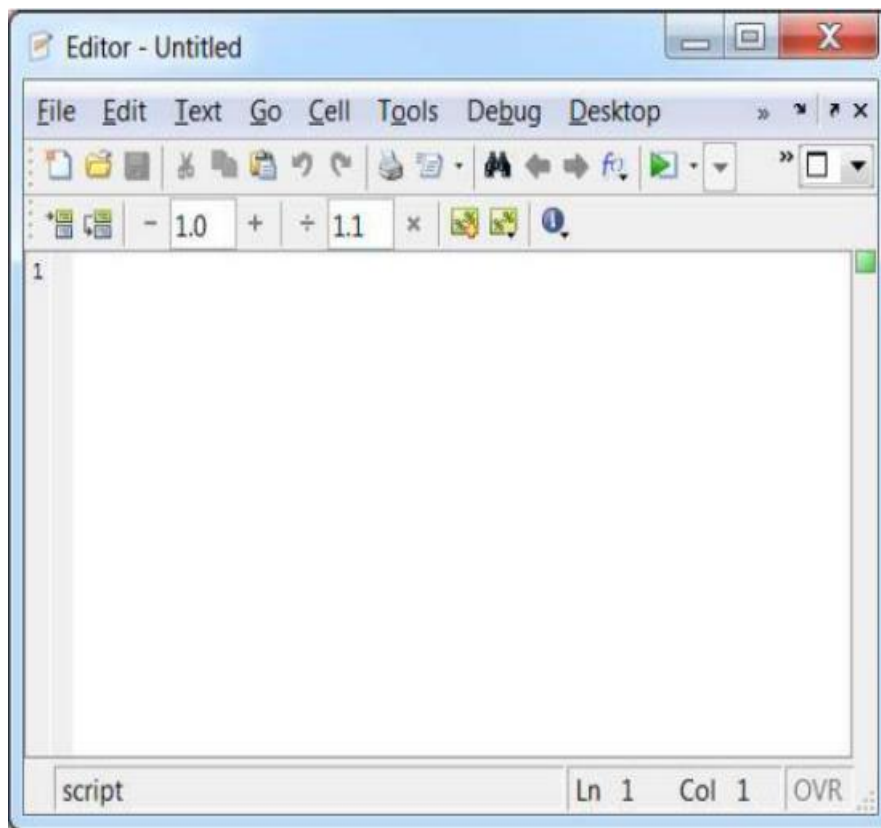
Ο χώρος εργασίας αποτελείται από όλες τις μεταβλητές που έχουν οριστεί στο παράθυρο εντολών και είναι προσβάσιμες από όλα τα προγράμματα. Κάθε μεταβλητή που αποθηκεύεται έχει διάφορες πληροφορίες όπως η τιμή της, το όνομά της, διάφορα στατιστικά στοιχεία και άλλα. Οι πληροφορίες αυτές γίνονται εμφανείς στον χρήστη μέσω του Workspace Browser. Δυνατότητες όπως η διαδραστική ανάγνωση και τροποποίηση των χαρακτηριστικών των αποθηκευμένων μεταβλητών (με διπλόκλικ επάνω στην επιθυμητή μεταβλητή και χρήση του Array Editor) και η γραφική απεικόνισή τους (plotting) είναι μερικές από τις παρεχόμενες από αυτόν ευκολίες. Η συμβολή του, δε, κατά την διαδικασία της αποσφαλμάτωσης (debugging) προγραμμάτων είναι καταλυτική.



Εικόνα 5.5: Τρέχων Κατάλογος (Current Directory).

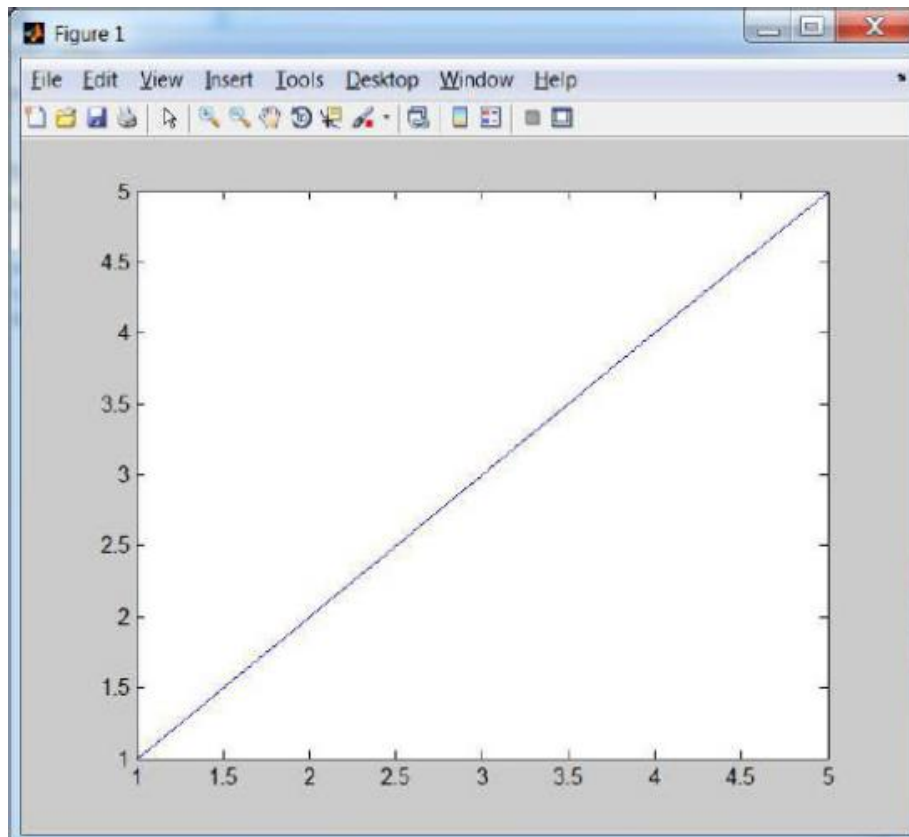
Κάθε λειτουργία η οποία στηρίζεται στην χρήση αρχείων, χρησιμοποιεί τον Τρέχοντα Κατάλογο (Current Directory) ως σημείο αναφοράς. Κάθε αρχείο που επιθυμούμε να εκτελέσουμε ή να χρησιμοποιήσουμε θα πρέπει αναγκαστικά να βρίσκεται μέσα στο συγκεκριμένο φάκελο εκτός και αν εμείς το ορίσουμε διαφορετικά. Στον κατάλογο αυτό παρουσιάζονται τα περιεχόμενα του, επιτρέποντας την περιήγηση του χρήστη στο σύστημα αρχείων του εκάστοτε υπολογιστή, το άνοιγμα (με το κατάλληλο

προκαθορισμένο πρόγραμμα), την αναζήτηση και τροποποίηση σχετικών αρχείων ή καταλόγων καθώς και την διαχείριση m-files.



Εικόνα 5.6: Παράθυρο Εισαγωγής Κειμένου (Editor window).

Στο παράθυρο αυτό μπορούμε να δημιουργήσουμε μια ακολουθία εντολών και συναρτήσεων και να τα αποθηκεύσουμε ως m-file, έτσι ώστε να μπορούμε να τα εκτελέσουμε οποιαδήποτε στιγμή. Το παράθυρο αυτό εμφανίζεται είτε με το πάτημα του κουμπιού «new» είτε πληκτρολογώντας την εντολή «edit» στο παράθυρο εντολών. Αφού δημιουργήσουμε το αρχείο μπορούμε είτε να το σώσουμε στον τρέχων κατάλογο είτε να το εκτελέσουμε άμεσα με τη χρήση του κουμπιού «Run».



Εικόνα 5.7: Παράθυρο Σχημάτων (Figure window).

Στο Παράθυρο Σχημάτων σχεδιάζουμε γραφικές παραστάσεις με τη χρήση κατάλληλων σχεδιαστικών εντολών οι οποίες γράφονται είτε στη γραμμή εντολών είτε σε προγράμματα MATLAB.



Εικόνα 5.8: Παράθυρο Βοήθειας (Help Window).

Το παράθυρο αυτό είναι ένα πολύ χρήσιμο εργαλείο τόσο για αρχάριους όσο και για έμπειρους χρήστες καθώς παρέχει όλες τις απαραίτητες πληροφορίες για οποιαδήποτε συνάρτηση. Μπορούμε να το χρησιμοποιήσουμε γράφοντας την εντολή «help» και το όνομα της συνάρτησης για την οποία θα θέλαμε περισσότερες πληροφορίες.

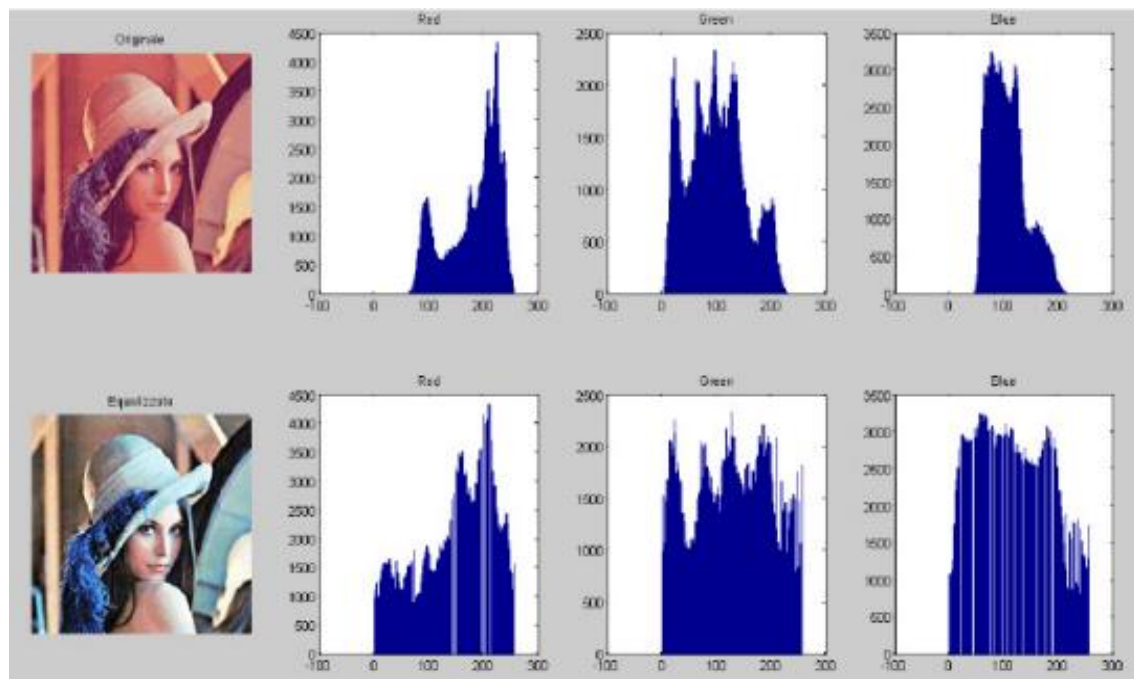
Αρχεία προγραμμάτων και συναρτήσεων

Το Matlab μας δίνει τη δυνατότητα να εργαστούμε με δύο τρόπους. Μπορούμε να χρησιμοποιήσουμε μεμονωμένες εντολές στο παράθυρο εντολών και να πάρουμε άμεσα αποτελέσματα είτε μέσω προγραμμάτων τα οποία γράφονται από το χρήστη στο παράθυρο εισαγωγής κειμένου. Τα προγράμματα αυτά αποθηκεύονται με την κατάληξη .m σαν αρχεία απλού κειμένου και ονομάζονται m-files. Τα ονόματα των m-file δεν επιτρέπεται να περιλαμβάνουν τελείες και συγκεκριμένους ειδικούς χαρακτήρες. Το πρόγραμμα δεν είναι απαραίτητο να περάσει το στάδιο της μεταγλώττισης(compiling), όπως γίνεται σε άλλες γλώσσες προγραμματισμού αλλά γίνεται αυτόματα πριν το στάδιο της εκτέλεσης.

Υπάρχουν δυο ειδών τύποι m-files. Τα αρχεία προγραμμάτων(scripts) και τα αρχεία συναρτήσεων(functions). Τα αρχεία προγραμμάτων είναι αυτόνομα, δεν δέχονται ορίσματα εισόδου και δεν επιστρέφουν μεταβλητές εξόδου. Χρησιμοποιούν τις

υπάρχουσες μεταβλητές που βρίσκονται στο χώρο εργασίας ή δημιουργούν καινούργιες. Τα αρχεία συναρτήσεων από την άλλη, δέχονται υποχρεωτικά ορίσματα εισόδου και μπορούν να επιστρέψουν μεταβλητές εξόδου. Το όνομα μιας συνάρτησης μαζί με τα ορίσματα εισόδου και τις μεταβλητές εξόδου γράφονται στην πρώτη γραμμή του αντίστοιχου m-file συνοδευόμενα από τη λέξη-κλειδί function, ενώ στην τελευταία γραμμή του αρχείου, γράφεται η λέξη-κλειδί end. Επίσης, το m-file που περιέχει μια συνάρτηση πρέπει να έχει το ίδιο όνομα με αυτή.

Τέλος να αναφέρουμε ότι μια συνάρτηση χρησιμοποιεί το δικό της χώρο μεταβλητών, ο οποίος δεν είναι προσπελάσιμος μετά την εκτέλεσή της. Οι συναρτήσεις μπορούν να κληθούν μέσα από scripts ή από άλλες συναρτήσεις για διάφορες τιμές των ορισμάτων εισόδου τους.

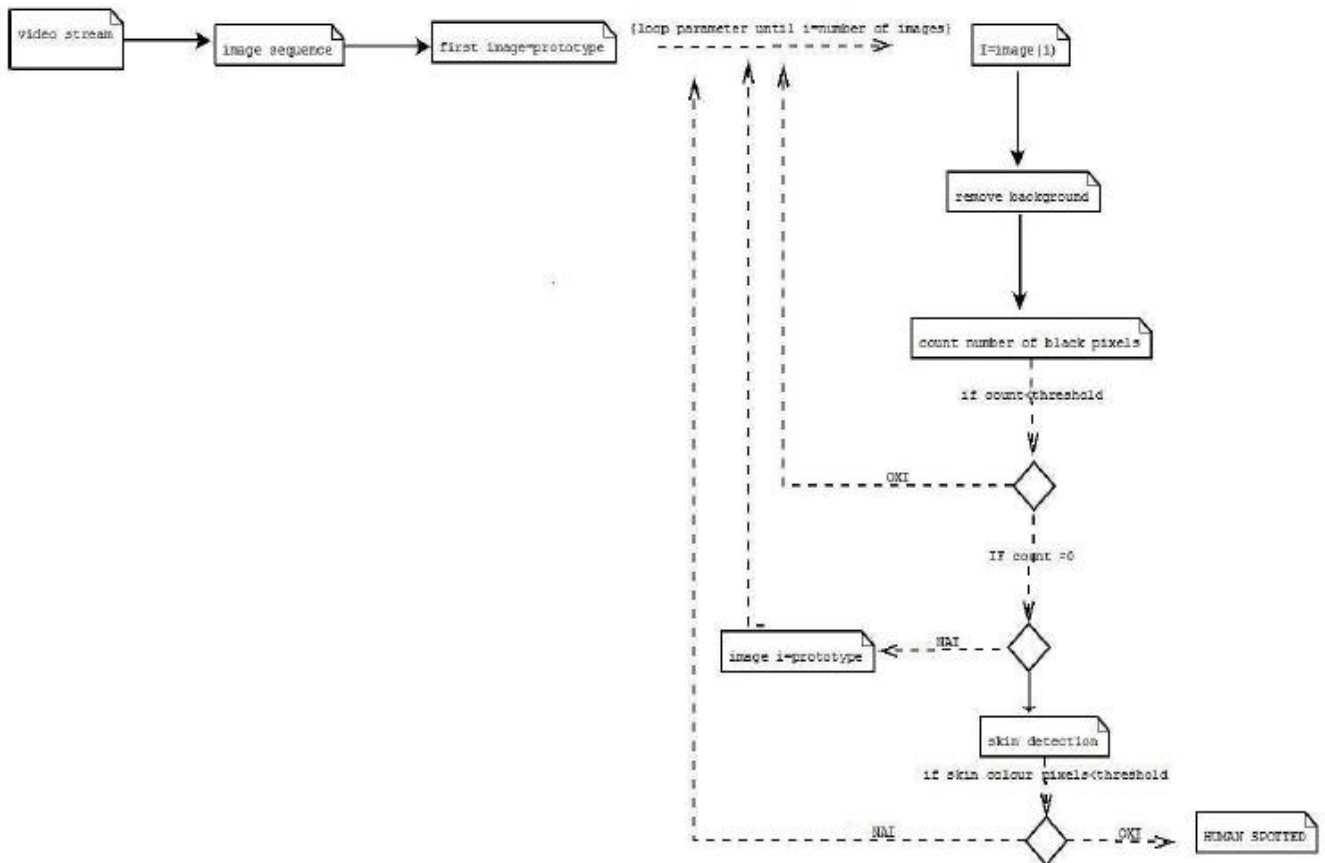


Εικόνα 5.9: Παράδειγμα χρήσης του Matlab στην επεξεργασία εικόνας.

5.2 Περιγραφή αλγορίθμου

Θα ξεκινήσουμε περιγράφοντας τον τρόπο με τον οποίο λειτουργεί ο αλγόριθμος και θα επισημάνουμε τις προδιαγραφές βάσει των οποίων είναι λειτουργικός, τις αδυναμίες του και τα δυνατά του σημεία. Θα εξηγήσουμε πως επιτυγχάνεται η ανίχνευση ανθρώπινης παρουσίας στο χώρο και θα δούμε διάφορα παραδείγματα του αλγορίθμου αυτού σε διάφορες σκηνές.

Ξεκινώντας, να επισημάνουμε ότι το πρόγραμμα λειτουργεί σε πλάνα τραβηγμένα από ακίνητη κάμερα, τόσο σε εσωτερικό όσο και σε εξωτερικό χώρο. Επιπλέον, είναι αρκετά ελαστικό με αλλαγές στον φωτισμό αλλά και στην εισαγωγή νέων αντικειμένων στη σκηνή χωρίς να επηρεάζεται η αποδοτικότητά του.



Εικόνα 5.10: Γραφική Αναπαράσταση Αλγορίθμου.

Αρχικά, ο χρήστης επιλέγει το βίντεο στο οποίο θέλει να εφαρμόσει τον αλγόριθμο για να ελέγξει για παρουσία ανθρώπου. Το βίντεο αυτό αναπαράγεται και χωρίζεται σε στιγμιότυπα για μετέπειτα επεξεργασία. Απαραίτητη προϋπόθεση είναι να μην υπάρχει στη σκηνή ανθρώπινη παρουσία στο πρώτο frame. Αφού δημιουργηθούν τα στιγμιότυπα ξεκινάει ο αλγόριθμος που ελέγχει τις εικόνες για τον εντοπισμό ανθρώπου. Ορίζει το πρώτο frame του βίντεο σαν πρωτότυπο και ξεκινάει την διαδικασία της αφαίρεσης background διαβάζοντας τις εικόνες και συγκρίνοντας τις με την πρώτη.

Στη συνέχεια ελέγχει το ποσοστό της εικόνας που έχει αλλάξει, και αν είναι μεγαλύτερο από ένα συγκεκριμένο κατώφλι που ορίζουμε εμείς προχωράει στο δεύτερο μέρος, δηλαδή τη διαδικασία για εντοπισμό pixel στο χρώμα του δέρματος.

Στην περίπτωση που το ποσοστό που άλλαξε είναι αρκετά μικρό θεωρούμε ότι δεν υπάρχει άνθρωπος στη σκηνή στο συγκεκριμένο frame και προχωράει στο επόμενο. Προκειμένου να αποφύγουμε λάθη που προκύπτουν από την παρουσία χρώματος παρόμοιου με αυτό του ανθρώπινου δέρματος, μελετάμε την εικόνα που έχει προκύψει από την αφαίρεση του background. Με αυτό τον τρόπο έχουμε μικρότερη πιθανότητα για εσφαλμένη ανίχνευση, αφού εργαζόμαστε μόνο πάνω στο κομμάτι της εικόνας που έχουμε εντοπίσει διαφορά σε σχέση με την αρχική μας. Αν ο αλγόριθμος εντοπίσει pixel με χρώμα κοντινό σε αυτό του δέρματος, θεωρεί ότι η αλλαγή στην εικόνα προέκυψε από ανθρώπινη παρουσία και τερματίζει τη λειτουργία του επιστρέφοντας στην εικόνα που εντόπισε τον άνθρωπο, καθώς και στην εικόνα που προέκυψε από την αφαίρεση του φόντου. Στην περίπτωση που δεν εντοπίσει pixel στο χρώμα του δέρματος, υποθέτουμε ότι υπήρξε μια αλλαγή στην εικόνα μας η οποία δεν ήταν αποτέλεσμα ανθρώπινης παρουσίας και συνεχίζει την εκτέλεση του προγράμματος ελέγχοντας το επόμενο frame.

Σε αυτό το σημείο, να κάνουμε μια περιγραφή των δυο βασικών προβλημάτων που προκύπτουν. Το ένα είναι το πρόβλημα της αλλαγής φωτισμού και το δεύτερο της μετακίνησης σημείων του background.

Στην πρώτη περίπτωση, όταν αλλάξει ο φωτισμός απότομα, το πρόγραμμα θα εντοπίσει μεγάλη αλλαγή στην εικόνα μας και θα ξεκινήσει τη διαδικασία για ανίχνευση δέρματος. Η πιθανότητα να βρει σε όλη την εικόνα pixel στο χρώμα του δέρματος είναι αρκετά μεγάλη λόγω των αντικειμένων που θα βρίσκονται στο χώρο και θα έχουν παρόμοιο χρώμα με αυτό του δέρματος, για παράδειγμα ξύλινα αντικείμενα. Για να αποφύγουμε αυτό το πρόβλημα εισάγουμε μια νέα συνθήκη στον κώδικα έτσι ώστε όταν ανιχνεύσει αλλαγή σε όλα τα pixel της εικόνας να θεωρεί ότι η αλλαγή αυτή προήλθε από αλλαγή του φωτισμού. Σε αυτή την περίπτωση θεωρεί ως πρωτότυπη την εικόνα, η οποία προέκυψε από την αλλαγή φωτισμού και συνεχίζει να ελέγχει τις επόμενες βάσει αυτής.

Το δεύτερο πρόβλημα αφορά σκηνές που δεν είναι ελεγχόμενες. Όπως είναι οι σκηνές εξωτερικού χώρου. Σε ένα τέτοιο περιβάλλον υπάρχουν αρκετοί αντιπερισπασμοί, όπως είναι τα φύλλα των δέντρων, η κίνηση οχημάτων και διάφορα ζώα. Για να κάνουμε το σύστημα μας πιο ελαστικό σε ένα τέτοιο περιβάλλον προσπαθήσαμε να κάνουμε τον αλγόριθμο της αφαίρεσης background να αλλάζει πιο ομαλά από ότι θα άλλαζε με την αφαίρεση της μιας εικόνας από την άλλη. Η συνάρτηση που προκύπτει μετά από μορφοποίηση είναι η εξής:

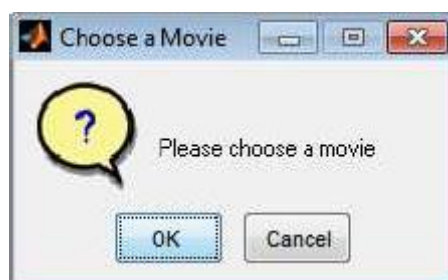
$$\text{Background} = (1-\alpha) * I + \alpha * I_b;$$

Στην συνάρτηση αυτή η εικόνα με τίτλο background, δημιουργείται από ένα ποσοστό του αρχικού frame και ένα ποσοστό του frame το οποίο εξετάζουμε για παρουσία ανθρώπου. Με αυτό τον τρόπο σκοπεύουμε να κάνουμε τον αλγόριθμό μας πιο ελαστικό σε μικρές αλλαγές, που προκύπτουν από κινήσεις που δεν παρουσιάζουν ενδιαφέρον και εντάσσονται στα στοιχεία του background. Μετά από αυτή την περιγραφή θα δούμε κατά πόσο είναι λειτουργικός ο αλγόριθμος σε πραγματικά περιβάλλοντα και θα εντοπίσουμε τυχόν αδυναμίες του αλγόριθμου.

5.3 Παρουσίαση της εφαρμογής

Θα ξεκινήσουμε την παρουσίαση της εφαρμογής με την δοκιμή του αλγόριθμου στο πιο απλό περιβάλλον. Σε αυτή την περίπτωση θα μελετήσουμε πως λειτουργεί σε ένα ελεγχόμενο εσωτερικό χώρο χωρίς κανένα περισπασμό.

Σε αυτό το πλάνο θα έχουμε μια σταθερή κάμερα με σταθερό και ελεγχόμενο φωτισμό και το αντικείμενο ενδιαφέροντος θα περάσει μπροστά από το πλάνο προκειμένου να δούμε αν θα καταφέρει να το αναγνωρίσει. Όταν πατήσουμε το πλήκτρο για να εκτελεστεί ο κώδικας εμφανίζεται ένα παράθυρο για να μπορέσει ο χρήστης να διαλέξει το βίντεο που επιθυμεί.



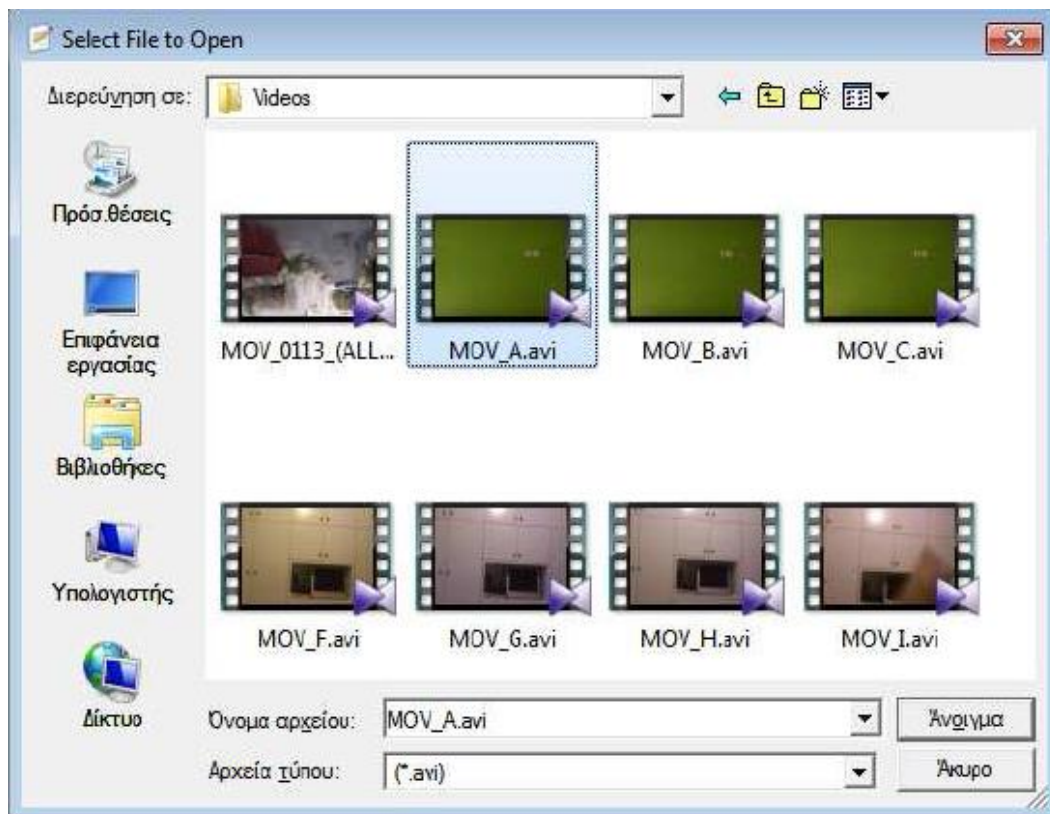
Εικόνα 5.11: Παράθυρο επιλογής βίντεο.

Τα βίντεο πρέπει να είναι σε μορφή που υποστηρίζεται από το Matlab. Ακολουθεί ένας πίνακας που δείχνει τα format που υποστηρίζονται από το Matlab.

All Platforms	AVI (.avi), Motion JPEG 2000 (.mj2)
All Windows [®]	MPEG-1 (.mpg), Windows Media [®] Video (.wmv, .asf, .asx), and any format supported by Microsoft [®] DirectShow [®] .
Windows 7	MPEG-4, including H.264 encoded video (.mp4, .m4v), Apple QuickTime Movie (.mov), and any format supported by Microsoft Media Foundation.
Macintosh	MPEG-1 (.mpg), MPEG-4, including H.264 encoded video (.mp4, .m4v), Apple QuickTime Movie (.mov), and any format supported by QuickTime as listed on http://support.apple.com/kb/HT3775 .
Linux [®]	Any format supported by your installed plug-ins for GStreamer 0.10 or above, as listed on http://gstreamer.freedesktop.org/documentation/plugins.html , including Ogg Theora (.ogg).

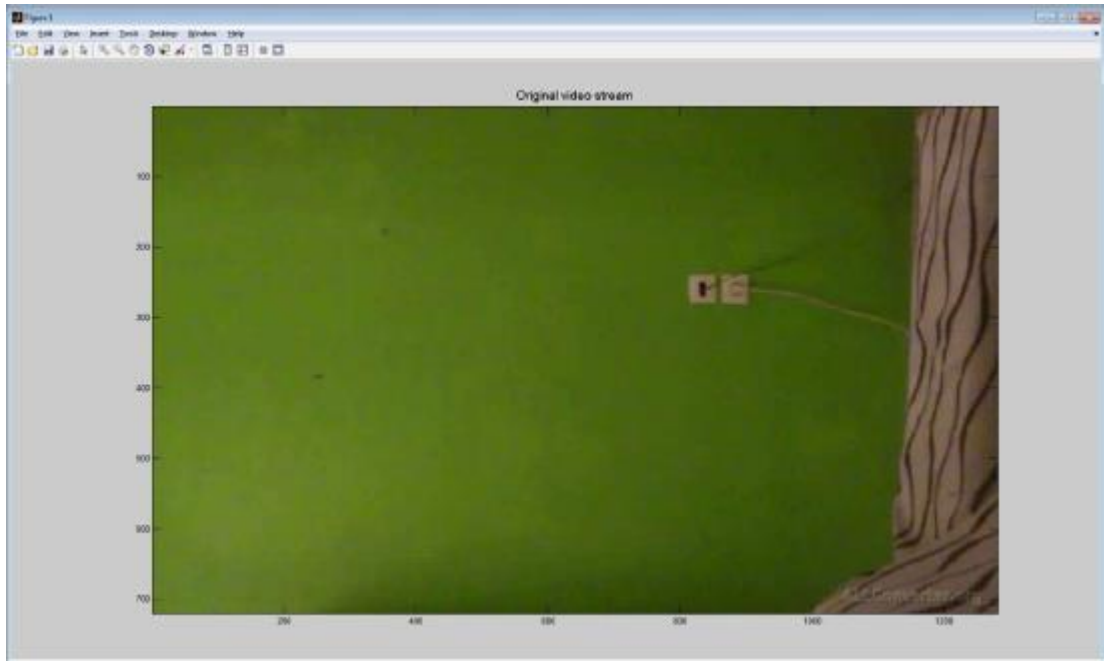
Εικόνα 5.12: Πίνακας υποστηριζόμενων μορφών βίντεο για το Matlab.

Στη συνέχεια ο χρήστης διαλέγει το βίντεο που επιθυμεί, το οποίο εισάγεται στο Matlab για να αναλυθεί και να εξεταστεί για τυχόν παρουσία ανθρώπου.

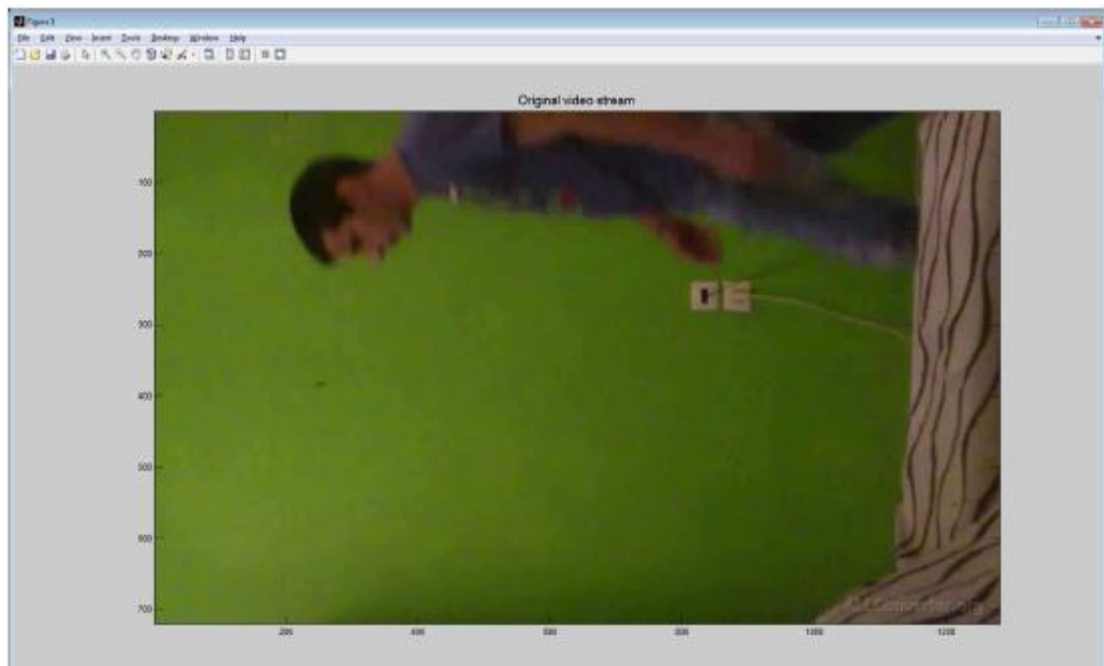


Εικόνα 5.13: Επιλογή αρχείου βίντεο προς επεξεργασία.

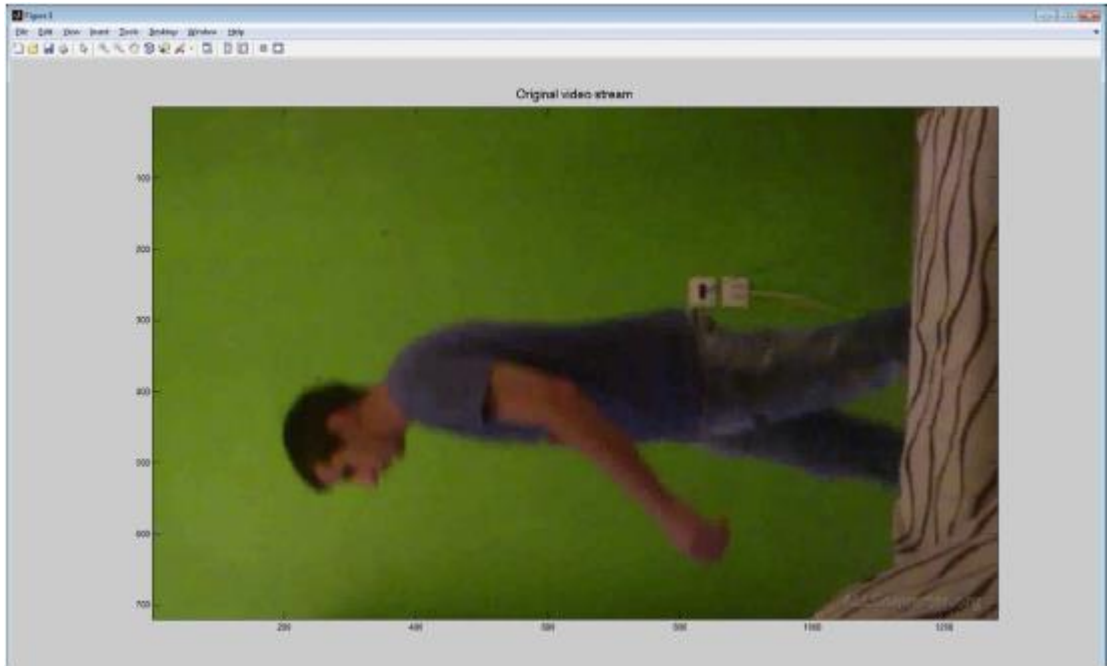
Αφού επιλέξουμε το αρχείο που επιθυμούμε ξεκινάει η αναπαραγωγή του βίντεο, ενώ ταυτόχρονα αποθηκεύεται η ακολουθία εικόνων.



Εικόνα 5.14: Στιγμιότυπο 1 από την αναπαραγωγή βίντεο1.

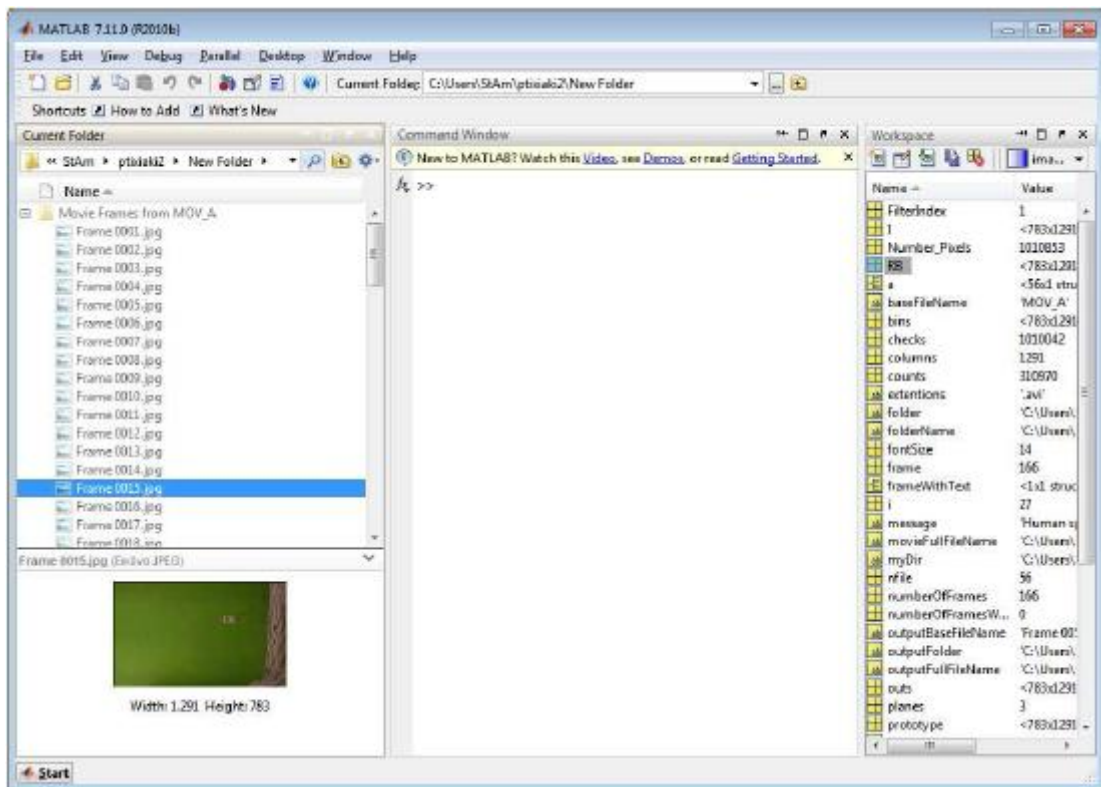


Εικόνα 5.15: Στιγμιότυπο 2 από την αναπαραγωγή βίντεο1.



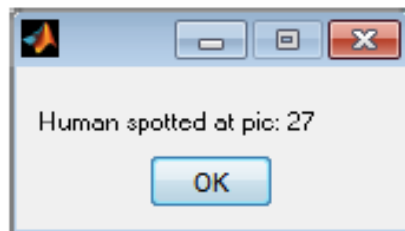
Εικόνα 5.16: Στιγμιότυπο 3 από την αναπαραγωγή βίντεο1.

Βλέπουμε στην παρακάτω εικόνα ότι στο φάκελο εργασίας μας στο Matlab έχει δημιουργηθεί ένας καινούργιος φάκελος με το όνομα του βίντεο και μέσα σε αυτόν βρίσκονται όλα τα στιγμιότυπα από το βίντεο.

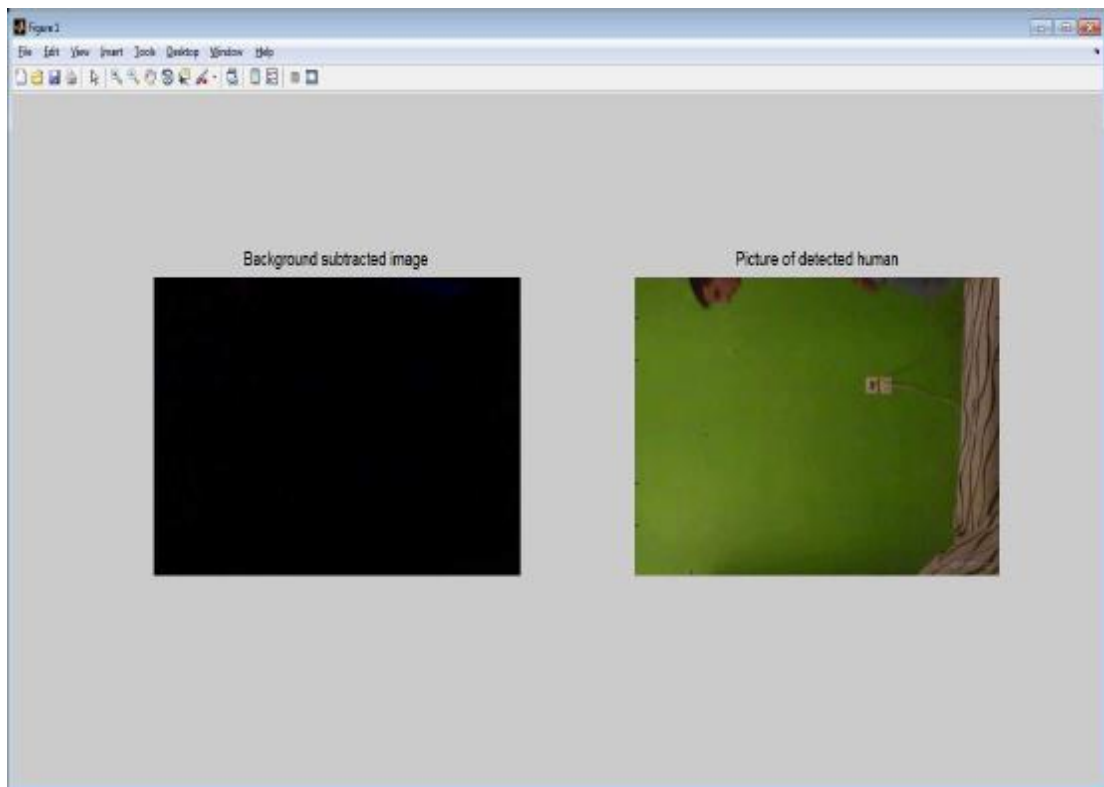


Εικόνα 5.17: Στιγμιότυπο από το περιβάλλον εργασίας του Matlab.

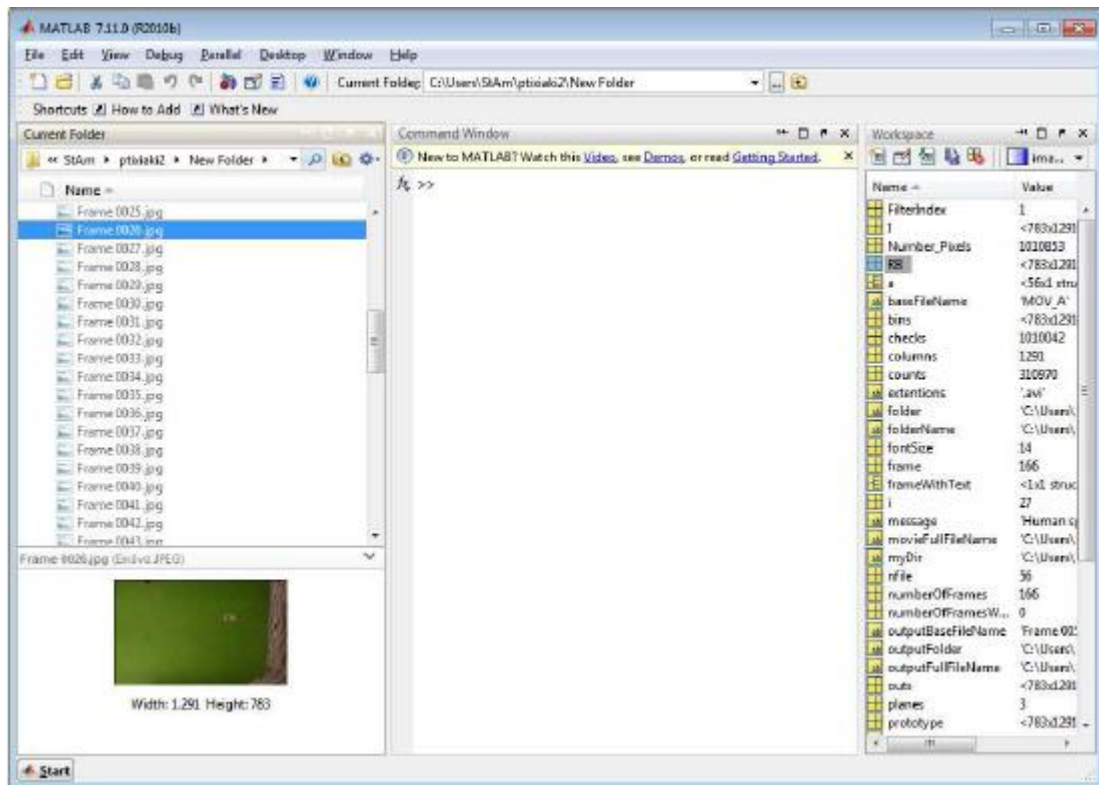
Μόλις τελειώσει η αναπαραγωγή του βίντεο ξεκινάει η ανίχνευση για ανθρώπινη παρουσία, με τον τρόπο που περιγράψαμε παραπάνω. Αν καταφέρει να εντοπίσει ανθρώπινη παρουσία εμφανίζει ένα μήνυμα με τον αριθμό της εικόνας που εντόπισε τον άνθρωπο, καθώς και την αρχική εικόνα μαζί με αυτήν που προέκυψε από την αφαίρεση του background.



Εικόνα 5.18: Μήνυμα ανίχνευσης ανθρώπου στο βίντεο1.



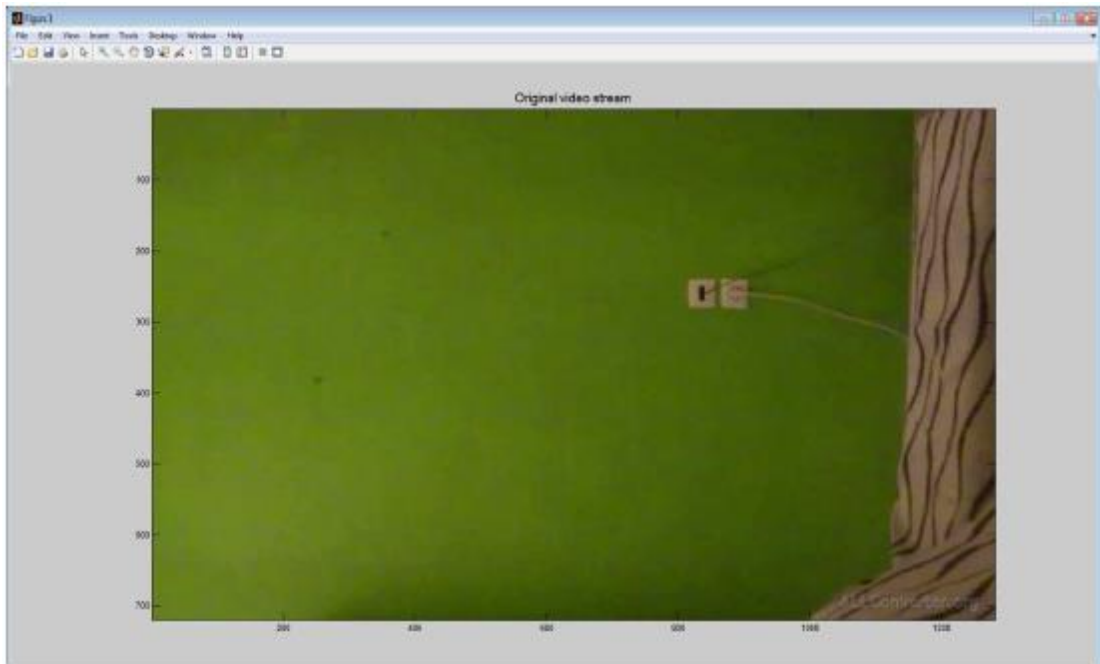
Εικόνα 5.19: Η εικόνα που εντοπίστηκε ανθρώπινη παρουσία με και χωρίς αφαίρεση background στο βίντεο1.



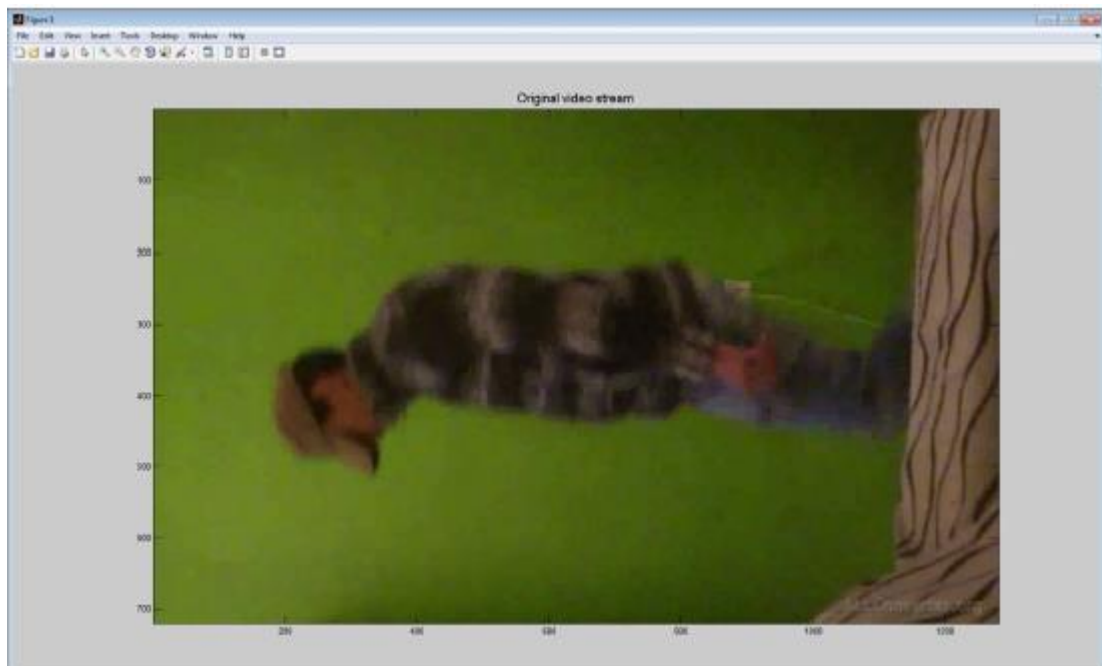
Εικόνα 5.20: Στιγμιότυπο από το περιβάλλον εργασίας του Matlab μετά την εκτέλεση της εφαρμογής για το βίντεο1.

Βλέπουμε στην παραπάνω εικόνα ότι το frame 27 είναι πράγματι η πρώτη φορά που εμφανίζεται η ανθρώπινη παρουσία. Συνεπώς, συμπεραίνουμε ότι ο αλγόριθμος είναι αρκετά αποδοτικός σε περιβάλλοντα πάρα πολύ απλά χωρίς περισπασμούς.

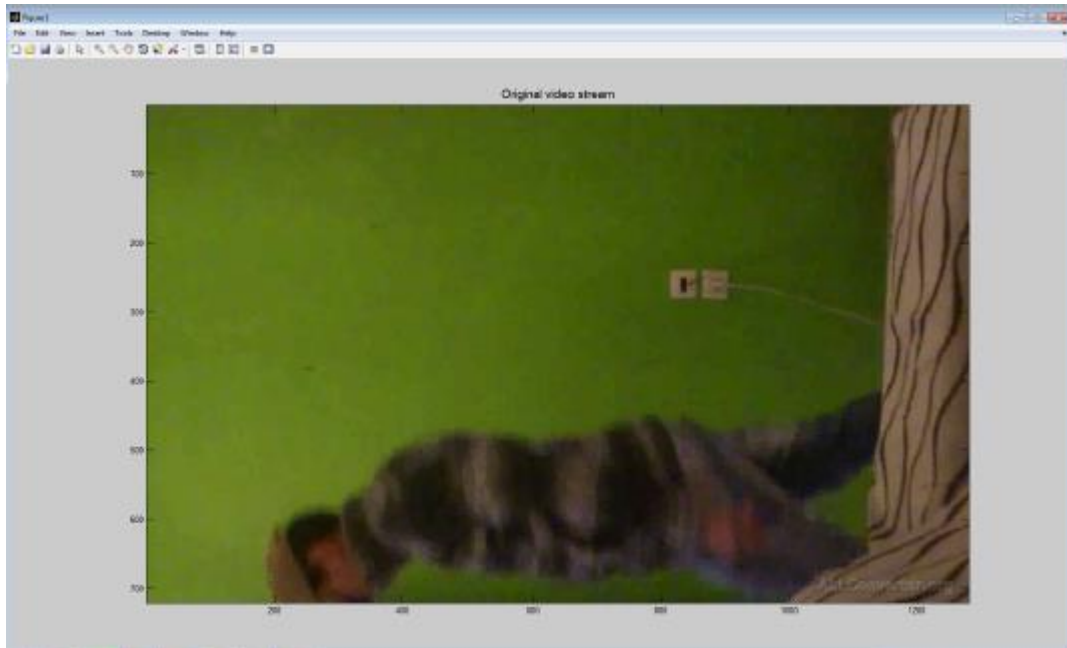
Εν συνεχεία, θα δοκιμάσουμε το αν μπορεί να ανιχνεύσει ανθρώπινη παρουσία όταν είναι λιγότερα τα pixels του δέρματος. Στο επόμενο βίντεο προσπαθούμε να μειώσουμε τις περιοχές που εμφανίζεται δέρμα.



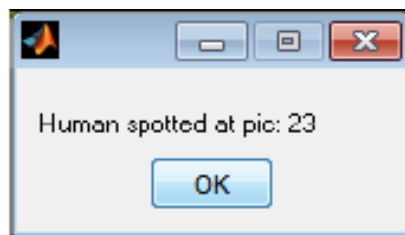
Εικόνα 5.21: Στιγμιότυπο 1 από την αναπαραγωγή βίντεο2.



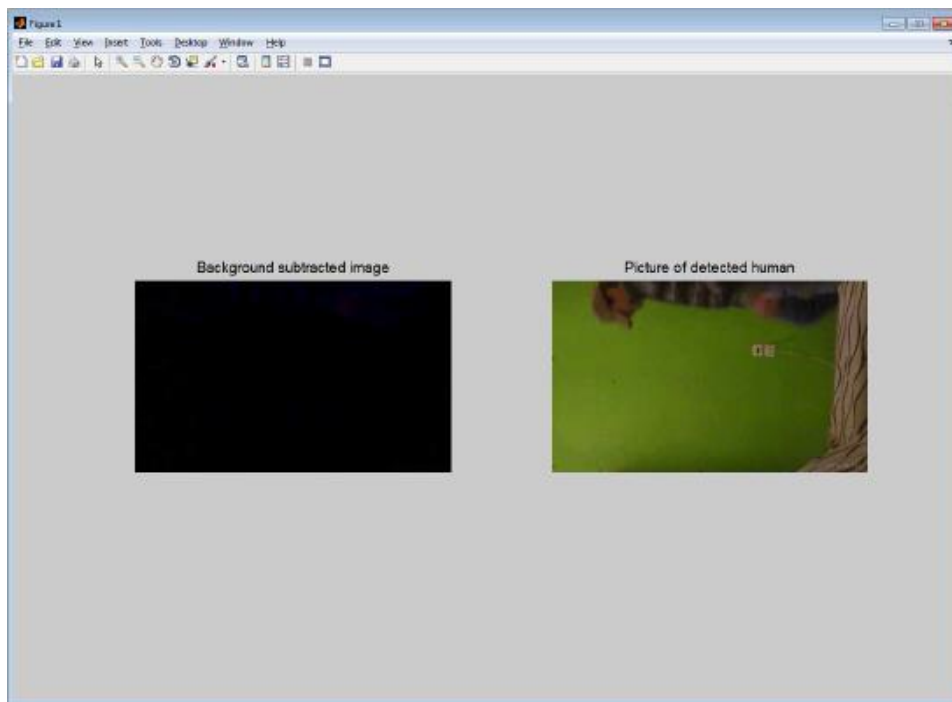
Εικόνα 5.22: Στιγμιότυπο 2 από την αναπαραγωγή βίντεο2.



Εικόνα 5.23: Στιγμιότυπο 3 από την αναπαραγωγή βίντεο2.



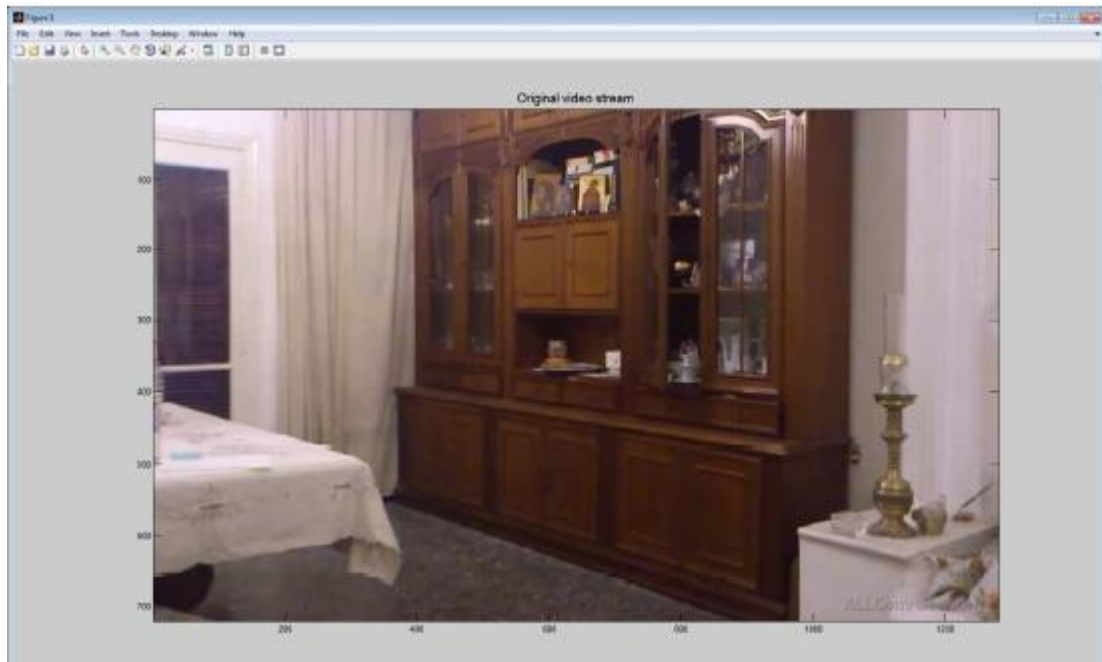
Εικόνα 5.24: Μήνυμα ανίχνευσης ανθρώπου στο βίντεο2.



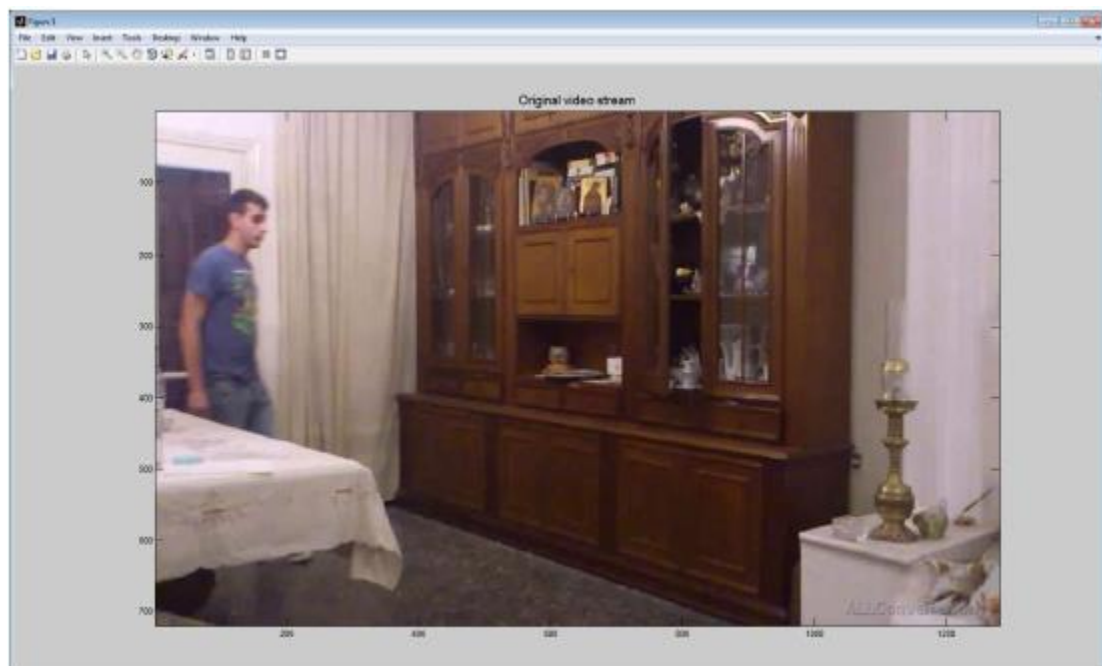
Εικόνα 5.25: Η εικόνα που εντοπίστηκε ανθρώπινη παρουσία στο βίντεο2 με και χωρίς αφαίρεση Background.

Βλέπουμε ότι και με τις περισσότερες περιοχές δέρματος καλυμμένες ο αλγόριθμος καταφέρνει με επιτυχία να ανιχνεύσει την ανθρώπινη παρουσία.

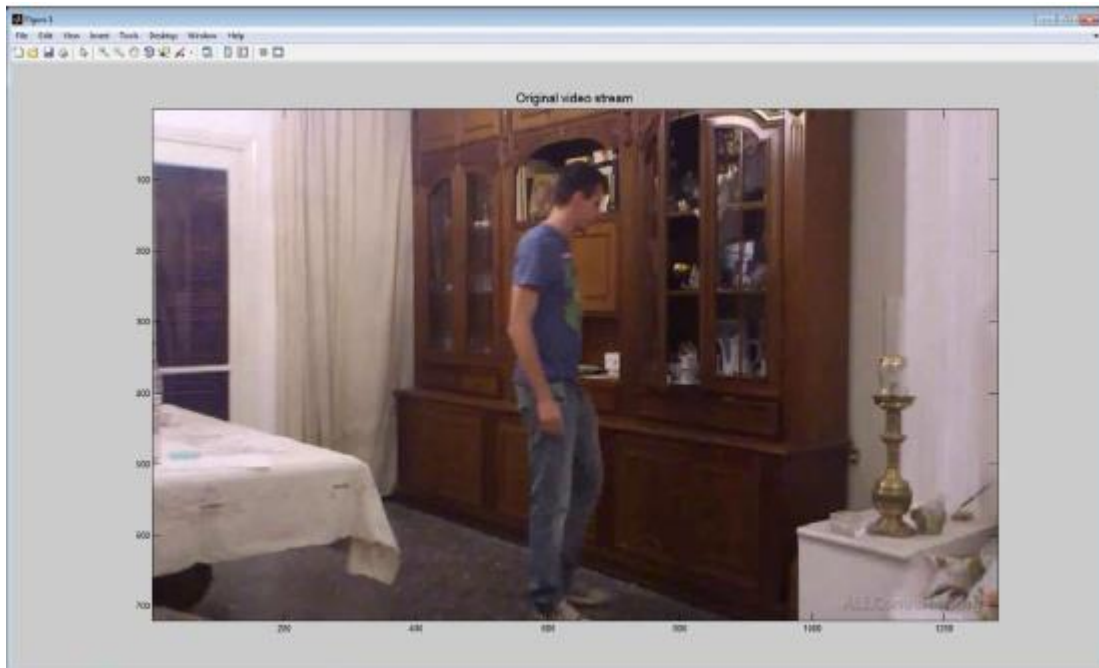
Στο επόμενο στάδιο θα δοκιμάσουμε το πρόγραμμα με ένα βίντεο τραβηγμένο από ένα πιο περίπλοκο εσωτερικό χώρο, όπου θα υπάρχουν πολλά αντικείμενα στο χρώμα του δέρματος.



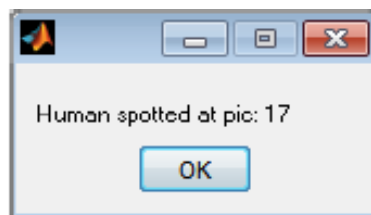
Εικόνα 5.26: Στιγμιότυπο 1 από την αναπαραγωγή βίντεο3.



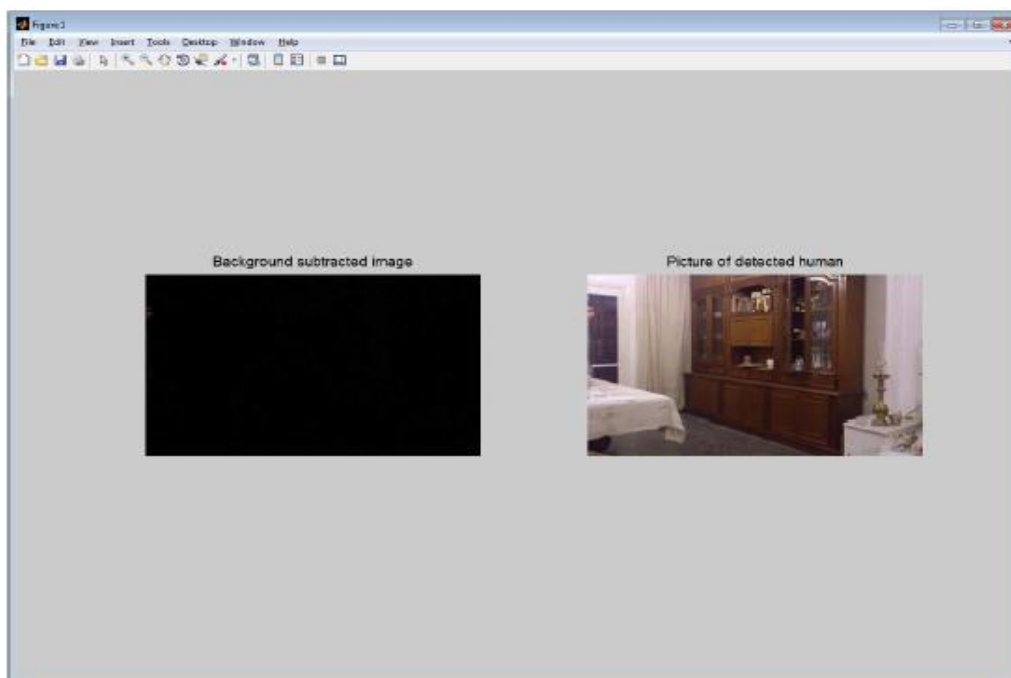
Εικόνα 5.27: Στιγμιότυπο 2 από την αναπαραγωγή βίντεο3.



Εικόνα 5.28: Στιγμιότυπο 3 από την αναπαραγωγή βίντεο3.

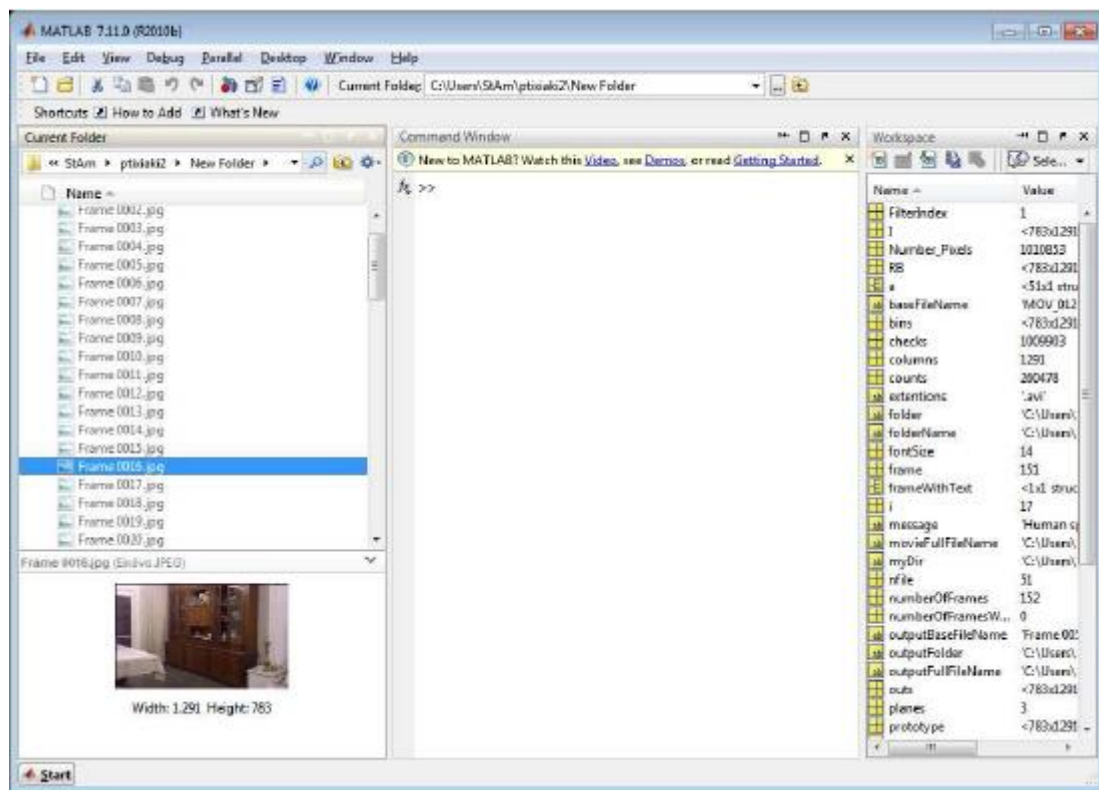


Εικόνα 5.29: Μήνυμα αντίγνωσης ανθρώπου στο βίντεο3.



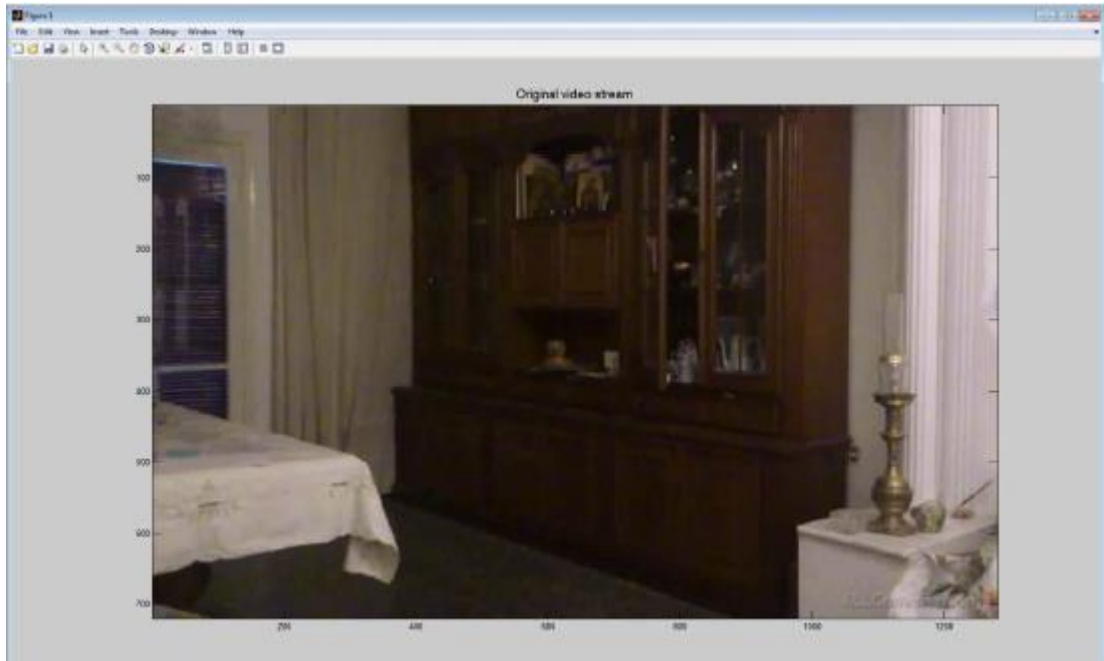
Εικόνα 5.30: Η εικόνα που εντοπίστηκε ανθρώπινη παρουσία στο βίντεο3 με και χωρίς αφαίρεση Background.

Βλέπουμε ότι παρόλη την πολυπλοκότητα του φόντου ο αλγόριθμος με επιτυχία εντοπίζει την ανθρώπινη παρουσία, μόλις εμφανίζεται, και δεν επηρεάζεται από τις σκιάσεις και τις ανακλάσεις που προκύπτουν σταδιακά στην εικόνα.

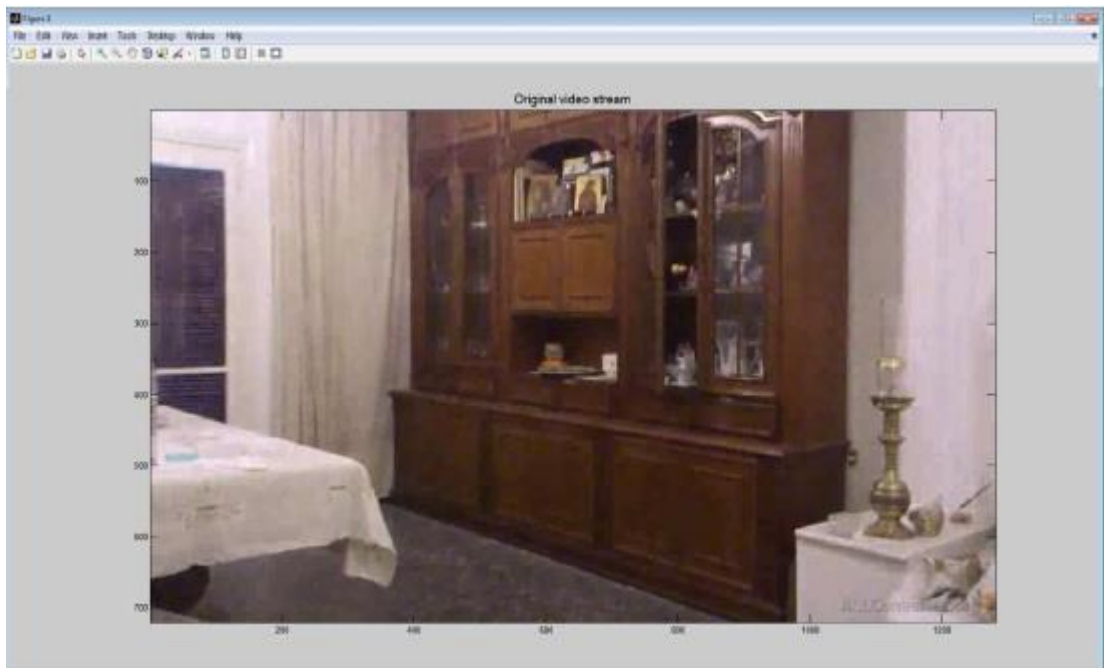


Εικόνα 5.31: Στιγμιότυπο από το περιβάλλον εργασίας του Matlab μετά την εκτέλεση της εφαρμογής για το βίντεο3.

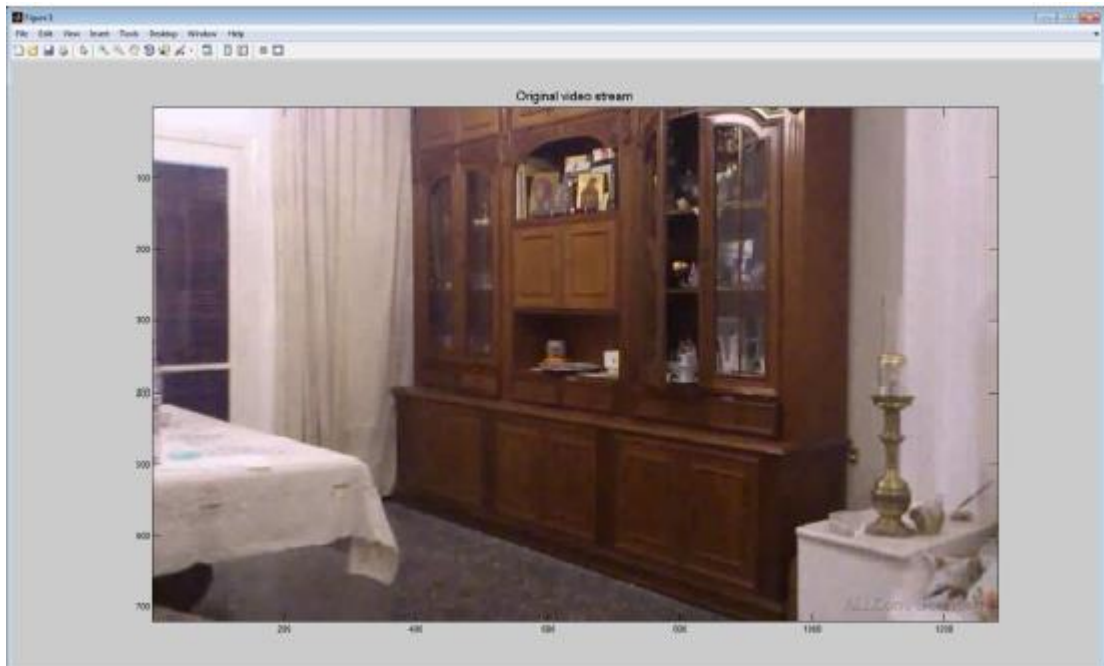
Στη συνέχεια θα δοκιμάσουμε την αποτελεσματικότητά του σε περιβάλλον με αλλαγή φωτισμού. Στο δωμάτιο θα αυξήσουμε το φωτισμό και στη συνέχεια θα εμφανιστεί ο άνθρωπος.



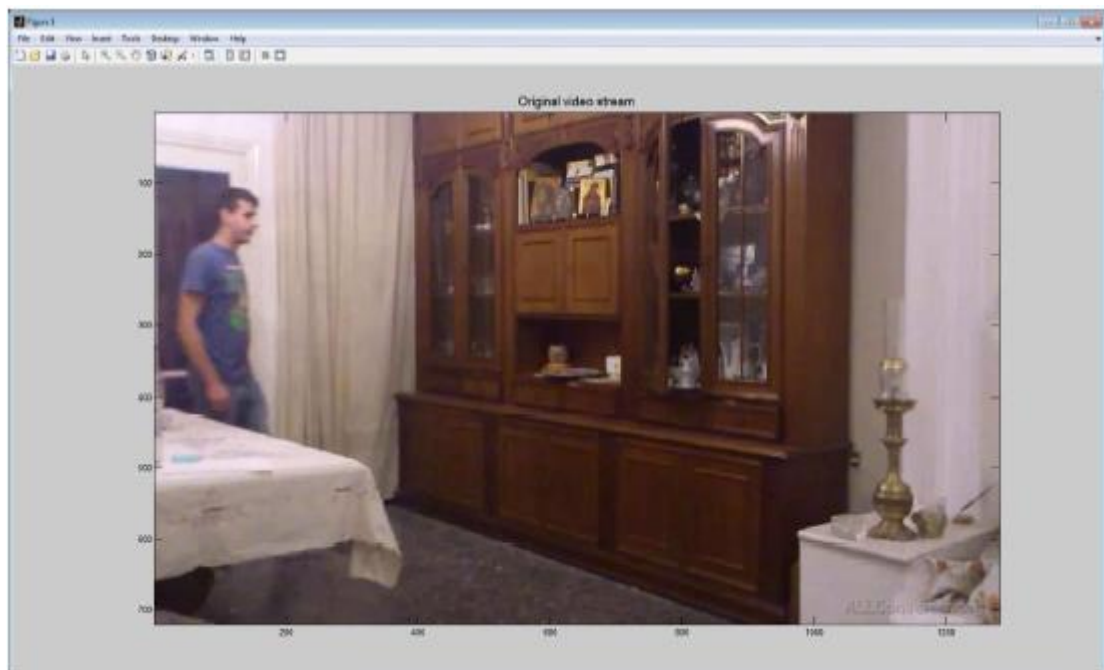
Εικόνα 5.32: Στιγμιότυπο 1 από την αναπαραγωγή βίντεο4.



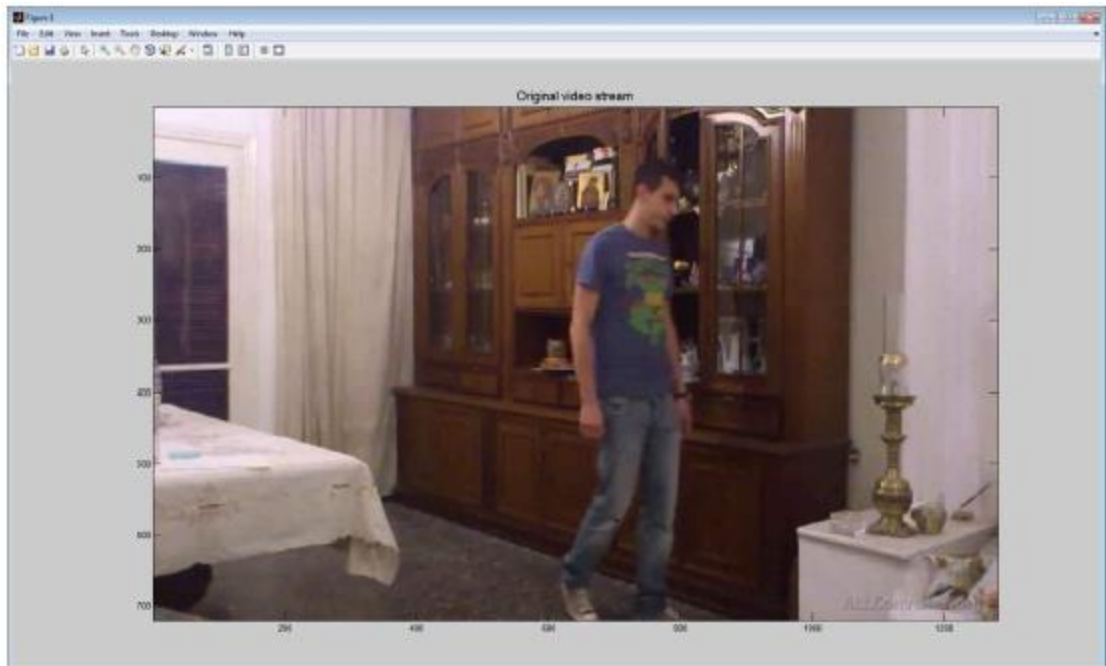
Εικόνα 5.33: Στιγμιότυπο 2 από την αναπαραγωγή βίντεο4.



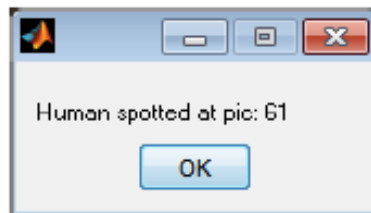
Εικόνα 5.34: Στιγμιότυπο 3 από την αναπαραγωγή βίντεο4.



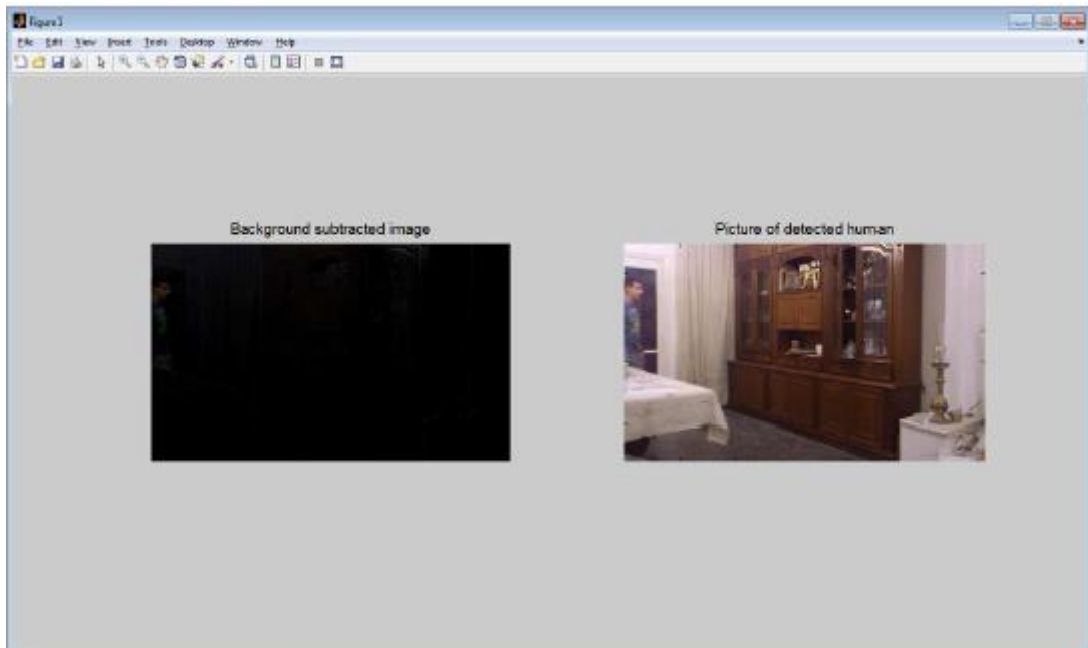
Εικόνα 5.35: Στιγμιότυπο 4 από την αναπαραγωγή βίντεο4.



Εικόνα 5.36: Στιγμιότυπο 5 από την αναπαραγωγή βίντεο4.

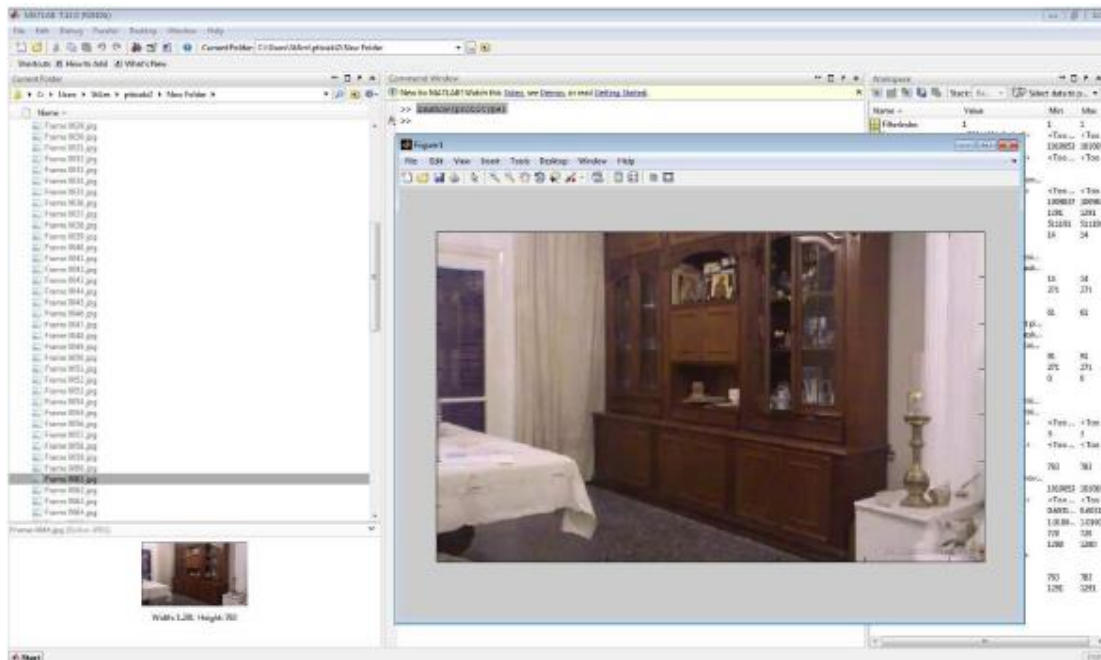


Εικόνα 5.37: Μήνυμα ανίχνευσης ανθρώπου στο βίντεο4.



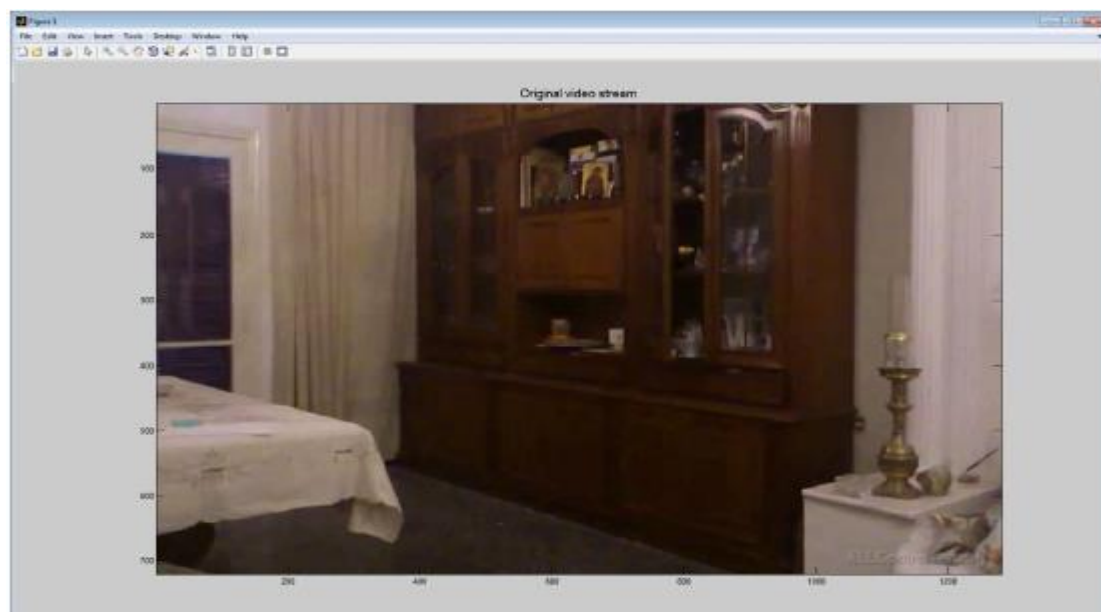
Εικόνα 5.38: Η εικόνα που εντοπίστηκε ανθρώπινη παρουσία στο βίντεο4 με και χωρίς αφαίρεση Background.

Παρατηρούμε ότι παρόλο που αλλάζει ο φωτισμός το πρόγραμμα δεν κάνει εσφαλμένη ανίχνευση σε pixel που έχουν χρώμα παρόμοιο με αυτό του ανθρώπινου δέρματος. Αφού αλλάζει ο φωτισμός όλα τα pixel τις εικόνες αλλάζουν χρώμα και δεν ξεκινάει τον αλγόριθμο για ανίχνευση δέρματος αλλά θεωρεί ως πρωτότυπη εικόνα την καινούργια και συνεχίζει την εκτέλεση του όπως βλέπουμε και στην εικόνα που ακολουθεί.

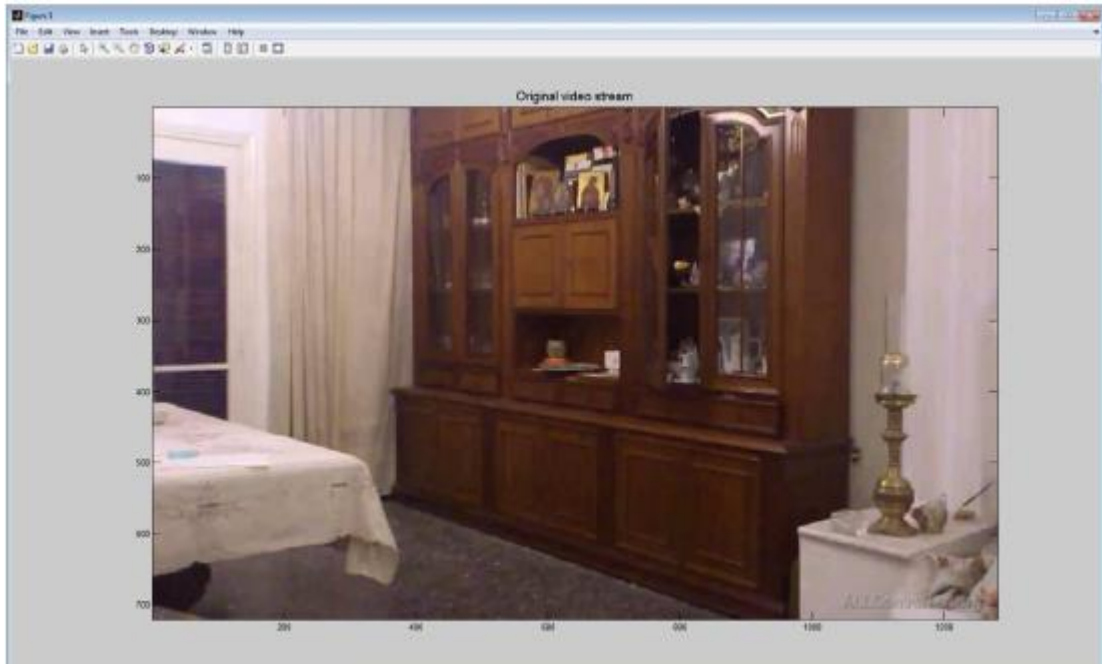


Εικόνα 5.39: Εναλλαγή αρχικής εικόνας όταν αλλάζει ο φωτισμός.

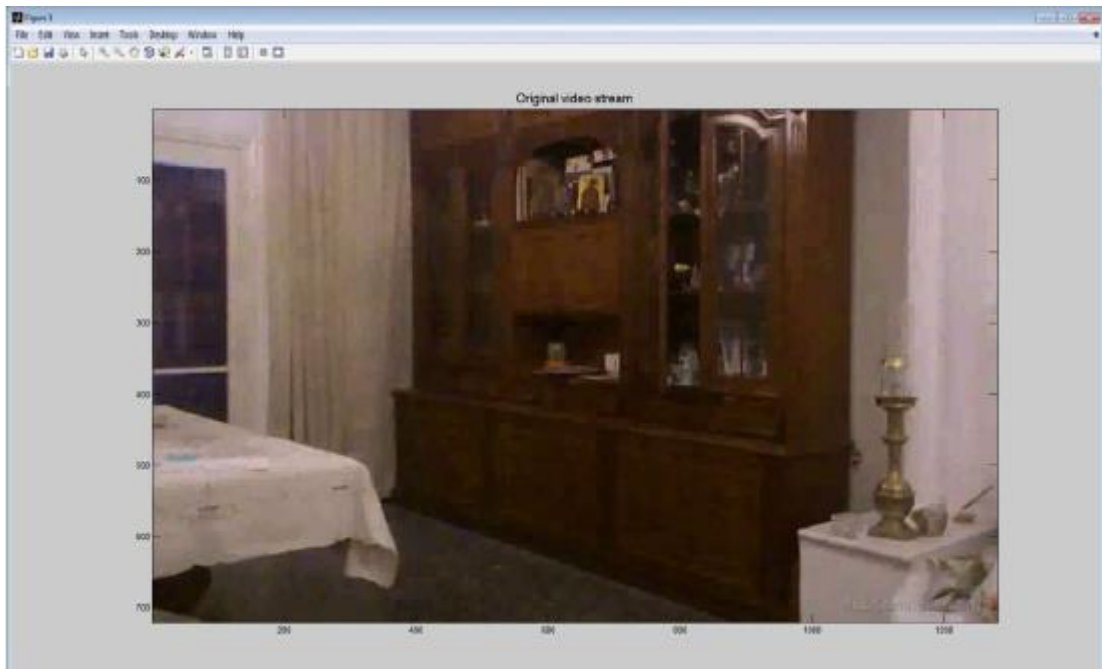
Τελειώνοντας με τις σκηνές εσωτερικού χώρου θα δοκιμάσουμε τι θα γίνει στην περίπτωση που ο φωτισμός δεν μένει σταθερός αλλά αλλάζει συνεχώς.



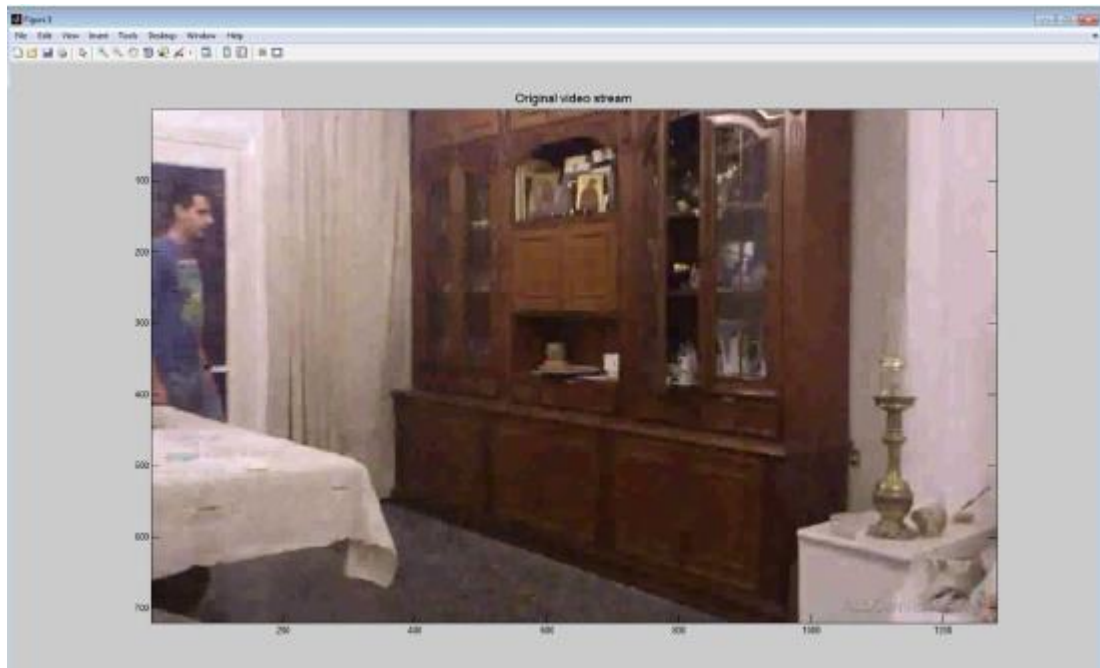
Εικόνα 5.40: Στιγμιότυπο 1 από την αναπαραγωγή βίντεο5.



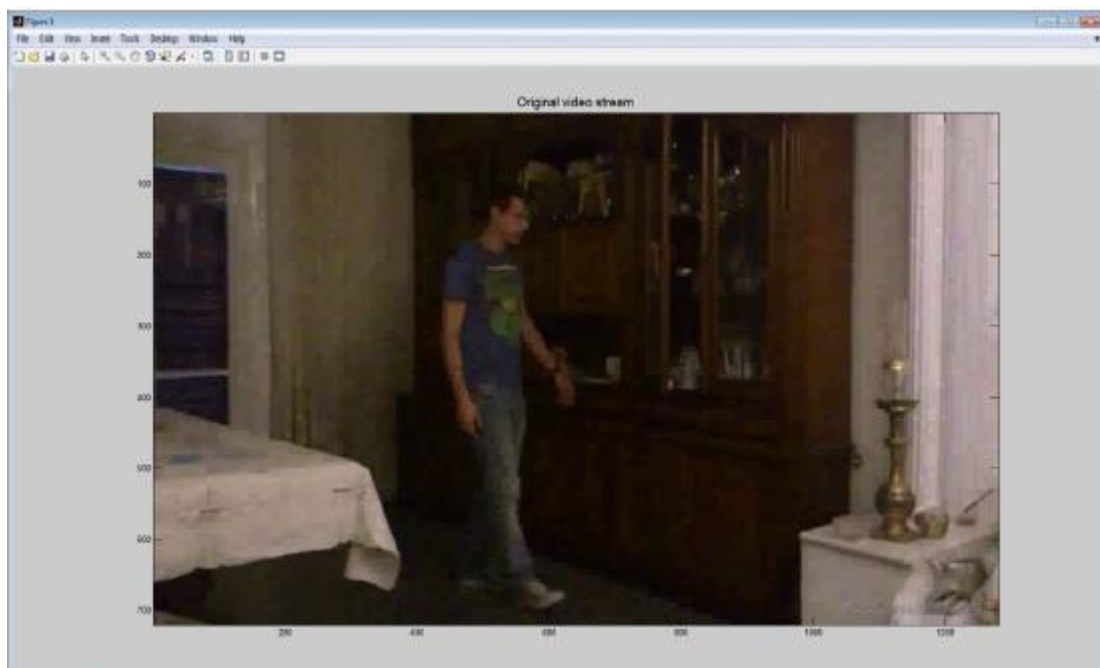
Εικόνα 5.41: Στιγμιότυπο 2 από την αναπαραγωγή βίντεο5.



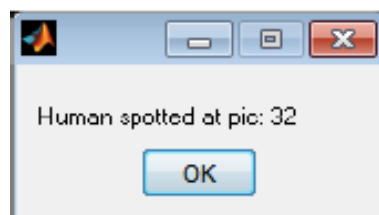
Εικόνα 5.42: Στιγμιότυπο 3 από την αναπαραγωγή βίντεο5.



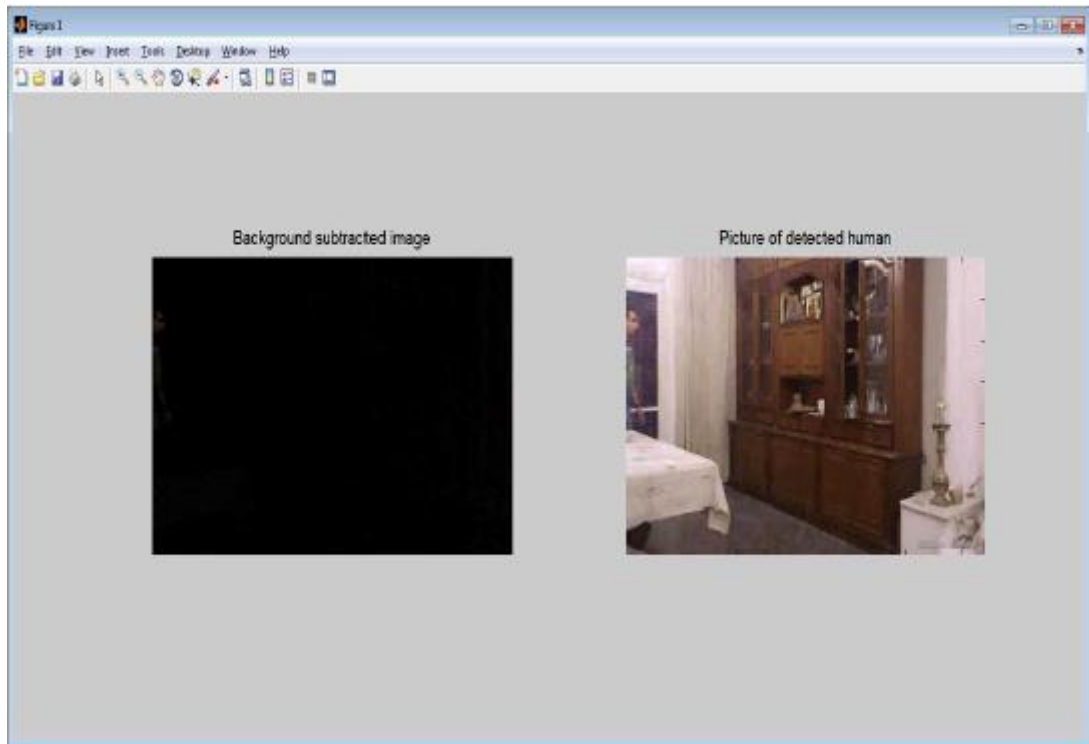
Εικόνα 5.43: Στιγμιότυπο 4 από την αναπαραγωγή βίντεο5.



Εικόνα 5.44: Στιγμιότυπο 5 από την αναπαραγωγή βίντεο5.

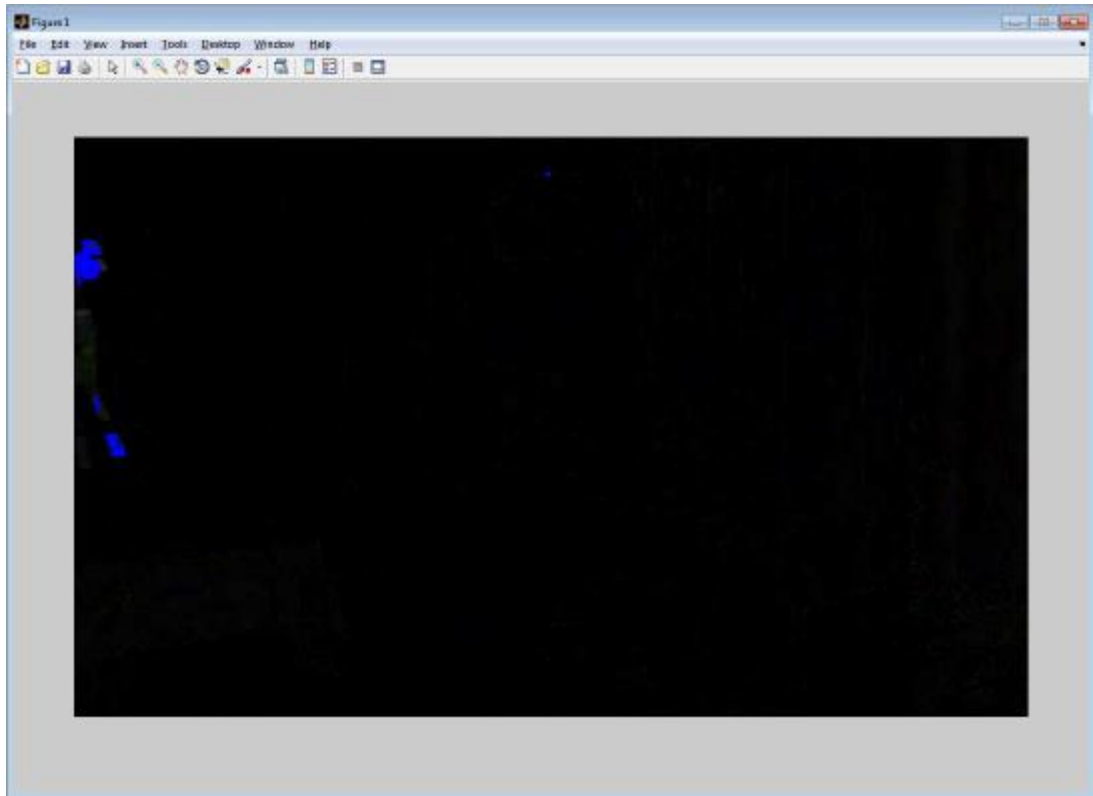


Εικόνα 5.45: Μήνυμα ανίχνευσης ανθρώπου στο βίντεο5.



Εικόνα 5.46: Η εικόνα που εντοπίστηκε ανθρώπινη παρουσία στο βίντεο5 με και χωρίς αφαίρεση Background.

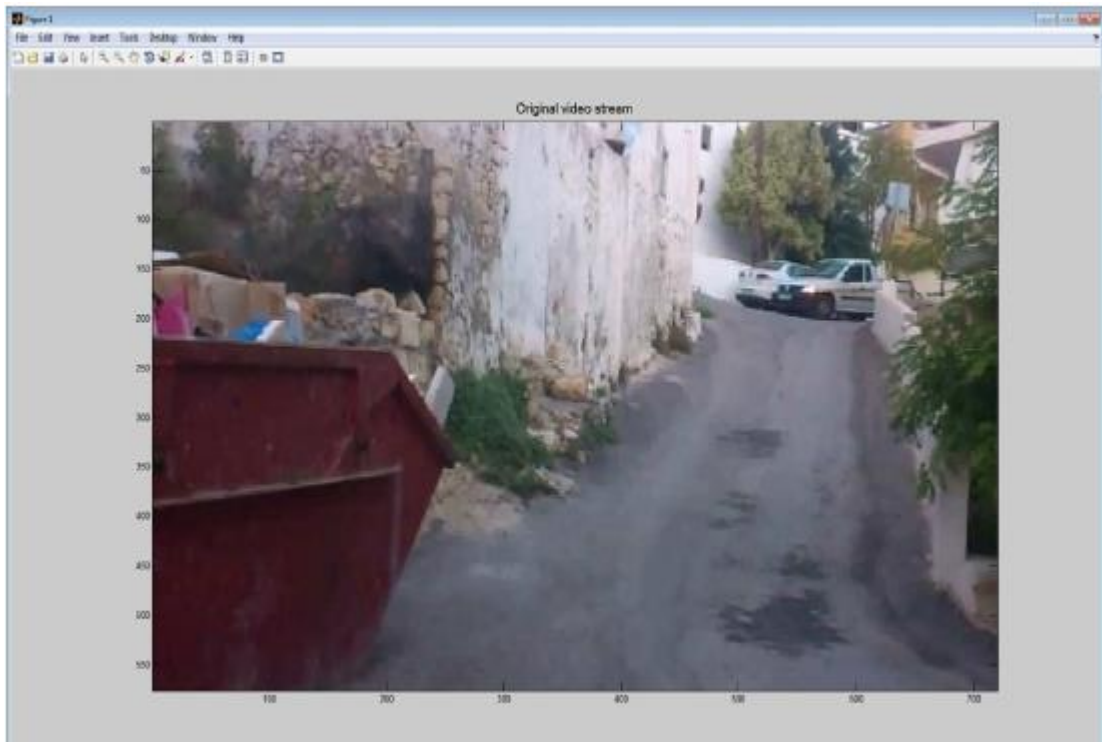
Βλέπουμε ότι η αφαίρεση του background είναι αρκετά αποτελεσματική, παρόλες τις εναλλαγές του φωτισμού και ο αλγόριθμος εντοπισμού ριχελ στο χρώμα του δέρματος είναι αρκετά ικανοποιητικός βοηθώντας τις να φτάσουμε σε ορθή ανίχνευση ανθρώπου.



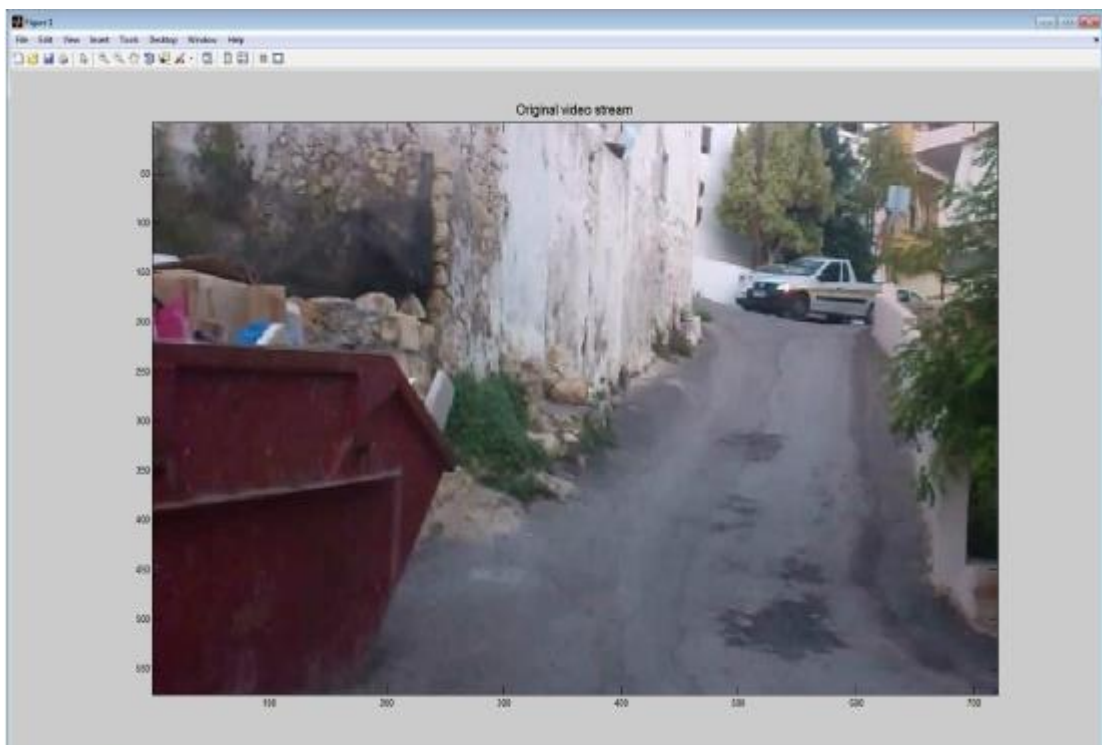
Εικόνα 5.47: Εντοπισμός pixel στο χρώμα του δέρματος.

Στα παραπάνω βίντεο είδαμε την αποδοτικότητα του αλγορίθμου σε εσωτερικούς χώρους. Παρατηρήσαμε ότι καταφέραμε να εντοπίσουμε ανθρώπινη παρουσία, πολύ αποτελεσματικά, σε τις τις περιπτώσεις. Θα συνεχίσουμε δοκιμάζοντας την εφαρμογή σε εξωτερικούς χώρους, σε ένα λιγότερο ελεγχόμενο περιβάλλον με αρκετούς αντιπερισπασμούς.

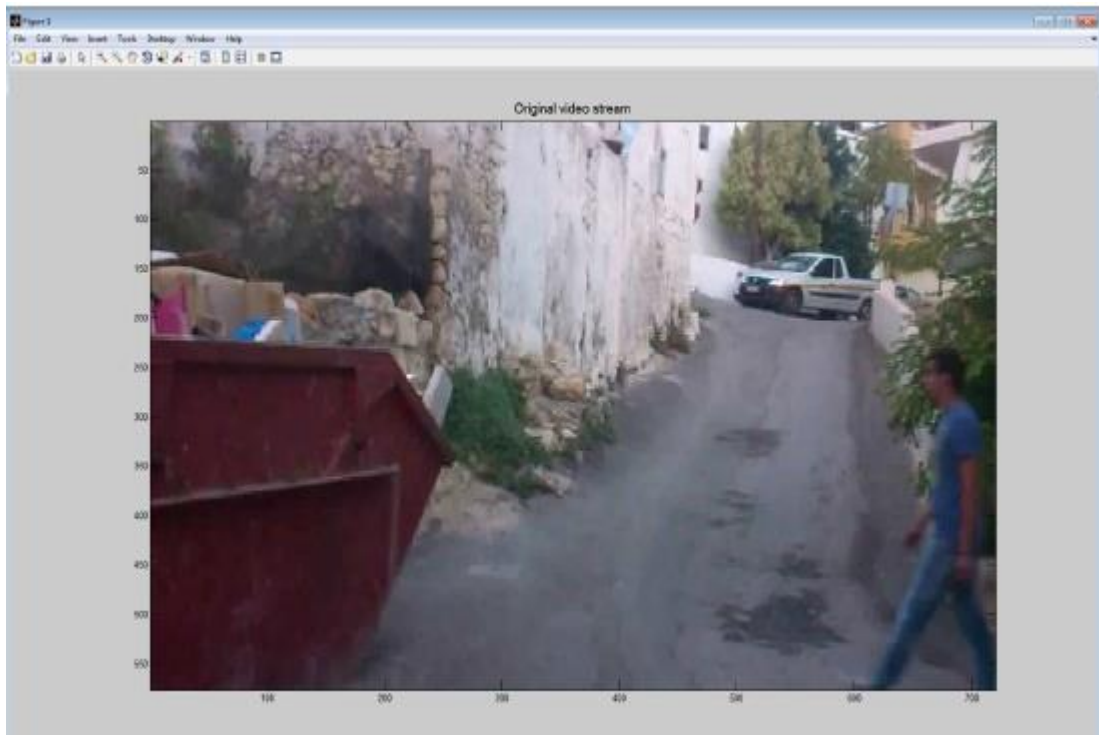
Στη συγκεκριμένη σκηνή βλέπουμε ότι, εκτός από την κίνηση του ανθρώπου έχουμε και την κίνηση διερχόμενου οχήματος. Ωστόσο, η εφαρμογή καταφέρνει να προσπεράσει τη μετακίνηση του οχήματος χωρίς να φτάσει σε εσφαλμένη αναγνώριση.



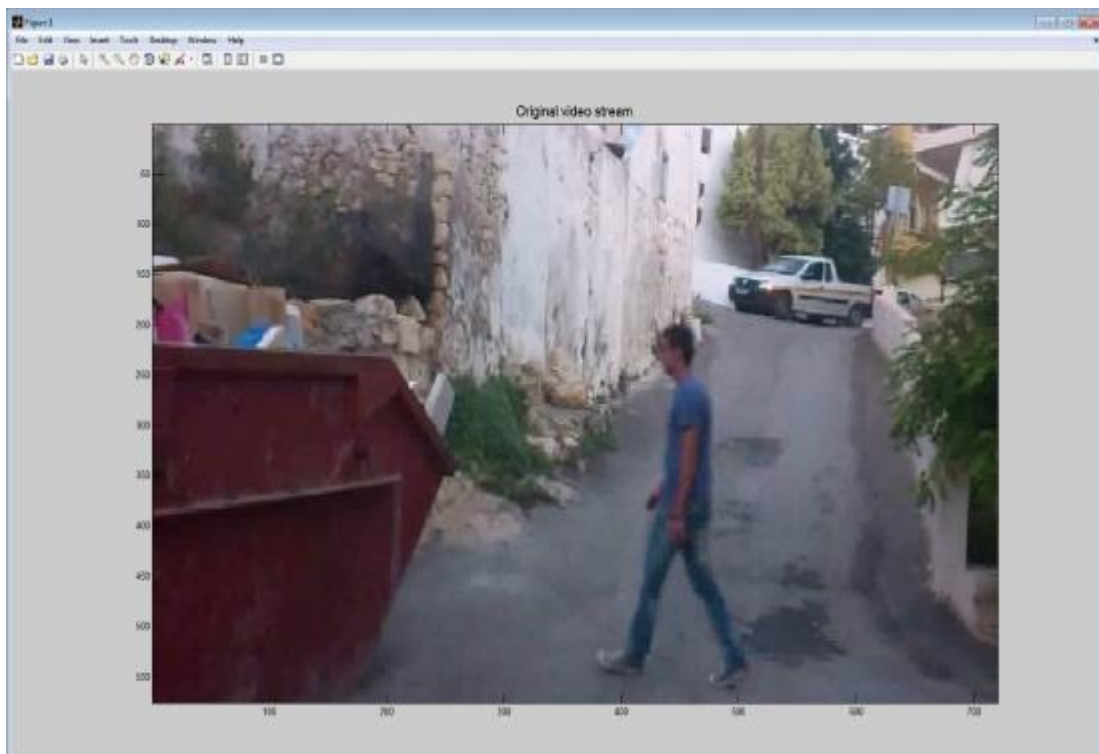
Εικόνα 5.48: Στιγμιότυπο 1 από την αναπαραγωγή βίντεο6.



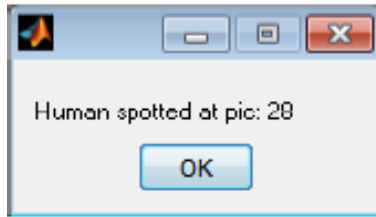
Εικόνα 5.49: Στιγμιότυπο 2 από την αναπαραγωγή βίντεο6.



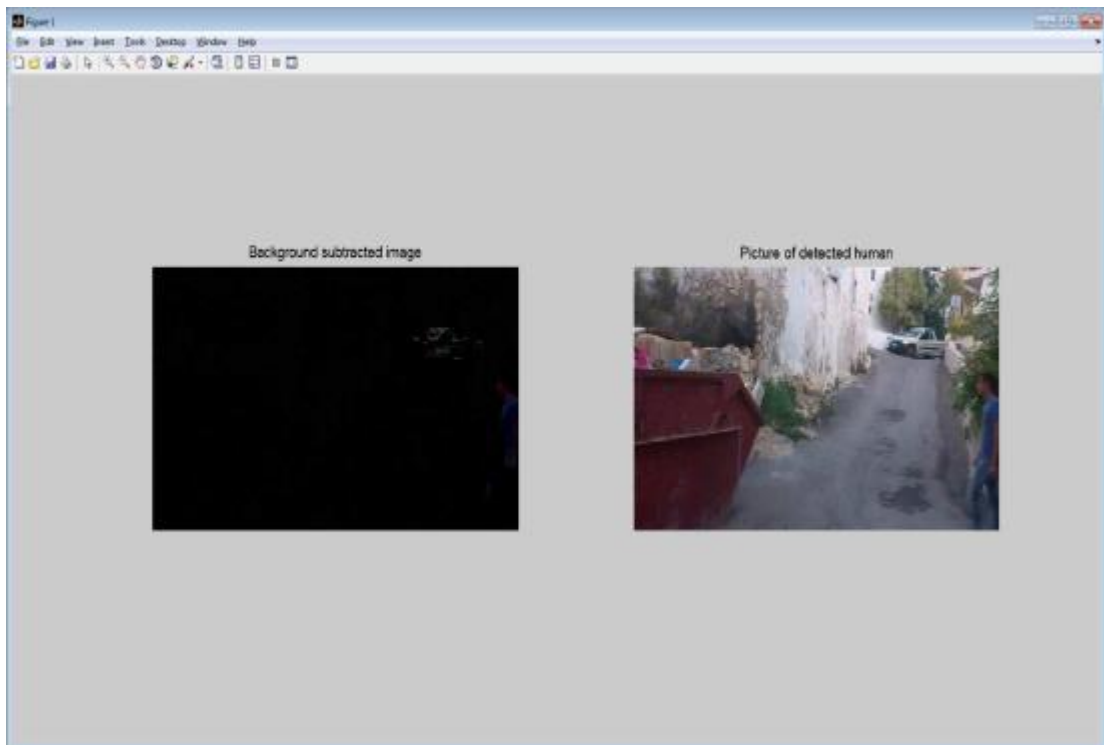
Εικόνα 5.50: Στιγμιότυπο 3 από την αναπαραγωγή βίντεο6.



Εικόνα 5.51: Στιγμιότυπο 4 από την αναπαραγωγή βίντεο6.



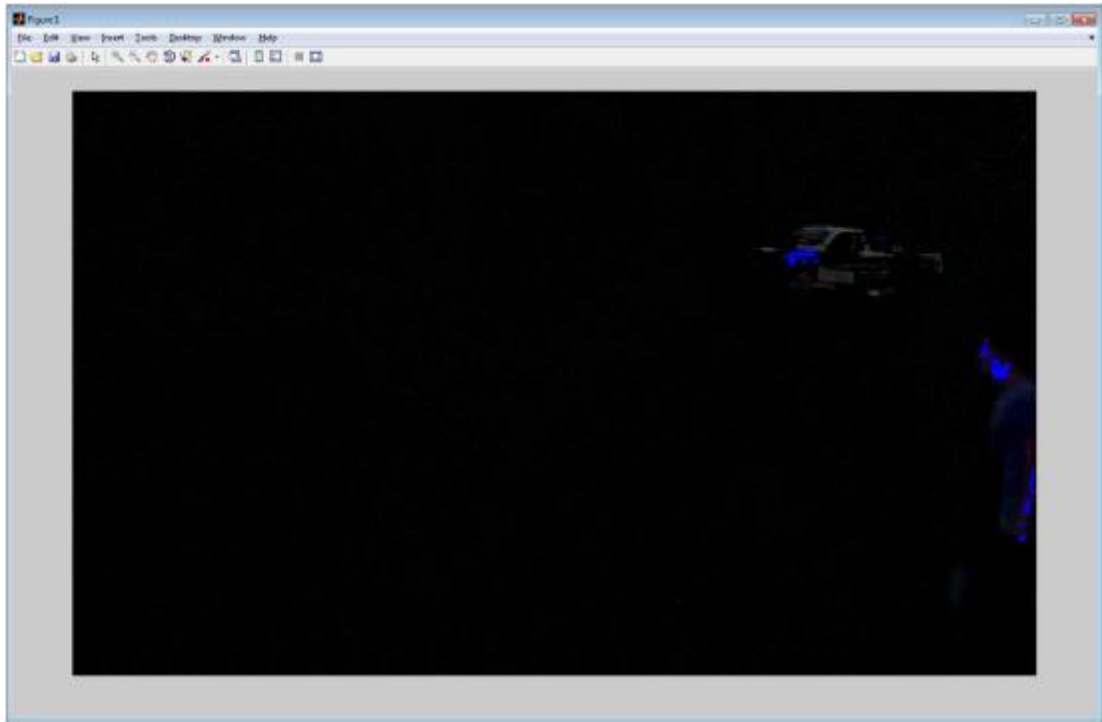
Εικόνα 5.52: Μήνυμα ανίχνευσης ανθρώπου στο βίντεο6.



Εικόνα 5.53: Η εικόνα που εντοπίστηκε ανθρώπινη παρουσία στο βίντεο6 με και χωρίς αφαίρεση Background.

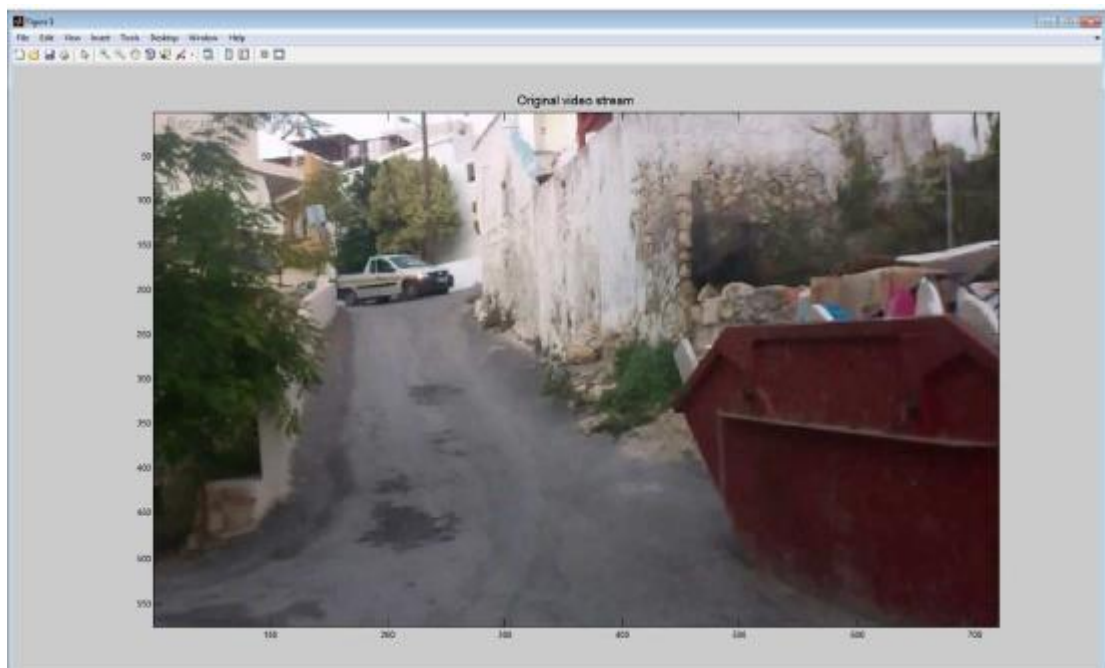
Βλέπουμε στο συγκεκριμένο βίντεο ότι η εφαρμογή είναι αρκετά ελαστική με την μετακίνηση αντικειμένων που δεν παρουσιάζουν ενδιαφέρον, όπως αυτή των φύλλων των δέντρων και η μετακίνηση του οχήματος αρκεί να μην έχουν χρώμα παρόμοιο με αυτό του ανθρώπινου δέρματος.

Στην εικόνα που ακολουθεί βλέπουμε ότι εντοπίζει pixels που είναι στο χρώμα του δέρματος αλλά αυτά δεν είναι αρκετά σε αριθμό για να προκαλέσουν εσφαλμένη ανίχνευση.

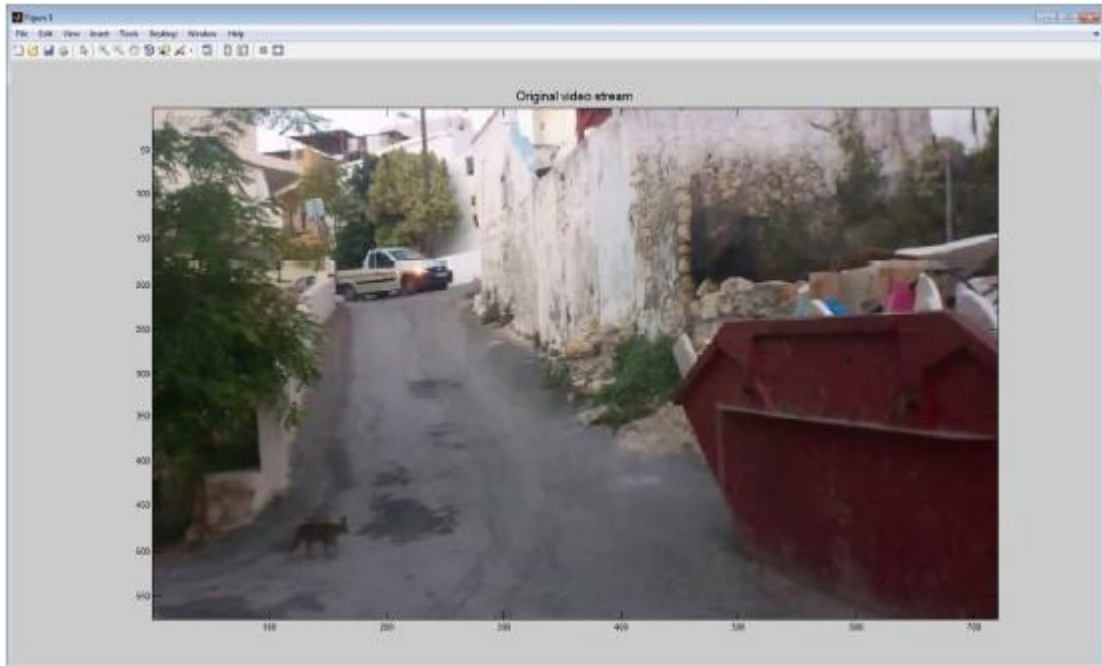


Εικόνα 5.54: Εντοπισμός ριxel στο χρώμα του δέρματος για το βίντεο6.

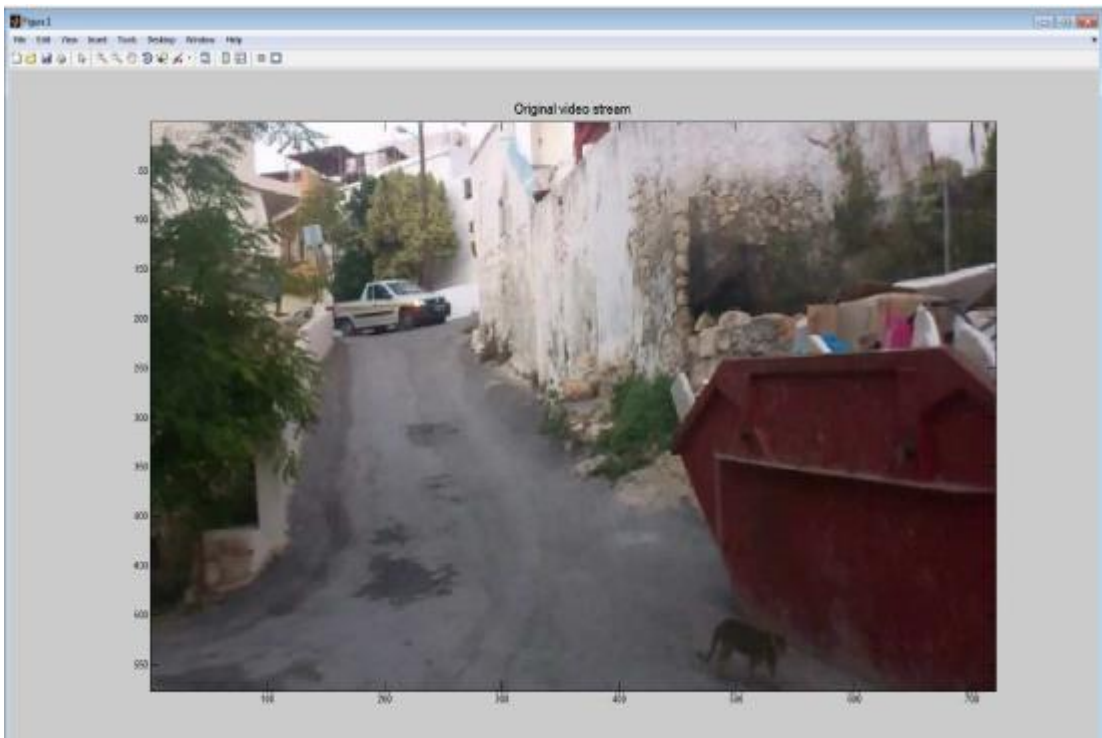
Τελειώνοντας να επισημάνουμε ότι η εφαρμογή δεν επηρεάζεται από μικρές αλλαγές στο foreground (προσκήνιο), όπως είναι κινήσεις που προκαλούνται από μικρά ζώα. Στο παράδειγμα που ακολουθεί έχουμε δυο αλλαγές που προκύπτουν στην εικόνα. Η πρώτη είναι από τα φώτα ενός αυτοκινήτου και η δεύτερη από ένα ζώο που περνάει από το πλάνο.



Εικόνα 5.55: Στιγμιότυπο 1 από την αναπαραγωγή βίντεο7.

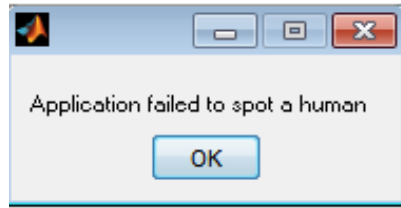


Εικόνα 5.56: Στιγμιότυπο 2 από την αναπαραγωγή βίντεο7.

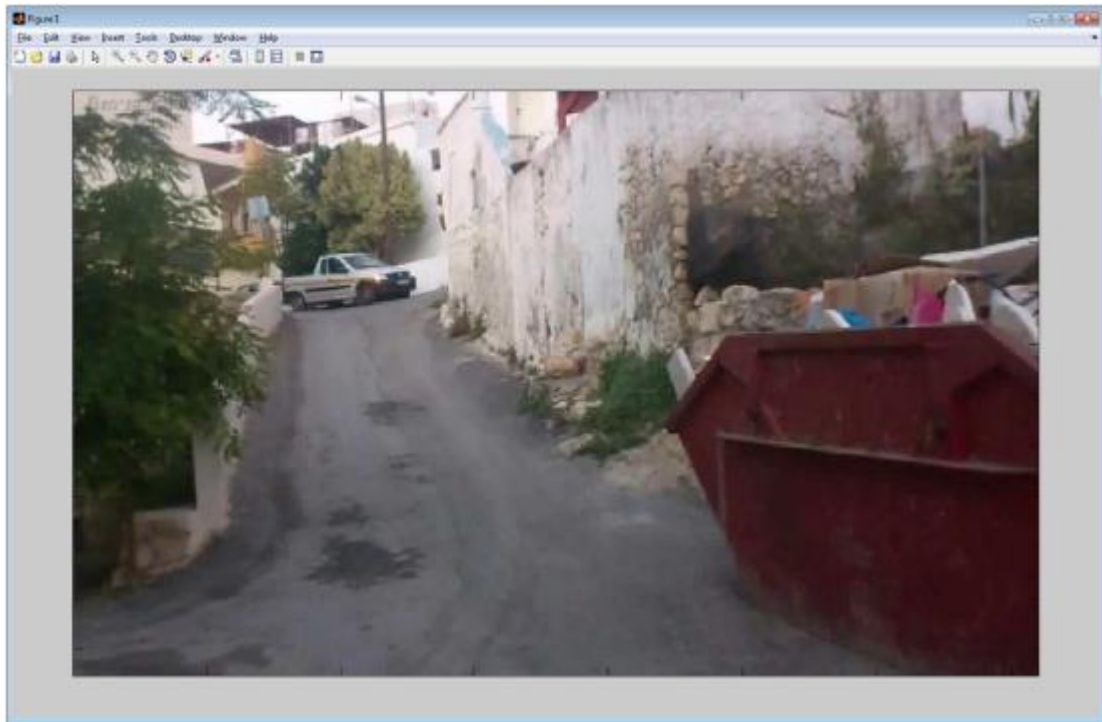


Εικόνα 5.57: Στιγμιότυπο 3 από την αναπαραγωγή βίντεο7.

Η εφαρμογή δεν εντοπίζει ανθρώπινη παρουσία και εμφανίζει το παρακάτω μήνυμα καθώς και την αρχική εικόνα.



Εικόνα 5.58: Μήνυμα αντίχτυσης ανθρώπου στο βίντεο7.



Εικόνα 5.59: Εικόνα επιστροφής εφαρμογής σε περίπτωση που δεν εντοπιστεί άνθρωπος.

ΣΥΜΠΕΡΑΣΜΑΤΑ

Στην παρούσα πτυχιακή εργασία πραγματοποιήθηκε έρευνα πάνω στις τεχνολογίες και τους τρόπους που χρησιμοποιούνται προκειμένου να επιτευχθεί ανίχνευση ανθρώπινης παρουσίας.

Οι τεχνολογίες που αναλύσαμε εκτενέστερα ήταν η ανίχνευση προσώπου, η ανίχνευση δέρματος, η ανίχνευση κίνησης και τέλος ο εντοπισμός χαρακτηριστικών σημείων.

Έγινε μια περιγραφή του Matlab το οποίο επιλέχθηκε σαν εργαλείο για τη δημιουργία της εφαρμογής μας και τέλος υλοποιήσαμε έναν αλγόριθμο ο οποίος βασίζεται στην αφαίρεση background και στην ανίχνευση δέρματος προκειμένου να εντοπίσει ανθρώπινη παρουσία σε ένα βίντεο.

Σαν άνθρωποι είμαστε ικανοί πολύ εύκολα να αναγνωρίζουμε αντικείμενα που έχουμε ξαναδεί αλλά και να μάθουμε καινούργια αντικείμενα. Μπορούμε να τα αναγνωρίσουμε ανεξάρτητα από το φωτισμό ή την οπτική γωνία που τα κοιτάμε, το ίδιο εύκολα μπορούμε να αναγνωρίσουμε και ανθρώπους σε ένα βίντεο. Αποδείχθηκε ότι δεν είναι τόσο απλό για ένα πρόγραμμα να κάνει το ίδιο πράγμα, καθώς υπάρχουν πολλές μεταβλητές που επηρεάζουν την απόδοσή του.

Καταφέραμε να επιτύχουμε το στόχο μας δημιουργώντας μια απλή εφαρμογή η οποία είναι ικανή να ανιχνεύσει με υψηλό ποσοστό ακρίβειας ανθρώπινη παρουσία σε μερικώς ελεγχόμενα περιβάλλοντα και να διαχειριστεί με επιτυχία κάποιες από τις πιο συνήθεις μεταβολές που μπορούν να εμφανιστούν σε ένα βίντεο.

ΒΙΒΛΙΟΓΡΑΦΙΑ

- [1] Aggarwal J.K., Cai Q., Human Motion Analysis: A Review, *Computer Vision and Image understanding*, vol. 73, no. 3, pp. 428-440, 1999.
- [2] Meghna S., *Robust Tracking and Human Activity Recognition*, Master of Science University of Alberta, 2004.
- [3] Yilmaz A., Javed O., and Shah M. 2006., Object tracking: A survey. *ACM Comput. Surv.* 38, 4, Article 13, 2006.
- [4] Ζουρνής-Καρούζος Ν., Τεχνικές ανίχνευσης κίνησης σε προκαθορισμένες περιοχές της εικόνας σε πραγματικό χρόνο, Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης, 2007.
- [5] Meha J. P., Bhumika B., A Comparative Study of Object Tracking Techniques, *International Journal of Innovative Research in Science, Engineering and Technology*, 2015.
- [6] Aherne F., Thacker N., and Rockett P., The Bhattacharyya metric as an absolute similarity measure for frequency coded data, *Kybernetika*, vol. 34, no. 4, pp. 363–368, 1998.
- [7] Rosenhahn B., Kersting U. G., He1 L., etc, A Silhouette Based Human Motion Tracking System, *Communication and Information Technology Research Technical Report*.
- [8] Kalman R., A new approach to linear filtering and prediction problems, *Journal of Basic Engineering*, pp. 35–45, 1960.
- [9] Negenborn R., *Robot Localization and Kalman Filters On finding your position in a noisy world*, Diploma Thesis, Utrecht University, 2003.