

ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΜΑΚΕΔΟΝΙΑΣ
ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

**Κατασκευή Εφαρμογής Ανίχνευσης Κινούμενων
Αντικειμένων ή Αντικειμένων που Εναποτέθηκαν,
με Χρήση Όρασης Μηχανής Πραγματικού Χρόνου**

Όνομα Φοιτητή:
Σκρέτας Αστέριος

Επιβλέπων Καθηγητής:
Dr. ΔΑΣΥΓΓΕΝΗΣ ΜΗΝΑΣ

Εργαστήριο Ψηφιακών Συστημάτων και Αρχιτεκτονικής Υπολογιστών

8 ΝΟΕΜΒΡΙΟΥ, ΚΟΖΑΝΗ 2017

Στους γονείς μου

Περίληψη

Αντικείμενο της παρούσας διπλωματικής εργασίας είναι ο αυτοματοποιημένος εντοπισμός αντικειμένων που εγκαταλείπονται από τον κάτοχό τους σε πραγματικό χρόνο, με χρήση τεχνικών μηχανικής όρασης. Ο συγκεκριμένος στόχος είναι μείζονος σημασίας σε εποχές που υφίσταται έξαρση της τρομοκρατίας, της εγκληματικότητας και άλλων παραβατικών συμπεριφορών. Για το λόγο αυτό το πεδίο της μηχανικής όρασης έχει προσελκύσει ιδιαίτερα το ενδιαφέρον της επιστημονικής κοινότητας. Μεγάλη πρόκληση του πεδίου αποτελεί η ανίχνευση συμβάντων (event detection) όπως αυτή της εγκατάλειψης ενός αντικειμένου όπου το πρόβλημα δεν είναι εύκολο να καθοριστεί αυστηρά. Το σύστημα που αναπτύχθηκε κάνει χρήση μιας μεθόδου ανίχνευσης φόντου διπλού μοντέλου σε συνδυασμό με μια μηχανή πεπερασμένων καταστάσεων (PFSM) για τον εντοπισμό αντικειμένων που ακινητοποιούνται και τέλος η επαλήθευση εγκατάλειψής τους μέσω σημασιολογικής ανάλυσης που στηρίζεται σε οπίσθια παρακολούθηση (Back-tracing Verification). Το σύστημα περιλαμβάνει επίσης χωροχρονικούς περιορισμούς (spatial-temporal constraints) καθώς και την προϋπόθεση της αλληλεπίδρασης ανθρώπου-αντικείμενου. Επιπροσθέτως, παρέχεται η δυνατότητα στον χρήστη του προσδιορισμού της ελάχιστης απόστασης του κατόχου από

το αντικείμενο, του εύρους μεγέθους του υποψήφιου εγκαταλελειμμένου αντικειμένου και του ελάχιστου χρονικού διαστήματος που θεωρείται εγκατάλειψη. Η ρύθμιση των παραμέτρων του συστήματος έγινε μέσω μιας αλληλουχίας προκαταρκτικών πειραμάτων. Στη συνέχεια για την επαλήθευση της ορθής λειτουργίας του συστήματος εκτελέστηκαν πειράματα τα οποία χωρίζονται σε 2 κατηγορίες. Η πρώτη κατηγορία αποτελείται από τη σύγκριση της απόδοσης της εφαρμογής α) της μεθόδου διπλού φόντου, β) του συνδυασμού της με το (PFSM) και γ) του συνδυασμού των α) και β) με την επαλήθευση οπίσθια παρακολούθησης. Η δεύτερη κατηγορία εξετάζει διαφορετικές περιπτώσεις ρύθμισης του συστήματος από τον χρήστη. Τα αποτελέσματα ανέδειξαν τον συνδυασμό των τριών προαναφερθέντων υποσυστημάτων ως την καλύτερη μέθοδο ανίχνευσης εγκαταλελειμμένου αντικειμένου. Στη δεύτερη κατηγορία πειραμάτων αποδεικνύεται ότι το σύστημα ανταποκρίνεται ικανοποιητικά στις διαφορετικές ρυθμίσεις από τον χρήστη.

Λέξεις Κλειδιά: ανίχνευση φόντου διπλού μοντέλου, ανίχνευση συμβάντων, μηχανή πεπερασμένων καταστάσεων, οπίσθια παρακολούθηση, χωροχρονικός περιορισμός.

Abstract

A major challenge in the field of image pattern recognition is the recognition of the so-called "event detection". Such a task is particularly difficult due to dynamic environment and a multitude of variables that render the processing procedure extremely non deterministic. The hereby thesis makes use of one particular methodology, the double background model in combination with a state machine in order to detect items that stay at rest. Finally, the verification procedure involves semantic analysis which is based on backtrack following. Our system involves a spacial-temporal constraints as well as the presupposition of the human-object interaction. Furthermore, a series of detection choices are provided to the user such as, the distance between human and object, the size range of the object and the minimum time span after which a left-behind object is registered as such. The system parameter initialization was established by a series of preliminary experiments. Next, the proper functional verification was implemented by running experiments that can be grouped into two categories: Firstly a cross-check between the three algorithmic procedures was performed a) Double background methodology b) PFSM and c) the combination of the former with the latter one. The second category examines various case experiments whose parameters are

set directly by the user. The results showed that the combination of the three methodologies to produce the best results for determining a left-behind object. In the second category of the experiments, an interesting result emerges, which suggests that the system responds adequately to the user parameter input. Overall the outcome of the work can be considered a success since it can provide the expected results up to a certain level of uncertainty.

Keys Words: event detection, PFSM, spatial-temporal constraints computer vision, backtracing verification

Δήλωση Πνευματικών Δικαιωμάτων

Δηλώνω ρητά ότι, σύμφωνα με το άρθρο 8 του Ν. 1599/1986 και τα άρθρα 2,4,6 παρ. 3 του Ν.1256/1982, η παρούσα Διπλωματική Εργασία με τίτλο ΑΝΙΧΝΕΥΣΗ ΕΓΚΑΤΑΛΕΛΕΙΜΜΕΝΟΥ ΑΝΤΙΚΕΙΜΜΕΝΟΥ καθώς και τα ηλεκτρονικά αρχεία και πηγαίοι κώδικες που αναπτύχθηκαν ή τροποποιήθηκαν στα πλαίσια αυτής της εργασίας και αναφέρονται ρητώς μέσα στο κείμενο που συνοδεύουν, και η οποία έχει εκπονηθεί στο Τμήμα Μηχανικών Πληροφορικής και Τηλεπικοινωνιών του Πανεπιστημίου Δυτικής Μακεδονίας, υπό την επίβλεψη του μέλους του Τμήματος κ.Δασυγένη Μηνά αποτελεί αποκλειστικά προϊόν προσωπικής εργασίας και δεν προσβάλλει κάθε μορφής πνευματικά δικαιώματα τρίτων και δεν είναι προϊόν μερικής ή ολικής αντιγραφής, οι πηγές που χρησιμοποιήθηκαν περιορίζονται στις βιβλιογραφικές αναφορές και μόνον. Τα σημεία όπου έχω χρησιμοποιήσει ιδέες, κείμενο, αρχεία ή / και πηγές άλλων συγγραφέων, αναφέρονται ευδιάκριτα στο κείμενο με την κατάλληλη παραπομπή και η σχετική αναφορά περιλαμβάνεται στο τμήμα των βιβλιογραφικών αναφορών με πλήρη περιγραφή. Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα. Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και μόνο.

Copyright (C)

Όνοματεπώνυμο Φοιτητή: ΣΚΡΕΤΑΣ ΑΣΤΕΡΙΟΣ

Επιβλέποντα: Δρ. ΔΑΣΥΓΕΝΗΣ ΜΗΝΑΣ

Έτος: 2017

Πόλη: ΚΟΖΑΝΗ

Υπογραφή Φοιτητή

Ευχαριστίες

Για την υλοποίηση της παρούσας διπλωματικής εργασίας ευχαριστώ πολύ τον επιβλέποντα καθηγητή μου, κύριο Δασυγένη Μηνά που χωρίς τη συμβολή του δεν θα ήταν εφικτή η διεκπεραίωση της εργασίας. Επίσης ευχαριστώ την οικογένεια μου, που με την στήριξή της ολοκλήρωσα έναν σημαντικό ακαδημαϊκό κύκλο της ζωής μου.

Περιεχόμενα

1	Εισαγωγή	1
1.1	Μέσα Οπτικής Παρατήρησης	2
1.2	Επιδιώξεις από την Επιτήρηση ενός Χώρου	3
1.3	Ανοικτά Προβλήματα Προκλήσεις	4
1.4	Μέθοδοι και τεχνικές εντοπισμού εγκαταλελειμμένων αντικειμένων	5
1.5	Εφαρμογές - Κίνητρα	7
1.6	Σύνοψη	9
2	Εισαγωγικές Έννοιες	11
2.1	Βίντεο	11
2.2	Φόντο-Προσκήνιο	15
2.3	Σύνοψη	17
3	Ανίχνευση κίνησης αντικειμένων σε βίντεο	19
3.1	Εισαγωγή	19
3.2	Προκλήσεις	20
3.3	Αφαίρεση Φόντου	21
3.4	Αφαίρεση φόντου μέσω στατιστικών μεθόδων	24

3.5	Μοντέλο Μείξης Γκαουσιανών (Mixture of Gaussians) . .	26
3.5.1	Εισαγωγή	26
3.6	Μαθηματική περιγραφή του Μοντέλου	28
3.6.1	Ανίχνευση φόντου με MMΓ	35
3.7	Σύνοψη	42
4	Μεθοδολογία	43
4.1	Εισαγωγή	43
4.2	Ανίχνευση προσκηνίου	44
4.3	Συνδυασμός μακροπρόθεσμου και βραχυπρόθεσμου μοντέλου φόντου	47
4.4	Μοντέλο μηχανής πεπερασμένων καταστάσεων βασισμένη στα εικονοστοιχεία	52
4.5	Επαλήθευση οπίσθιας παρακολούθησης	54
4.6	Σημειολογική ανάλυση εγκαταλελειμμένου αντικειμένου	57
4.7	Εκτέλεση Μεθοδολογίας	58
4.7.1	Με είσοδο βίντεο	59
4.7.2	Με είσοδο βίντεο πραγματικού χρόνου	64
4.7.3	Με είσοδο βίντεο και ορισμό διαφορετικών MAXSFG-MINSFG	67
4.8	Σύνοψη	69
5	Πειραματική Διαδικασία	71
5.1	Σύνοψη	76
6	Σύνοψη - Συμπεράσματα	77
7	Μελλοντική Εργασία	79

Κατάλογος σχημάτων

1.1	Κατηγορίες Καμερών	3
2.1	Αναπαράσταση Εικονοστοιχείων	12
2.2	RGB	13
2.3	Καρέ Βίντεο.	16
3.1	(Αριστερά) Μοντελοποίηση των δεδομένων με κατανομή Γκάους. (Δεξιά) Μοντελοποίησή τους με ΜΜΓ 2 συνιστωσών.	27
3.2	Μοντέλο μείξης Γκαουσιανών 2 συνιστωσών.	28
4.1	Αλγόριθμος ανίχνευσης προσκηνίου.	44
4.2	Μοντέλο Μηχανής Πεπερασμένων Καταστάσεων	51
4.3	Εισαγωγή Πηγής Εισόδου Βίντεο	59
4.4	Επιλογή Περιοχής ενδιαφέροντος	59
4.5	Αριστερά: Βραχυπρόθεσμο Μοντέλο $n+5$, Δεξιά: Μακροπρόθεσμο $n+5$	60
4.6	Αριστερά: Βραχυπρόθεσμο Μοντέλο $n+10$, Δεξιά: Μακροπρόθεσμο $n+10$	60

4.7	Αριστερά: Βραχυπρόθεσμο Μοντέλο $n+5$, Δεξιά: Μακροπρόθεσμο $n+5$	61
4.8	Αριστερά: Βραχυπρόθεσμο Μοντέλο $n+10$, Δεξιά: Μακροπρόθεσμο $n+10$	61
4.9	Αριστερά: Βραχυπρόθεσμο Μοντέλο $n+20$, Δεξιά: Μακροπρόθεσμο $n+20$	62
4.10	Ορισμός στατικού αντικειμένου	63
4.11	Αριστερά:Εμφάνιση Μηνύματος Συναγερμού Δεξιά: Σύνοψη	64
4.12	Αρχικό Φόντο	65
4.13	Αριστερά: RGB Απεικόνιση, Κέντρο: Βραχυπρόθεσμο Μοντέλο $n+5$ Δεξιά: Μακροπρόθεσμο $n+5$	65
4.14	Αριστερά: Βραχυπρόθεσμο Μοντέλο $n+5$, Δεξιά: Μακροπρόθεσμο $n+5$	66
4.15	Αριστερά: Βραχυπρόθεσμο Μοντέλο $n+5$, Δεξιά: Μακροπρόθεσμο $n+5$	66
4.16	Αριστερά: Στατικό Αντικείμενο, Δεξιά: Σύνοψη αντικειμένου	67
4.17	Αριστερά: Καρέ Εγκατάλειψης, Δεξιά: Χτύπημα Συναγερμού	67
4.18	Αριστερά: Βραχυπρόθεσμο Μοντέλο, Κέντρο: Μακροπρόθεσμο Μοντέλο, Δεξιά: Στατικό αντικείμενο	68
4.19	Αριστερά: Βραχυπρόθεσμο Μοντέλο, Κέντρο: Μακροπρόθεσμο Μοντέλο, Δεξιά: Στατικό αντικείμενο	68
5.1	Είσοδος Βίντεο	74

5.2	Αριστερά: RGB μοντέλο Κέντρο: Μακροπρόθεσμο Μοντέλο Δεξιά: Βραχυπρόθεσμο Μοντέλο	74
5.3	Αριστερά: RGB μοντέλο Κέντρο: Μακροπρόθεσμο Μοντέλο Δεξιά: Βραχυπρόθεσμο Μοντέλο	75
5.4	Χαρακτηρισμός περιοχής ως στατικό αντικείμενο	75
5.5	Ορισμός αντικειμένου ως εγκαταλελειμμένο	76



Κεφάλαιο 1

Εισαγωγή

Τα τελευταία χρόνια αυξάνεται όλο και πιο πολύ η χρησιμοποίηση συστημάτων ασφαλείας επιτήρησης χώρου (κάμερες παρακολούθησης). Δημόσιοι χώροι όπως αεροδρόμια, σταθμοί μετρό, εμπορικά κέντρα, αθλητικές εγκαταστάσεις αποτελούν σημεία μετακίνησης μεγάλου αριθμού ανθρώπων, κάνοντας επιτακτική την συνεχή παρακολούθησή τους. Ο κλάδος που απασχολεί αυτό το πεδίο, υπάγεται σε αυτόν της μηχανικής όρασης (computer vision). Η μηχανική όραση σχετίζεται με την θεωρία και την τεχνολογία που εμπλέκονται στη σχεδίαση και κατασκευή συστημάτων που αναλύουν δεδομένα από ψηφιακές εικόνες. Τα εν λόγω δεδομένα μπορούν να είναι φωτογραφίες, βίντεο, καρτέ εικόνων. Η μηχανική όραση επιδιώκει να εφαρμόσει θεωρίες και μοντέλα στην κατασκευή μηχανικών συστημάτων με δυνατότητα όρασης. Ένα τέτοιο σύστημα στον κλάδο αυτό αποτελεί και η ανίχνευση συμβάντων στο χώρο (οπτική παρατήρηση) όπως είναι και η εγκατάλειψη ενός αντικειμένου από τον κάτοχό του.

1.1 Μέσα Οπτικής Παρατήρησης

Υπάρχουν πολλά είδη καμερών που δύναται να χρησιμοποιηθούν (εικόνα 1.1). Για να εποπτευθεί ένας χώρος για λόγους ασφαλείας απαιτείται ένα σύστημα εποπτείας. Ένα σύστημα εποπτείας στην πιο απλή του μορφή αποτελείται από μία κάμερα και μία οθόνη, μεταξύ τους συνδεδεμένα. Ολοκληρωμένο σύστημα εποπτείας χώρων (CCTV) ονομάζεται ένα σύστημα που αποτελείται από τις συσκευές που εποπτεύουν τον χώρο (κάμερες), τις συσκευές καταγραφής (DVRs), τις συσκευές διαχείρισης και διάφορα παρελκόμενα που απαιτούνται. Τα συστήματα χωρίζονται σε ενσύρματα και ασύρματα. Επίσης χωρίζονται σε Αναλογικά και IP συστήματα. Συνήθως το κάθε είδος απαιτεί διαφορετικές συσκευές (κάμερες και DVRs). Υπάρχουν βέβαια και ειδικές συσκευές με την χρήση των οποίων μπορούμε να διασυνδέσουμε συστήματα διαφορετικής κατηγορίας. Δηλαδή μπορούμε να συνδέσουμε μια αναλογική κάμερα σε ένα IP σύστημα ή μια IP κάμερα σε ένα αναλογικό σύστημα. Επίσης μπορούμε να συνδέσουμε μια ασύρματη κάμερα σε ένα ενσύρματο σύστημα ή μια ενσύρματη κάμερα σε ένα ασύρματο σύστημα. Υπάρχουν επίσης και υβριδικά DVRs, στα οποία μπορούμε να συνδέσουμε κάμερες όλων των κατηγοριών. Επίσης υπάρχει διαχωρισμός στις κάμερες ανάλογα με την μορφή τους ή τον τρόπο στήριξής τους. Αυτές οι κατηγορίες είναι οι παρακάτω: Κάμερες Θόλου (DOME)-Κάμερες με βάση αδιάβροχες - Κάμερες Σώμα (δεν περιλαμβάνουν φακό)-Κάμερες ειδικού τύπου-Κρυφές κάμερες -Κινούμενες κάμερες. Όλα τα παραπάνω είδη καμερών μπορεί να είναι: Αναλογικές, IP, Ασύρματες ή Ενσύρματες. Επίσης όλα τα είδη μπορεί να έχουν ενσωματωμένα LED

υπερύθρων για να φωτίζουν την νύκτα. Αυτές οι κάμερες χαρακτηρίζονται IR κάμερες. Ο φωτισμός του χώρου με IR LED δεν είναι ορατός με γυμνό μάτι. Υπάρχουν και προβολείς Υπερύθρων, των οποίων η απόσταση που φωτίζουν είναι ανάλογη με το πλήθος ή την κατηγορία των LEDs που έχουν.



Σχήμα 1.1: Κατηγορίες Καμερών

1.2 Επιδιώξεις από την Επιτήρηση ενός Χώρου

Κάθε χώρος επιτηρείται για διαφορετικό σκοπό αναλόγως με τις απαιτήσεις του παρατηρητή. Σε κάποιους χώρους επιδιώκεται να προσμετράται το πλήθος των ατόμων που τους διαπερνούν, σε άλλους η συχνότητα ροής, ο όγκος των ατόμων και των αντικειμένων, αν κάποιο αντικείμενο σταματάει να κινείται ή αν κάποιος κάτοχος ενός αντικειμένου εγκαταλείψει το τελευταίο στην περιοχή που επιτηρείται και απο-

μακρυνθεί από εκείνη. Ένα εγκαταλειφθέν αντικείμενο σε μια περιοχή ευρείας μετακίνησης ατόμων αποτελεί υποψήφια απειλή για την ασφάλεια του χώρου. Έχουν για αυτό το λόγο αναπτυχθεί αρκετά συστήματα ασφαλείας που προσπαθούν να διαπιστώνουν τέτοιου τύπου απειλές. Για το λόγο αυτό είναι αναγκαία η ανάπτυξη μεθόδων που έχουν ως στόχο την ταξινόμηση των αντικειμένων σε δυο κατηγορίες, προσκηνίου (foreground) και φόντου (background). Ο ορισμός είναι βέβαια ασαφής δεδομένου ότι δεν υπάρχει ευθεία τοποθέτηση για το πότε ένα αντικείμενο εγκαταλείπεται πραγματικά. Παρόλα αυτά έχουν προταθεί αρκετές εναλλακτικές συστηματικές προσεγγίσεις που επιδιώκουν το προαναφερθέν. Κάποιες στηρίζονται σε απλούστερες παραδοχές, όπως η διακοπή κίνησης του αντικειμένου, άλλες που συνδυάζουν σημασιολογική ανάλυση και άλλες χρησιμοποιώντας χωρο-χρονικές πληροφορίες. Στη παρούσα διπλωματική εργασία προτείνεται μια υβριδική μέθοδος για τον εντοπισμό και για την επαλήθευση ενός εγκαταλελειμμένου αντικειμένου.

1.3 Ανοικτά Προβλήματα Προκλήσεις

Η μεγαλύτερη πρόκληση στην ανίχνευση αντικειμένων είναι η κατηγοριοποίηση, ενός αντικειμένου ως εγκαταλελειμμένο [1]. Παρουσιάζονται πολλές δυσκολίες για την κατηγοριοποίηση, ιδιαιτέρως όταν το αντικείμενο καλύπτεται μερικώς ή πλήρως από φυσικά ή τεχνητά εμπόδια όπως δέντρα, κτίρια, άλλα κινούμενα αντικείμενα, παρεμβολή θορύβου, καιρικές συνθήκες, επίπεδα φωτισμού, η πολυκοσμία σε ένα χώρο και άλλα. Επιπροσθέτως υπάρχει και το ζήτημα πότε ένα αντικείμενο θεω-

ρείται εγκαταλελειμμένο στο χώρο πραγματικά και πότε όχι. Υπάρχει η πιθανότητα κάτι να αφηθεί στο χώρο προσωρινά, από τον ιδιοκτήτη του και το σύστημα να το εκλάβει ως ύποπτο αντικείμενο λανθασμένα. Για τους παραπάνω λόγους έχουν ερευνηθεί και μελετηθεί διάφορες τεχνικές για την αντιμετώπιση των συγκεχυμένων καταστάσεων.

1.4 Μέθοδοι και τεχνικές εντοπισμού εγκαταλελειμμένων αντικειμένων

Υπάρχουν πολλές προσεγγίσεις για να επιτευχθεί ο εντοπισμός των αντικειμένων όπως είναι η συνεχής ροή και η κατάτμηση του φόντου με βάση τα οποία μπορεί το φόντο και το προσκήνιο να διαχωριστούν και στη συνέχεια να συγκριθούν για να εξαχθούν συμπεράσματα. Με το διαχωρισμό αυτό, αντικείμενα που έμειναν στάσιμα για κάποιο χρονικό διάστημα θέτουν υποψηφιότητα να χαρακτηριστούν ως εγκαταλελειμμένα, απομακρυσμένα αντικείμενα ή ως ακινητοποιημένα για κάποια χρονική στιγμή μέχρι να ξεκινήσουν εκ νέου να κινούνται στο χώρο. Η αφαίρεση του φόντου είναι μια πιο ευρεία διαδεδομένη τεχνική που εξετάζει την πιθανότητα κάποιο αντικείμενο να εγκαταλείφθηκε. Γι αυτό το λόγο πάνω σε αυτό το πεδίο έχουν μελετηθεί και εφαρμοστεί διάφορες τεχνικές και μεθοδολογίες. Μια από αυτές είναι ο εντοπισμός εγκαταλελειμμένων αντικειμένων με μια μέθοδο βασισμένη σε εικονοστοιχεία (pixel) που χρησιμοποιεί διπλά προσκήνια για την εξαγωγή περιοχών στατικής εικόνας [2]. Ανάλογα με την εφαρμογή αυτές οι περιοχές υποδεικνύουν αντικείμενα που δεν αποτελούν το αρχικό φόντο

αλλά μεταφέρθηκαν στη σκηνή σε μεταγενέστερο χρόνο.

Άλλες προσεγγίσεις βασίζονται στην αφαίρεση του φόντου το οποίο μοντελοποιείται σε τρία Γκαουσιανά μοντέλα τα οποία χρησιμοποιούνται για την ανίχνευση στατικών περιοχών [3]. Άλλη μια μέθοδος για την ανίχνευση εγκαταλελειμμένων αντικειμένων είναι η επιτήρηση του αντίστοιχου χώρου με χιλιάδες κάμερες. Η προσέγγιση αυτή βασίζεται στη μοντελοποίηση χρονικά στατικών αντικειμένων με την χρήση Μοντέλων Μείξης Γκαουσιανών (Gaussian mixtures) το οποίο θα αναλυθεί εκτενώς στη συνέχεια. Μέσω αυτής της προσέγγισης γίνεται ορισμένη τμηματοποίηση του φόντου εντοπίζοντας ταυτόχρονα το στατικό προσκήνιο. Ανάλογα με το μέγεθος του αντικειμένου και το χρόνο που παρέμεινε ανενεργό το αντικείμενο χαρακτηρίζεται ως ανενεργό [4]. Ακόμα μια μεθοδολογία που έχει χρησιμοποιηθεί είναι η ανίχνευση δυο μοντέλων φόντου που εκπαιδεύονται με διαφορετικούς ρυθμούς και ταξινομεί τα εικονοστοιχεία με τη βοήθεια μιας μηχανής πεπερασμένων καταστάσεων. Το φόντο διαμορφώνεται από δυο μείγματα γκαουσιανών μοντέλων με πανομοιότυπες παραμέτρους εκτός του ρυθμού εκμάθησης [5].

Χρησιμοποιήθηκε για την εκπόνηση της διπλωματικής εργασίας λειτουργικό σύστημα Windows 10. Το compile του κώδικα έγινε με το πρόγραμμα Microsoft Visual Studio χρησιμοποιώντας την βιβλιοθήκη OpencV.

Στην παρούσα διπλωματική εργασία θα γίνει λόγος για την μεθοδολογία που χρησιμοποιήθηκε για τον επιτυχή εντοπισμό αντικειμένου που έχει αφεθεί σε ένα χώρο από τον ιδιοκτήτη του. Αρχικά θα γίνει μια αναφορά στις διάφορες τεχνικές που υπάρχουν για να επιτευχθεί το επιθυμητό αποτέλεσμα. Στη συνέχεια θα κάνουμε λεπτομερή ανα-

φορά στη σειρά των τεχνικών που εφαρμόστηκαν στην εργασία για την επίτευξη του στόχου. Τέλος εξήχθησαν τα συμπεράσματα από αυτή την διαδικασία και το κατά πόσο η μεθοδολογία που εφαρμόστηκε είναι ακριβής.

1.5 Εφαρμογές - Κίνητρα

Το σύστημα εντοπισμού που υλοποιήθηκε για τον εντοπισμό αντικειμένων που εγκαταλείφθηκαν στο χώρο δύναται να έχει πολλές εφαρμογές σε πολλούς τομείς και πτυχές της καθημερινότητας. Η πιο διαδεδομένη χρήση του μπορεί να λάβει χώρα σε δημόσιους χώρους, κτίρια, αεροδρόμια όπου εξαιτίας της μεταφοράς μεγάλου όγκου ανθρώπων δυσχεραίνεται η επιτήρηση τους. Έτσι αν ένα αντικείμενο σε τέτοιους χώρους βρίσκεται για κάποιο χρονικό διάστημα χωρίς την επιτήρησή του κατόχου του, ένας συναγερμός (alarm) θα ήταν επιθυμητός για τον άμεσο εντοπισμό και απομάκρυνσή του. Το ίδιο χρήσιμο είναι για την καταγραφή της κίνησης στους αυτοκινητόδρομους μιας πόλης, όπου εγκαταλελειμμένα αντικείμενα πέτρες, ξύλα, κορμοί δέντρων θα μπορούσαν να δημιουργήσουν προβλήματα στην κυκλοφορία και στην ασφαλή οδήγηση. Επιπρόσθετα σε χώρους όπως οι πίστες των αεροδρομίων που η επιτήρηση τους από τον ανθρώπινο παράγοντα είναι δύσκολη, ένα τέτοιο σύστημα θα διευκόλυνε την έγκαιρη απομάκρυνση αντικειμένων που θα μπορούσαν να δημιουργήσουν προβλήματα στην άρτια τεχνική λειτουργία των αεροπλάνων. Επίσης στο χώρο της υγείας θα αποτελούσε χρήσιμο εργαλείο στα χέρια των γιατρών. Είναι γνωστό, ότι πολλές χειρουργικές επεμβάσεις πραγματοποιούνται ενδοσκοπικά

με χρήση κάμερας. Έχουν υπάρξει περιπτώσεις όπου γιατροί άφησαν κατά λάθος ιατρικά εργαλεία εντός του σώματος του ασθενή. Έτσι με την εφαρμογή μιας τέτοιας μεθόδου στις ενδοσκοπικές επεμβάσεις θα περιορίζονται δραστικά η εμφάνιση τέτοιων περιστατικών, αφού ο χειρουργός θα ειδοποιούνταν ότι έχει εγκαταλείψει λανθασμένα κάποιο του εργαλείο σε περιοχή μη αποδεκτή. Σε χώρους στρατοπέδων ακόμα θα ήταν χρήσιμη η εγκατάσταση ενός τέτοιου συστήματος παρακολούθησης. Το μοντέλο θα εκπαιδευόταν με τους χώρους του στρατοπέδου με αποτέλεσμα, αν καταγραφόταν για παράδειγμα τις βραδυνές κυρίως ώρες κάτι που δεν θα ταυτιζόταν με το γενικό φόντο του στρατοπέδου θα σήμαινε συναγερμό. Με αυτό τον τρόπο θα μειώνονταν σε ένα μεγάλο βαθμό η περιφρούρηση του στρατοπέδου από τον ανθρώπινο παράγοντα. Θα μπορούσε επίσης το συγκεκριμένο σύστημα να ενισχύσει και να αναβαθμίσει ήδη υπάρχοντα συστήματα ασφαλείας επιτήρησης χώρου. Σε κτίρια και δημόσιους χώρους όπου επιτηρούνται από μια πληθώρα καμερών ασφαλείας είναι αδύνατο το ανθρώπινο μάτι που παρακολουθεί τις καταγραφές τους να μπορεί να έχει ολική αντίληψη για το τι συμβαίνει στο κάθε χώρο ταυτόχρονα, ιδίως όταν αυτοί κατακλύζονται από πλήθος κόσμου. Έτσι λοιπόν ένα τέτοιο σύστημα θα πρόσφερε την βοήθεια στον άνθρωπο να ελέγχει γρήγορα και άμεσα σκηνές του χώρου όπου έχει σημάνει συναγερμός ότι κάποιο αντικείμενο εγκαταλείφθηκε. Τέλος, λόγω των πολλών τρομοκρατικών και εγκληματικών χτυπημάτων στις μέρες μας κρίνεται απαραίτητη και επιτακτική η ενίσχυση όλο και περισσότερων των συστημάτων ασφαλείας χώρου. Συμπερασματικά, θα λέγαμε ότι εξαιτίας των πολλών εφαρμογών της διαδικασίας ανίχνευσης αντικειμένου, τα τελευταία χρόνια έχει ενταθεί η έρευνα

στον τομέα αυτό που όμως εξακολουθεί να αποτελεί ένα ανοιχτό πρόβλημα της επιστημονικής περιοχής και μια ταυτόχρονα πρόκληση για τελειοποίηση των μεθόδων και των προγραμμάτων που θα διαθέτουν όσο το δυνατόν καλή απόδοση και μεγάλη ευχρηστία. Στα πλαίσια αυτής της προσπάθειας εντάσσεται και η παρούσα διπλωματική εργασία στην οποία περιγράφεται η μέθοδος και η τεχνική που υλοποιήθηκε για τον εντοπισμό κινούμενων και εγκαταλελειμμένων αντικειμένων.

Η εκπόνηση αυτής της εργασίας οδήγησε σε επισταμένη ενασχόληση με τον κλάδο της μηχανικής όρασης, επί του οποίου διενεργείται ολοένα και περισσότερο πληθώρα ερευνητικών δραστηριοτήτων. Ιδιαίτερο ενδιαφέρον παρουσίασε η ενασχόληση με την βιβλιοθήκη `openCV` η οποία παρέχει όλα τα χρήσιμα εργαλεία για την ανίχνευση κίνησης.

1.6 Σύνοψη

Συνοψίζοντας, για να εποπτευθεί ένας χώρος για λόγους ασφαλείας απαιτείται ένα σύστημα εποπτείας. Ο πιο διαδεδομένος τρόπος να εποπτευθεί ένας χώρος είναι η χρησιμοποίηση καμέρων. Μια από τις βασικότερες επιδιώξεις στο τομέα της ασφαλούς επιτήρησης μιας σκηνής είναι ο εντοπισμός αντικειμένων που εναποτέθηκαν σε μια περιοχή και εγκαταλείφθηκαν από τον ιδιοκτήτη τους. Στο συγκεκριμένο πεδίο της

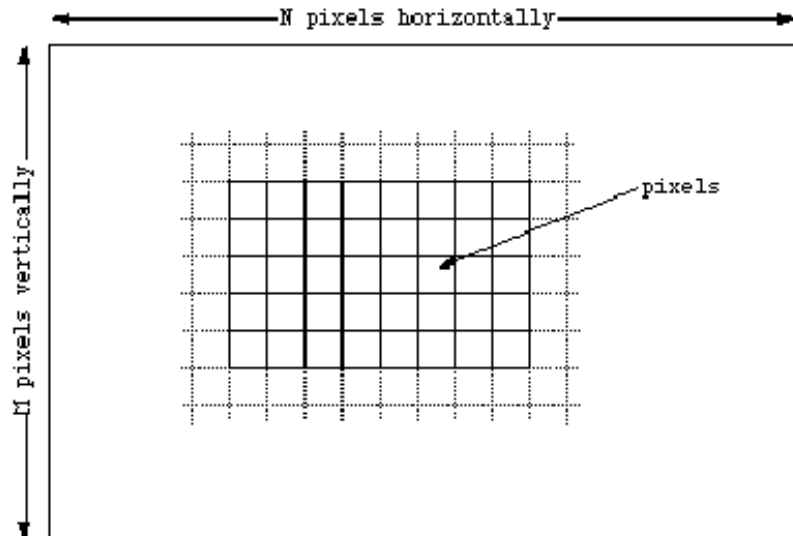
μηχανικής όρασης έχουν χρησιμοποιηθεί πολλές τεχνικές και μέθοδοι εντοπισμού εγκαταλελειμμένων αντικειμένων. Μέθοδοι όπως η κατάτμηση της εικόνας, η χρήση γκαουσιανών μοντέλων για την ανίχνευση στατικών αντικειμένων και η ανίχνευση δυο μοντέλων φόντου που εκπαιδεύονται με διαφορετικούς ρυθμούς και ταξινομεί τα εικονοστοιχεία με τη βοήθεια μιας μηχανής πεπερασμένων καταστάσεων είναι από τις πιο διαδεδομένες. Υπάρχει πληθώρα εφαρμογών που χρησιμοποιείται καθώς επίσης πολλαπλά κίνητρα για την επίτευξη του εν λόγω στόχου.

Κεφάλαιο 2

Εισαγωγικές Έννοιες

2.1 Βίντεο

Η έννοια του βίντεο-κινούμενης εικόνας βασίζεται στη σύνθεση συνεχών εικόνων-καρέ που έχουν ληφθεί από το εκάστοτε μέσο λήψης όπως μια φωτογραφική μηχανή ή βιντεοσκοπική κάμερα. Μια εικόνα απαρτίζεται από πολλά σημεία που ονομάζονται εικονοστοιχεία (pixel) και αποτελούν τα θεμελιώδη στοιχεία της εικόνας. Στην οθόνη ενός υπολογιστή οι εικόνες αναπαρίστανται με υποδιαίρεση σε ένα δυσδιάστατο πίνακα. Κάθε κελί του πίνακα ονομάζεται εικονοστοιχείο και εμπεριέχει μια χρωματική τιμή (2.1).



Σχήμα 2.1: Αναπαράσταση Εικονοστοιχείων

Η εικόνα αναλύεται με βάση ένα ορθογώνιο πλέγμα το (bitmap) που αποτελείται από οριζόντιες σειρές με μικρές υποδιαιρέσεις. Κάθε εικονοστοιχείο του πλέγματος αναπαρίστανται με τη θέση του στο πίνακα αυτό με τις τιμές x, y . Ένα απόσπασμα βίντεο αποτελείται από διαδοχικές εικόνες, οι οποίες προβάλλονται γρήγορα και έτσι τα αντικείμενα εμφανίζονται ως κινούμενα. Οι διαδοχικές εικόνες, που αποτελούν το απόσπασμα βίντεο, ονομάζονται πλαίσια ή καρέ (frames). Η εναλλαγή των καρέ δεν πραγματοποιείται με τον ίδιο ρυθμό σε όλα τα υπολογιστικά συστήματα και μέσα καταγραφής. Ο ρυθμός καρέ (frame rate per second, fps) είναι η συχνότητα (ταχύτητα) με την οποία εμφανίζονται διαδοχικές εικόνες. Ο όρος μπορεί επίσης να ονομάζεται συχνότητα πλαισίου και να εκφράζεται σε hertz. Ο ρυθμός αυτός είναι σημαντικός γιατί καθορίζει την ομαλότητα της εικόνας. Η χρονική ευαισθησία και η ανάλυση της ανθρώπινης όρασης ποικίλει ανάλογα με τον τύπο και

τα χαρακτηριστικά των οπτικών ερεθισμάτων και διαφέρει μεταξύ των ατόμων. Το ανθρώπινο οπτικό σύστημα μπορεί να επεξεργαστεί 1 έως 5 εικόνες ανά δευτερόλεπτο και να τις αντιληφθεί μεμονωμένα.

Επιπρόσθετες σημαντικές παράμετροι του βίντεο είναι η ανάλυση (video resolution) που είναι η οριζόντια και κάθετη διάσταση της εικόνας. Εκφράζεται σε εικονοστοιχεία. Υπάρχουν διάφορες μορφές συστατικών σημάτων με βάση τα οποία αναπαράγεται η πληροφορία της εικόνας στο δέκτη. Η πιο ευρέως διαδεδομένη είναι η μορφή RGB.



Σχήμα 2.2: RGB

Αποτελεί ένα προσθετικό πρότυπο στο οποίο τα χρώματα κόκκινο, πράσινο, μπλε συνδυάζονται με διάφορους τρόπους για να αναπαράχθούν και τα άλλα χρώματα στην εικόνα. Ένα χρώμα στο πρότυπο χρώματος RGB (RED, GREEN, BLUE) μπορεί να περιγραφεί με το προσδιορισμό του πόσο κάθε ένα από το κόκκινο, πράσινο και μπλε χρώμα συμπεριλαμβάνεται. Κάθε ένα μπορεί να ποικίλει μεταξύ του ελάχιστου (καθόλου χρώμα) και του μεγίστου (πλήρης ένταση). Εάν όλα τα χρώματα είναι στο ελάχιστο, το αποτέλεσμα είναι μαύρο. Το μοντέλο αυτό απλοποιεί αρκετά την σχεδίαση γραφικών συστημάτων, αλλά δημιουργεί δυσκολίες στην ανάπτυξη αλγορίθμων λόγω της συ-

σχέτισης των χρωματικών συνιστωσών του. Για παράδειγμα η εξισορρόπηση ιστογράμματος λειτουργεί μόνο επί της φωτεινότητας μίας εικόνας και πραγματοποιείται πιο εύκολα στο μοντέλο HSI. Έτσι πολλές φορές είναι η αναγκαία η μετατροπή ενός χρωματικού μοντέλου σε ένα άλλο. Χρησιμοποιείται συνήθως σε έγχρωμες οθόνες και κάμερες. Η μετατροπή RGB σε κλίμακα γκρι γίνεται για το σύστημα PAL με τη σχέση: $Gray=0.259R + 0.600G + 0.117B$ ή μπορεί να δοθεί ως ο μέσος όρος των τριών χρωματικών συνιστωσών $Gray=0.333R + 0.333G + 0.333B$ Οι εικόνες ή τα καρέ που εξάγονται από ακολουθίες εικόνων είναι σήματα δύο διαστάσεων και στην αναλογική τους μορφή μπορούν να παρασταθούν από μια πραγματική συνάρτηση δύο μεταβλητών που θα απεικονίζει τη φωτεινότητα (τιμή έντασης) της εικόνας για την περίπτωση εικόνας grayscale. Η ψηφιακή εικόνα προκύπτει από δειγματοληψία της αναλογικής και στη συνέχεια κβαντισμό των εντάσεων των σημείων σε διακριτές τιμές.

Η κωδικοποίηση των ψηφιακών εικόνων είναι απαραίτητη ώστε να επιτυγχάνεται συμπίεση του όγκου των δεδομένων με σκοπό τη γρήγορη μετάδοση τους από διαύλους περιορισμένης χωρητικότητας ή τον περιορισμό των χώρων αποθήκευσής τους. Η συμπίεση μπορεί να επιτευχθεί αν γίνει εκμετάλλευση του πλεονασμού (redundancy) που υπάρχει σε κάθε εικόνα. Οι μέθοδοι κωδικοποίησης διακρίνονται σε αυτές που δεν εισάγουν σφάλμα κατά την αποκωδικοποίηση (lossless) και σε εκείνες που έχουν απώλειες (lossy). Μια εικόνα μπορεί να είναι δυαδική binary image ή μονοχρωματική με αποχρώσεις του γκρι (gray level-gray scale image) ή έγχρωμη. Μια ψηφιακή εικόνα αποχρώσεων του γκρι διαστάσεων $M*N$, παριστάνεται από έναν δισδιάστατο πίνακα ακεραίων

αριθμών: $I(i,j)$ με $i=1\dots M$ και $j=1\dots N$ όπου

$$0 \leq I(i,j) \leq L - 1 \quad (2.1)$$

. Το L ισούται συνήθως με μια δύναμη του 2, δηλαδή $L = 2^k$, με συνήθεστερη τιμή το $k = 8$, που αντιστοιχεί σε 256 αποχρώσεις του γκρι. Η τιμή $I(i,j)$ αντιπροσωπεύει τη φωτεινότητα του εικονοστοιχείου (i,j) . Η απλούστερη μορφή μιας εικόνας είναι η δυαδική μορφή. Μια δυαδική εικόνα έχει μόνο δύο στάθμες φωτεινότητας, που συνήθως είναι το μαύρο και το άσπρο $L = 2$ και $k = 1$. Το μαύρο αντιστοιχεί στην τιμή 0 και το άσπρο στην τιμή 1. Μια δυαδική εικόνα καταλαμβάνει μικρότερη μνήμη και η επεξεργασία της απαιτεί μικρότερο υπολογιστικό κόστος. Σε δυαδική μορφή μπορούν να απεικονισθούν σημαντικές πληροφορίες όπως είναι το εμβαδόν και η θέση των αντικειμένων, η μορφή αντικειμένων κ.α. Ακόμα μια έννοια που μας απασχολεί είναι και η ευκρίνεια μιας εικόνας. Το πόσο δηλαδή μπορούμε να βλέπουμε τις λεπτομέρειες της. Ισούται με το πλήθος των εικονοστοιχείων ανά μονάδα επιφανείας, και συνήθως μετριέται σε (pixels / in) ή διαφορετικά σε d.p.i (dots per inch). Είναι φανερό ότι η ευκρίνεια εξαρτάται τόσο από τις διαστάσεις, όσο και από το πλήθος των αποχρώσεων της κάθε εικόνας.

2.2 Φόντο-Προσκήνιο

Τα εικονοστοιχεία μιας εικόνας που καταγράφονται από μια κάμερα μπορούν να διακριθούν σε εικονοστοιχεία προσκήνιου και σε εικονοστοιχεία φόντου. Φόντο αποτελεί το σύνολο των εικονοστοιχείων που παρουσιάζουν μικρο-διακυμάνσεις στις τιμές τους για ένα ορισμένο χρονικό

διάστημα. Τα υπόλοιπα εικονοστοιχεία είναι εικονοστοιχεία προσκήνιου. Με άλλα λόγια είναι τα εικονοστοιχεία που μεταβάλλονται λόγω της κίνησης σε εκείνη την περιοχή. Στο παράδειγμα της εικόνας 2.3 γίνεται εύκολα αντιληπτό ότι τα εικονοστοιχεία του χώρου όπου δεν μεταβάλλονται οι τιμές τους αποτελούν το φόντο σε αντίθεση με τα εικονοστοιχεία που κάθε φορά απαρτίζουν τα κυκλωμένα άτομα και βρίσκονται σε μεταβολή εξαιτίας της κίνησης τους στο χώρο. Τα άτομα αυτά αποτελούν το προσκήνιο της εικόνας.



Σχήμα 2.3: Καρέ Βίντεο.

2.3 Σύνοψη

Συνοπτικά, σε αυτό το κεφάλαιο αναλύθηκε η έννοια του βίντεο-κινούμενης εικόνας. Δόθηκε ο ορισμός του εικονοστοιχείου, ο τρόπος με τον οποίο απεικονίζεται στην οθόνη ενός υπολογιστή, ο ορισμός του καρέ, και τα χαρακτηριστικά που το διέπουν όπως είναι ο ρυθμός ανανέωσης και η συχνότητα με την οποία εμφανίζονται διαδοχικές εικόνες. Η πληροφορία ενός εικονοστοιχείου αναπαράγεται με διάφορες μορφές συστατικών σημάτων με την πιο ευρέως διαδεδομένη την RGB. Βασικός διαχωρισμός για την ανίχνευση κινούμενων αντικειμένων αποτελεί ο διαχωρισμός του φόντου με το προσκήνιο.

Κεφάλαιο 3

Ανίχνευση κίνησης αντικειμένων σε βίντεο

3.1 Εισαγωγή

Τα αυτοματοποιημένα συστήματα παρακολούθησης έχουν διαδοθεί ευρέως, λόγω της εφαρμογής τους στο πεδίο της ασφάλειας και της άμυνας. Τα εν λόγω συστήματα είναι κατάλληλα για ένα ευρύ φάσμα εφαρμογών που αφορούν την ασφάλεια και την επιτήρηση δημόσιων χώρων. Παραδείγματα αποτελούν η παρακολούθηση αεροδρομίων, ανίχνευση κυκλοφοριακής συμφόρησης καθώς και η πρόβλεψή της, παρακολούθηση στάσεων Μαζικών Μέσων Μεταφοράς, μαζικές εκδηλώσεις, επιτήρηση συνόρων κ.α. Το θεμελιώδες πρόβλημα στην ανάλυση της κίνησης κατά την παρακολούθηση ενός βίντεο, είναι η ανίχνευση του κινούμενου αντικειμένου και η αφαίρεση του φόντου. Τα απαιτούμενα βήματα της διαδικασίας είναι η παρακολούθηση, η κατάτμηση και τέ-

λος η ανίχνευση των κινούμενων αντικειμένων στο βίντεο. Πρωταρχική απαίτηση αποτελεί ο διαχωρισμός των κινούμενων αντικειμένων από το φόντο. Έπειτα τα κινούμενα αντικείμενα πρέπει να εντοπιστούν μέσα σε διαφορετικά καρέ. Το κάθε πλαίσιο έχει συνήθως τη δική του ταυτότητα ή αλλιώς το δικό του αναγνωριστικό.

Αρχικά μια στατική κάμερα παρατηρεί μια σκηνή. Το αποτέλεσμα της καταγραφής αποτελεί την είσοδο στο σύστημα ανίχνευσης κινούμενου αντικειμένου. Η διαδικασία της ανίχνευσης του κινούμενου αντικειμένου ανήκει ως μέθοδος στο πεδίο επεξεργασίας εικόνας υψηλού επιπέδου.

3.2 Προκλήσεις

Το έργο της αξιόπιστης ανίχνευσης και παρακολούθησης κινούμενων αντικειμένων σε βίντεο παρακολούθησης, δεν έχει ακόμη επιλυθεί πλήρως και έχει πολλά ζητήματα που απαιτούν λύσεις. Πολλές συνθήκες οδηγούν σε κακής ποιότητας ανίχνευση φόντου. Κάποιες εξ αυτών είναι η αλλαγή του φωτισμού, οι επαναλαμβανόμενες κινήσεις και η άτακτη κίνηση αντικειμένων που εμφανίζονται στο προσκήνιο. Η ανακριβής ανίχνευση φόντου επηρεάζει την ανίχνευση κίνησης αντικειμένων στο βίντεο, ειδικά σε ανίχνευση σε πραγματικό χρόνο, οδηγώντας σε αντίστοιχα αναξιόπιστα αποτελέσματα. Επομένως, για να οδηγηθούμε σε αξιόπιστη ανίχνευση κινούμενου αντικειμένου απαιτείται η ακριβής ανίχνευση του φόντου [6][7]. Οι πιο σημαντικές προκλήσεις είναι[8]:

1. **Αλλαγή φωτεινότητας.** Ένα αξιόπιστο μοντέλο ανίχνευσης κινούμενου αντικειμένου πρέπει να προσαρμόζεται στις σταδιακές

αλλά και στις ξαφνικές αλλαγές της εμφάνισης του περιβάλλοντος. Επίσης απαιτείται συνεπής προσαρμογή και στις προοδευτικές αλλαγές φωτισμού, συμπεριλαμβανομένης της αλλαγής της έντασης φωτός σε εξωτερικούς χώρους κατά τη διάρκεια της ημέρας (π.χ., αλλαγές καιρού, πρωί-απόγευμα κτλ.).

2. **Πολυπλοκότητα φόντου.** Οι υπό παρακολούθηση σκηνές αποτελούνται συνήθως από δυναμικά αντικείμενα. Το φόντο μιας σκηνής μπορεί να αποτελείται από κινούμενα αντικείμενα όπως κινούμενα δέντρα, κινούμενες σημαίες, αυτοκίνητα κτλ.
3. **Κινούμενο αντικείμενο.** Εάν ένα αντικείμενο που ανήκει στο προσκήνιο εγκαταλείψει μια σκηνή, η περιοχή που καταλάμβανε στο πλαίσιο θα πρέπει πλέον να θεωρηθεί φόντο.
4. **Θόρυβος βίντεο.** Οι μεταβολές φωτεινότητας και χρώματος που μπορεί να είναι αποτέλεσμα σκίασης ή άλλων εξωτερικών παραγόντων, αποτελούν θόρυβο. Ενθόρυβα δεδομένα μπορούν να δημιουργηθούν λόγω της επίδρασης αισθητήρων αλλά και λόγω της συμπίεσης των δεδομένων (υποβάθμιση βίντεο) χάριν εξοικονόμησης πόρων.

3.3 Αφαίρεση Φόντου

Η παρακολούθηση μιας σκηνής συνιστά ένα βίντεο το οποίο είναι είσοδος στο σύστημα ανίχνευσης κίνησης αντικειμένων. Το πιο σημαντικό βήμα είναι η ανίχνευση και η αφαίρεση φόντου σε κάθε καρέ. Το πρόβλημα της τεχνικής είναι ότι παρουσιάζει πολύ μεγάλη ευαισθησία σε

δυναμικές αλλαγές του σκηνικού λόγω φωτισμού και άλλων εξωγενών παραγόντων. Το συγκεκριμένο πρόβλημα προσεγγίζεται με ποικίλες μεθόδους. Το προσκήνιο συνήθως αποτελείται από κινούμενα αντικείμενα όπως άνθρωποι, ζώα, οχήματα κτλ. Ο διαχωρισμός προσκηνίου και φόντου οδηγεί το σύστημα να εστιάζει σε συγκεκριμένες περιοχές των σκηνών και απλοποιεί τη διαδικασία επεξεργασίας του βίντεο.

Η διαδικασία αφαίρεσης φόντου εφαρμόζεται κυρίως σε βίντεο με στατικό φόντο ή σε συστήματα παρακολούθησης όπου η κάμερα είναι στατική.

Υπάρχουν διάφορες μέθοδοι ανίχνευσης στατικού φόντου. Μια κατηγορία μεθόδων εφαρμόζει αφαίρεση διπλού φόντου [9]. Σύμφωνα με τη συγκεκριμένη κατηγορία υπάρχουν δύο είδη φόντου, το μακροπρόθεσμο (long-term) και το βραχυπρόθεσμο (short-term) φόντο. Στην πρώτη περίπτωση όλα τα αντικείμενα που περιέχονται στο φόντο θεωρούνται μέρος της σκηνής και ότι δεν είναι στατικά. Το μακροπρόθεσμο φόντο αρχικοποιείται με το αποτέλεσμα της εφαρμογής διαμέσου (median) φίλτρου στα αρχικά καρέ. Στη συνέχεια το φόντο ανανεώνεται. Το μακροπρόθεσμο φόντο συλλαμβάνει μια πιο σταθερή αναπαράσταση της σκηνής και προσαρμόζεται αργά στις αλλαγές. Αντιθέτως, το βραχυπρόθεσμο φόντο προσαρμόζεται γρήγορα για να επιτρέψει υψηλής ευαισθησίας ανίχνευση. Αυτό το φόντο αποτελείται από τα πιο πρόσφατα δείγματα φόντου [10] και έχει ανανεωθεί με τμήματα της τρέχουσας εικόνας. Το προσκήνιο εντοπίζεται μέσω του υπολογισμού της διαφοράς των δύο φόντων. Η συγκεκριμένη μέθοδος δεν έχει καλή απόδοση σε περιπτώσεις όπου υπάρχουν στατικές σκηνές όπως ακίνητοι άνθρωποι, αυτοκίνητα, πλήθος ατόμων.

Μια άλλη κατηγορία μεθόδων ανίχνευσης φόντου επικεντρώνεται στη συλλογή των καρέ, τη δημιουργία δυαδικού προσκηνίου και στη συνέχεια στην παρακολούθηση των περιοχών που ανήκουν στο προσκηνίο. Με αυτόν τον τρόπο ανιχνεύεται ένα στατικό φόντο. Αυτό μπορεί να επιτευχθεί μέσω των εικονοστοιχείων που συσώρευσαν τις μέγιστες τιμές και μπορούν να θεωρηθούν στατικές περιοχές της εκάστοτε σκηνής [11].

Γενικά, το πρώτο καρέ που λαμβάνεται συνήθως χρησιμοποιείται ως εικόνα αναφοράς και θεωρείται **εικόνα φόντου αναφοράς**. Εικόνα φόντου αναφοράς μπορεί να προκύψει και από το μέσο όρο των εικόνων κατά τη διάρκεια του χρόνου λήψης των πρώτων καρέ. Η βασική ιδέα είναι ο εντοπισμός διαφορών μεταξύ των εικόνων φόντου και μιας τρέχουσας εικόνας. Η διαδικασία αφορά την αφαίρεση της τρέχουσας εικόνας από την εικόνα φόντου εικονοστοιχείο-προς-εικονοστοιχείο. Στη συνέχεια ορίζεται ένα κατώφλι και αν υπάρχουν εικονοστοιχεία των οποίων η τιμή ξεπερνά το κατώφλι, τότε θεωρούνται ως εικονοστοιχεία προσκηνίου. Το αποτέλεσμα είναι η δημιουργία μιας δυαδικής εικόνας όπου τα ενεργά εικονοστοιχεία επισημαίνονται με “1” και τα ανενεργά με “0”. Προκειμένου να βελτιωθεί η διαδικασία ανίχνευσης τμήματος της εικόνας που ανήκει στο προσκηνίο, εφαρμόζονται μορφολογικοί μετασχηματισμοί στα πλαίσια της επεξεργασίας, όπως η διαστολή (dilation) και η μορφολογική διάβρωση (morphological erosion [12]. Η εικόνα φόντου αναφοράς δεν είναι στατική αλλά ανανεώνεται στο χρόνο ώστε να προσαρμόζεται στις δυναμικές αλλαγές που συμβαίνουν στις σκηνές. Όπως προαναφέρθηκε υπάρχουν διάφορες τεχνικές για την αφαίρεση φόντου. Η βασική ιδέα στηρίζεται σε μια απλή και ταυτόχρονα δια-

δεδομένη τεχνική [13], σύμφωνα με την οποία ένα εικονοστοιχείο στην περιοχή (x, y) ενός καρέ θεωρείται ως προσκίνητο εαν:

$$| I_t(x, y) - B_t(x, y) | > T, \quad (3.1)$$

όπου I_t το τρέχον καρέ, η τρέχουσα εικόνα φόντου και T το προκαθορισμένο κατώφλι. Η ανανέωση της εικόνας φόντου γίνεται σύμφωνα με τη σχέση

$$B_{t+1} = (1 - a)B_t + aI_t, \quad (3.2)$$

όπου a είναι ένας συντελεστής προσαρμογής ο οποίος λαμβάνει τιμές από 0 μέχρι 1. Όσο μεγαλύτερη η τιμή του συντελεστή, τόσο γρηγορότερα ενημερώνεται το σύστημα για τις νέες αλλαγές στη σκηνή όσον αφορά το φόντο. Ωστόσο, ο συντελεστής δεν πρέπει να λαμβάνει πολύ μεγάλες τιμές, επειδή μπορεί να δημιουργηθούν τεχνητές "ουρές" πίσω από τα κινούμενα αντικείμενα [13]. Στη συνέχεια ομαδοποιούνται τα εικονοστοιχεία που τελικά ανήκουν στο προσκίνητο και εφαρμόζονται μορφολογικοί μετασχηματισμοί όπως άνοιγμα και κλείσιμο [14], για να εξαλειφθούν περιοχές μικρού μεγέθους. Η συγκεκριμένη τεχνική είναι ιδιαίτερα ευαίσθητη στην αλλαγή της φωτεινότητας και στις δυναμικές αλλαγές που υφίσταται το φόντο.

3.4 Αφαίρεση φόντου μέσω στατιστικών μεθόδων

Η στατιστική μοντελοποίηση του φόντου είναι μια δημοφιλής προσέγγιση στο πρόβλημα ανίχνευσης κινούμενου αντικειμένου. Σύμφωνα

με τη συγκεκριμένη κατηγορία μεθόδων τα ανεξάρτητα κινούμενα αντικείμενα που ανήκουν στο προσκήνιο διαχωρίζονται από το στατιστικό φόντο, χωρίς να απαιτείται εκ των προτέρων πληροφορία σχετικά με τις ιδιότητες του προσκηνίου. Αρκετές διαφορετικές μέθοδοι απαρτίζουν το πεδίο [15][16][17][18]. Η γενική ιδέα είναι ο ορισμός ενός μοντέλου φόντου ως προς τα εικονοστοιχεία. Για κάθε εικονοστοιχείο σε μια νέα εικόνα εξετάζεται εάν προέρχεται από το μοντέλο φόντου του συγκεκριμένου εικονοστοιχείου ή όχι. Σύμφωνα με μια από τις πρώτες ακριβείς στατιστικές προσεγγίσεις [19] προτείνεται η μοντελοποίηση της διακύμανσης που βρέθηκε σε ένα σύνολο εικόνων φόντου με την μέγιστη και ελάχιστη τιμή έντασης και τη μέγιστη διαφορά μεταξύ διαδοχικών καρέ. Το σύστημα υπολογιστικής όρασης Pfinder [20] χρησιμοποιεί το μέσο όρο και τη διακύμανση της τιμής των εικονοστοιχείων. Αν το μόνο που είναι γνωστό για μια κατανομή είναι ο μέσος όρος και η διακύμανση, η πιο λογική υπόθεση που βασίζεται στη μέγιστη εντροπία είναι η κανονική κατανομή ή αλλιώς η κατανομή Γκάους (Gauss). Η υπόθεση τότε είναι ότι η τιμή του εικονοστοιχείου ακολουθεί μια κατανομή Γκάους και χρησιμοποιείται ένα μοντέλο πιθανοφάνειας (likelihood) για να συγκριθεί η πιθανοφάνεια για το εάν εικονοστοιχείο ανήκει σε φόντο με την πιθανοφάνεια να ανήκει σε προσκήνιο. Άλλες μέθοδοι προτείνουν μη-παραμετρικά μοντέλα για την ανίχνευση του φόντου [21], σύμφωνα με τα οποία ενσωματώνονται χωρικοί περιορισμοί για τη διαμόρφωση της διαδικασίας ταξινόμησης προσκηνίου. Στη δεύτερη φάση της προσέγγισής αυτής, οι τιμές των εικονοστοιχείων που θα μπορούσαν να ερμηνευθούν σύμφωνα με τις κατανομές που ακολουθούν τα γειτονικά εικονοστοιχεία, ταξινομούνται εκ νέου ως φόντο, επιτρέποντας μεγα-

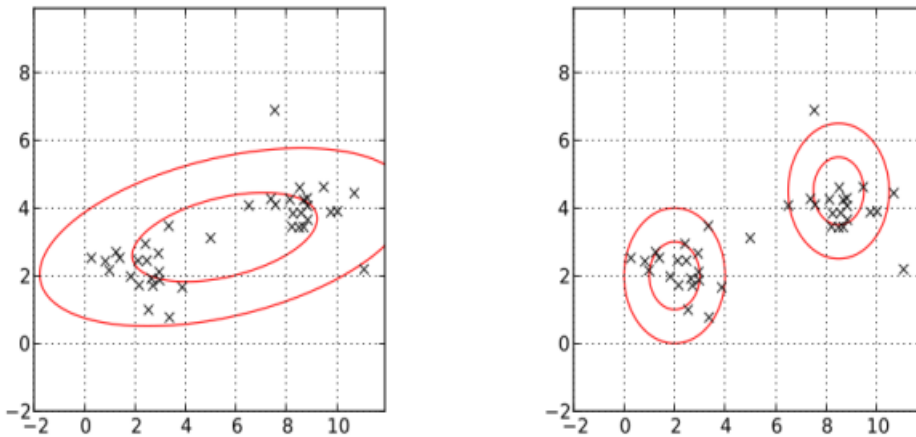
λύτερη ανθεκτικότητα σε δυναμικά φόντα. Οι Sheikh και Shah ενοποιούν τις χρονικές και χωρικές οντότητες σε ένα ενιαίο μοντέλο [22]. Το αποτέλεσμα είναι υψηλής ακρίβειας κατάτμηση των αντικειμένων ακόμη και αν υπάρξει δυναμικό φόντο. Μια διαφορετική προσέγγιση [23] λαμβάνει υπόψη γενικά στατιστικά χαρακτηριστικά αντί για τοπικούς περιορισμούς. Μια διαδεδομένη υποκατηγορία στατιστικής ανίχνευσης φόντου απο στηρίζεται στην ιδιοανάλυση όπως η Ανάλυση σε Κύριες Συνιστώσες (PCA) και ιδιοφόντου (eigenbackgrounds). Στόχος εδώ είναι να συλληφθεί από το μοντέλο η κυρίαρχη μεταβλητότητα του φόντου. Υπολογίζεται η διαφορά μεταξύ της αρχικής εικόνας και του τμήματος της εικόνας που ανήκει στο τμήμα του ιδιοφόντου. Αφού οριστεί κατώφλι διαχωρίζονται τα αντικείμενα εκείνα που ανήκουν στο φόντο από εκείνα που ανήκουν στο προσκήνιο.

3.5 Μοντέλο Μείξης Γκαουσιανών (Mixture of Gaussians)

3.5.1 Εισαγωγή

Μια ευρέως διαδεδομένη μέθοδος ανίχνευσης φόντου είναι το Μοντέλο Μείξης Γκαουσιανών (MMΓ). Είναι ένα πιθανοθεωρητικό μοντέλο που χρησιμοποιείται για την αναπαράσταση ενός κανονικά κατανοημένου υποπληθυσμού ενός συνολικού πληθυσμού. Το απλό μοντέλο Γκάους εφαρμόζεται στη μάθηση πιθανοτικών μοντέλων τα οποία λαμβάνουν τη μορφή απλών κατανομών λόγω της απλής φύσης των δεδομένων στα οποία εφαρμόζονται. Συχνά όμως τα δεδομένα που τίθενται προς

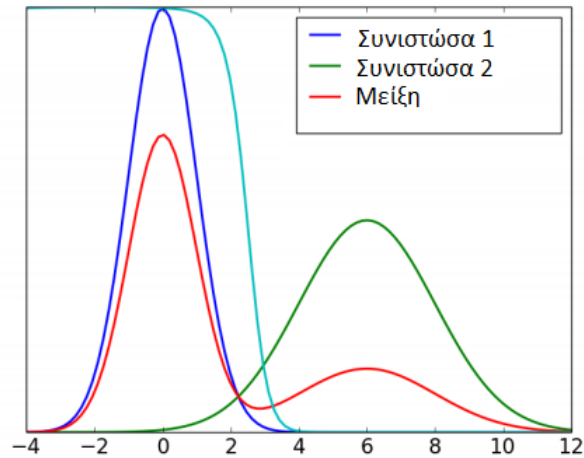
μοντελοποίηση είναι εξαιρετικά πολύπλοκα. Για παράδειγμα τα δεδομένα μπορεί να είναι πολυτροπικά (multimodal). Αυτό σημαίνει ότι οι κατανομές τους παρουσιάζουν δύο ή περισσότερες κορυφές (modes). Για παράδειγμα μπορεί να είναι δεδομένα που προέρχονται από δυο διαφορετικές διεργασίες που συμπεριφέρονται σύμφωνα με την κανονική κατανομή, αλλά τα διαστήματα των τιμών τους αλληλοεπικαλύπτονται. Για τη μοντελοποίηση τέτοιων δεδομένων χρησιμοποιείται το ΜΜΓ. Συγκεκριμένα τα δεδομένα μοντελοποιούνται ως μείξη διαφορετικών συνιστωσών (components), όπου κάθε συνιστώσα ακολουθεί μια κανονική κατανομή. Στο παράδειγμα που παρατέθηκε εφαρμόζεται ΜΜΓ δύο συνιστωσών. Η διαφορά μεταξύ της μοντελοποίησης αυτών των δεδομένων με κατανομή Γκάους και ΜΜΓ απεικονίζεται στο σχήμα 3.1



Σχήμα 3.1: (Αριστερά) Μοντελοποίηση των δεδομένων με κατανομή Γκάους. (Δεξιά) Μοντελοποίησή τους με ΜΜΓ 2 συνιστωσών.

Συνεπώς, θεωρείται για κάθε σημείο δείγματος ότι ανήκει σε μια από

τις συνιστώσες. Μια περίπτωση μονομεταβλητού ΜΜΓ δύο συνιστωσών απεικονίζεται στο παρακάτω σχήμα.



Σχήμα 3.2: Μοντέλο μείξης Γκαουσιανών 2 συνιστωσών.

Στόχος στη συνέχεια είναι να εξαχθεί συμπέρασμα για το είδος της κατανομής της κάθε συνιστώσας ξεχωριστά. Αυτή είναι μια διαδικασία που μαθαίνει μόνο του το μοντέλο. Δεδομένου ότι η ταξινόμηση των δειγμάτων στις αντίστοιχες συνιστώσες γίνεται αυτόματα χωρίς κάποια πληροφορία για τα δεδομένα, το μοντέλο συγκαταλέγεται στην κατηγορία της μάθησης χωρίς επίβλεψη (unsupervised learning).

3.6 Μαθηματική περιγραφή του Μοντέλου

Δεδομένου ότι γίνεται συναγωγή συμπερασμάτων για την κατάταξη των δεδομένων στην εκάστοτε συνιστώσα όπως προαναφέρθηκε, τα δεδομένα θεωρείται ότι περιγράφονται από λανθάνουσες μεταβλητές z .

Τέτοιου τύπου μεταβλητές δεν είναι ποτέ άμεσα παρατηρήσιμες. Οι τιμές τους συνάγονται από τον αλγόριθμο. Αντίθετα οι μεταβλητές που παρατηρούνται ευθέως και πάντα, καλούνται παρατηρήσεις και συμβολίζονται με x . Οι λανθάνουσες μεταβλητές αντιστοιχούν στις συνιστώσες μείξης. Οι τιμές που λαμβάνουν ανήκουν σε ένα διακριτό σύνολο $\{1, \dots, K\}$. Το πρώτο βήμα είναι η εφαρμογή δειγματοληψίας στις λανθάνουσες μεταβλητές z και στη ακολουθεί η δειγματοληψία των παρατηρήσεων x από μια κατανομή που εξαρτάται το z

$$p(z, x) = p(z)p(x | z), \quad (3.3)$$

όπου $p(z)$ η κατανομή Γκάους. Ο υπολογισμός της συνάρτησης πυκνότητας πιθανότητας των παρατηρήσεων x γίνεται με την εφαρμογή περιθωριοποίησης (marginalization) στο z .

$$p(x) = \sum_z p(z)p(x | z) \quad (3.4)$$

$$= \sum_{k=1}^K \Pr(z = k)p(x | z = k), \quad (3.5)$$

όπου K το πλήθος των συνιστωσών και $\Pr(z = k)$ η αναλογία μείξης. Η παραπάνω συνάρτηση πυκνότητας πιθανότητας αποτελεί ένα σταθμισμένο άθροισμα των συναρτήσεων πυκνότητας πιθανότητας των κατανομών των συνιστωσών. Η γενική ιδέα είναι ότι τα δεδομένα θα πρέπει να ομαδοποιηθούν σε συστάδες (clusters), όπου τα δεδομένα που ανήκουν στην ίδια συστάδα εμφανίζουν περισσότερες ομοιότητες σε σχέση με εκείνα που ανήκουν σε διαφορετικές συστάδες.

Οι βασικές παράμετροι σε ένα ΜΜΓ είναι τρεις. Η μέση τιμή μ_k και η τυπική απόκλιση σ_k που σχετίζεται με την εκάστοτε συνιστώσα k καθώς και οι αναλογίες μείξης π_k που ορίζονται ως $\Pr(z = k)$.

Στόχος είναι η μάθηση αυτών των παραμέτρων από το μοντέλο. Μια μέθοδος μάθησης των παραμέτρων είναι η εκτίμηση μέγιστης πιθανοφάνειας (maximum likelihood). Για την εκτίμηση της μέγιστης πιθανοφάνειας είναι απαραίτητος ο υπολογισμός των παραγώγων της λογαριθμημένης πιθανοφάνειας. Αυτή είναι μια διαδικασία άρρηκτα συνδεδεμένη με τον αλγόριθμο Μεγιστοποίησης Αναμενόμενης Τιμής (Expectation Maximization, EM).

Η μέγιστη πιθανοφάνεια υπολογίζεται σε σχέση με μια παράμετρο θ . Συγκεκριμένα μπορεί να είναι η μέση τιμή, η τυπική απόκλιση ή αναλογία μείξης. Συγκεκριμένα,

$$\begin{aligned}
 \frac{d}{d\theta} p(x) &= \frac{d}{d\theta} \log \sum_z p(z, x) \\
 &= \frac{\frac{d}{d\theta} \sum_z p(z, x)}{\sum_{z'} p(z', x)} \\
 &= \frac{\sum_z p(z, x) \frac{d}{d\theta} \log p(z, x)}{\sum_{z'} p(z', x)} \\
 &= \sum_z \left(\frac{p(z, x)}{\sum_{z'} p(z', x)} \right) \frac{d}{d\theta} \log p(z, x) \\
 &= \mathbb{E}_{p(z|x)} \left[\frac{d}{d\theta} \log p(z, x) \right]
 \end{aligned} \tag{3.6}$$

Η παράγωγος της περιθωριοποιημένης (marginal) λογαριθμημένης πιθανότητας $p(x)$ είναι η παράγωγος της λογαριθμημένης από κοινού πιθανότητας, όπου η αναμενόμενη τιμή είναι σε σχέση με την εκ των υστέρων κατανομή.

Ο υπολογισμός της αναμενόμενης τιμής της σχέσης (3.6) αθροιστικά για όλα τα δείγματα εκπαίδευσης, φέρει ως αποτέλεσμα την κλίση της λογαριθμημένης πιθανοφάνειας.

Στο παράδειγμα της MMΓ με $x = 2$ συνιστώσες που προαναφέρθηκε, ο υπολογισμός της κλίσης της λογαριθμημένης πιθανοφάνειας σε σχέση με τη μέση τιμή μ_1 της πρώτης συνιστώσας μείξης είναι:

$$\begin{aligned}
 \frac{\partial}{\partial \mu_1} \log \mathbf{p}(x) &= \mathbb{E}_{\mathbf{p}(z|x)} \left[\frac{\partial}{\partial \mu_1} \log \mathbf{p}(z, x) \right] \\
 &= \mathbb{E}_{\mathbf{p}(z|x)} \left[\frac{\partial}{\partial \mu_1} \log \mathbf{p}(z) + \frac{\partial}{\partial \mu_1} \log \mathbf{p}(x | z) \right] \\
 &= \Pr(z = 1 | x) \left[\frac{\partial}{\partial \mu_1} \log \Pr(z = 1) + \frac{\partial}{\partial \mu_1} \log \mathbf{p}(x | z = 1) \right] \\
 &+ \Pr(z = 2 | x) \left[\frac{\partial}{\partial \mu_1} \log \Pr(z = 2) + \frac{\partial}{\partial \mu_1} \log \mathbf{p}(x | z = 2) \right]
 \end{aligned} \tag{3.7}$$

Η συγκεκριμένη διαδικασία καλύπτει την περίπτωση ενός δείγματος εκπαίδευσης. Αντίστοιχα, για τον υπολογισμό της κλίσης της λογαριθμημένης πιθανοφάνειας, εφαρμόζεται για όλα τα δείγματα εκπαίδευσης η σχέση:

$$\frac{\partial \mathbf{l}}{\partial \mu_1} = \sum_{i=1}^N \Pr(z^{(i)} = 1 | x^{(i)}) \cdot \frac{x^{(i)} - \mu_1}{\sigma_1^2} \tag{3.8}$$

Ουσιαστικά υπολογίζεται ένα σταθμισμένο άθροισμα των κλίσεων των υπό συνθήκη πιθανοτήτων για όλα τα δείγματα εκπαίδευσης. Κάθε περίπτωση εκπαίδευσης είναι σταθμισμένη από την πιθανότητα να ανήκει στην πρώτη συνιστώσα μείξης.

Το επόμενο βήμα είναι η μεγιστοποίηση της λογαριθμημένης πιθανοφάνειας με την εφαρμογή της ανόδου κλίσης (gradient ascent). Αυτός είναι ενός τρόπος εκτίμησης των παραμέτρων του MMΓ.

Η πιο αποδοτική μέθοδος εκτίμησης παραμέτρων ενός MMΓ είναι ο αλγόριθμος EM. Αν εισάγουμε την έννοια μιας μεταβλητής $r_k^{(i)} = \Pr(z^{(i)} = 1 | x^{(i)})$ που θα περιγράφει τη βεβαιότητα με την οποία θα

ανήκει ένα δείγμα σε μία συνιστώσα μείξης, η σχέση (3.8) μπορεί να αναδιατυπωθεί ως εξής:

$$\frac{\partial \mathcal{L}}{\partial \mu_1} = \sum_{i=1}^N r_k^{(i)} \cdot \frac{x^{(i)} - \mu_1}{\sigma_1^2} \quad (3.9)$$

Εαν θέσουμε τη μερική παράγωγο ίση με 0 και λύσουμε ως προς μ_1 αυτό θα φέρει ως αποτέλεσμα:

$$\mu_1 = \frac{\sum_{i=1}^N r_k^{(i)} x^{(i)}}{\sum_{i=1}^N r_k^{(i)}} \quad (3.10)$$

Συνεπώς προκύπτει ένα σταθμισμένο άθροισμα δειγμάτων εκπαίδευσης. Αυτή είναι η βάση του αλγόριθμου EM. Ο αλγόριθμος αποτελείται από δύο βασικά δείγματα. Συγκεκριμένα:

- **Βήμα E.** Υπολογισμός των αναμενόμενων τιμών των λανθανόντων μεταβλητών

$$r_k^{(i)} \leftarrow \Pr(z^{(i)} = 1 \mid x^{(i)}) \quad (3.11)$$

- **Βήμα M.** Υπολογισμός παραμέτρων μέγιστης πιθανοφάνειας δοθέντων των αναμενόμενων τιμών

$$\theta \leftarrow \arg \max_{\theta} \sum_{i=1}^N \sum_{k=1}^K r_k^{(i)} [\log \Pr(z^{(i)} = k) + \log p(x^{(i)} \mid z^{(i)} = k)] \quad (3.12)$$

Τα βήματα επαναλαμβάνονται μέχρι να συγκλίνει ο αλγόριθμος και να καταλήξουμε σε ένα τοπικό βέλτιστο.

Το βήμα E για ένα MMF είναι:

$$\begin{aligned}
 r_k^{(i)} &\leftarrow \Pr(z^{(i)} = 1 \mid x^{(i)}) \\
 &= \frac{\Pr(z^{(i)} = k) \mathbf{p}(x^{(i)} \mid z^{(i)} = k)}{\sum_{k'} \Pr(z^{(i)} = k') \mathbf{p}(x^{(i)} \mid z^{(i)} = k')} \quad (3.13) \\
 &= \frac{\pi_k \mathcal{N}(x^{(i)}; \mu_k, \sigma_k)}{\sum_{k'} \pi_{k'} \mathcal{N}(x^{(i)}; \mu_{k'}, \sigma_{k'})}.
 \end{aligned}$$

Αντίστοιχα το βήμα M για ένα MMF χρειάζεται να προσδιοριστούν οι αναλογίες μείξης ως:

$$\pi_k \leftarrow \frac{1}{N} \sum_{i=1}^N r_k^{(i)}. \quad (3.14)$$

Στόχος τώρα είναι η μεγιστοποίηση του

$$\begin{aligned}
 &\sum_{i=1}^N \sum_{k=1}^K r_k^{(i)} \log \Pr(z^{(i)} = k) = \\
 &\sum_{i=1}^N \sum_{k=1}^K r_k^{(i)} \log \pi_k \quad (3.15)
 \end{aligned}$$

υπό τον περιορισμό της κανονικοποίησης $\sum_k \pi_k = 1$. Οδηγούμαστε εκεί, μέσω του υπολογισμού των συντελεστών Lagrange και θέτοντας τη μερική τους παράγωγο ίση με 0 ως εξής:

$$\begin{aligned}
 \mathcal{L} &= \sum_{i=1}^N \sum_{k=1}^K r_k^{(i)} \log \pi_k + \left(1 - \sum_{k=1}^K \pi_k\right) \\
 \frac{\partial \mathcal{L}}{\partial \pi_k} &= \frac{\sum_{i=1}^N r_k^{(i)}}{\pi_k} - \lambda. \quad (3.16)
 \end{aligned}$$

Αν θέσουμε τη μερική παράγωγο ίση με 0, παρατηρούμε ότι για κάθε k :

$$\lambda = \frac{\sum_{i=1}^N r_k^{(i)}}{\pi_k}. \quad (3.17)$$

Για να ισχύει κάτι τέτοιο το π_k πρέπει να είναι ανάλογο του $\sum_{i=1}^N r_k^{(i)}$. Συνεπώς,

$$\pi_k \leftarrow \frac{\sum_{i=1}^N r_k^{(i)}}{\sum_{k'=1}^K \sum_{i=1}^N r_{k'}^{(i)}} \quad (3.18)$$

$$= \frac{\sum_{i=1}^N r_k^{(i)}}{N}. \quad (3.19)$$

Τελικά στόχος είναι η μεγιστοποίηση της σχέσης:

$$\begin{aligned} \sum_{i=1}^N r_k^{(i)} \log p(x^{(i)} | z^{(i)} = k) = \\ \sum_{i=1}^N r_k^{(i)} \log \mathcal{N}(x^{(i)}; \mu_k, \sigma_k). \end{aligned} \quad (3.20)$$

Η ανανέωση έγκειται στην ανανέωση των παραμέτρων, ως εξής:

$$\begin{aligned} \mu_k &\leftarrow \frac{1}{\sum_{i=1}^N r_k^{(i)}} \sum_{i=1}^N r_k^{(i)} x^{(i)} \\ \sigma_k &\leftarrow \frac{1}{\sum_{i=1}^N r_k^{(i)}} \sum_{i=1}^N r_k^{(i)} (x^{(i)} - \mu_k)^2. \end{aligned} \quad (3.21)$$

Η συγκεκριμένη είναι μια πλήρης EM διαδικασία με εφαρμογή σε ΜΜΓ. Τέλος ιδιαίτερη βαρύτητα πρέπει να δίνεται στην επιλογή του πλήθους συνιστωσών K . Αν το K λάβει πολύ μικρή τιμή, υπάρχει ενδεχόμενο υποεκπαίδευσης, ενώ εάν είναι πολύ μεγάλο υπάρχει ενδεχόμενο υπερεκπαίδευσης. Η επιλογή μπορεί να γίνει είτε με εφαρμογή ρυθμίσεων

(tuning) και ελέγχου, είτε με τεχνικές διασταυρωμένης αξιολόγησης (cross validation). Όσον αφορά την αρχικοποίηση των συστάδων, μια καλή τεχνική είναι η τυχαία αρχικοποίηση των μέσων τιμών τους και η απόδοση ευρείας τιμής στην τυπική τους απόκλιση.

3.6.1 Ανίχνευση φόντου με MMΓ

Η μοντελοποίηση του φόντου μιας εικόνας ή ενός καρέ ενός βίντεο από ένα MMΓ είναι μια διαδικασία που στηρίζεται στα εικονοστοιχεία. Εάν κάθε εικονοστοιχείο λαμβάνει μια τιμή η οποία είναι το αποτέλεσμα ενός συγκεκριμένου φωτισμού σε μια συγκεκριμένη επιφάνεια, τότε μια και μόνο κατανομή Γκάους επαρκεί για να περιγράψει το σύνολο των εικονοστοιχείων. Σε πραγματικές συνθήκες όμως εμφανίζονται πολλαπλές επιφάνειες και οι συνθήκες φωτισμού ποικίλουν. Λύση σε τέτοιου είδους ζητήματα παρέχει ένα MMΓ [16][24]. Κάθε φορά που ανανεώνονται οι παράμετροι των Γκαουσιανών, οι Γκαουσιανές αξιολογούνται μέσω ευριστικών μεθόδων. Έτσι, δημιουργούνται υποθέσεις για το ποιες από τις Γκαουσιανές περιγράφει το φόντο. Οι τιμές των εικονοστοιχείων που δεν αντιστοιχούν στις Γκαουσιανές του φόντου, ομαδοποιούνται με τη χρήση συνδεδεμένων συνιστωσών. Τελικά οι συνδεδεμένες συνιστώσες τίθενται υπό παρακολούθηση από καρέ σε καρέ, μέσω της χρήσης υποθέσεων.

Αρχικά θεωρούμε τις τιμές που λαμβάνει ένα εικονοστοιχείο στο χρόνο, ως **διεργασία εικονοστοιχείου**. Η συγκεκριμένη διεργασία είναι μια χρονοσειρά (π.χ. βαθμωτή τιμή για εικόνες τις κλίμακας του γκρι ή διάνυσμα για έγχρωμες εικόνες). Στο χρόνο t η μόνη διαθέσιμη πληροφορία για ένα δοθέν εικονοστοιχείο, είναι οι προηγούμενες τιμές

που έλαβε

$$\{X_1, \dots, X_t\} = \{I(x_0, y_0, i) : 1 \leq i \leq t\}, \quad (3.22)$$

όπου I είναι η αλληλουχία των εικόνων.

Η τιμή κάθε εικονοστοιχείου αντιπροσωπεύει μία μέτρηση της ακτινοβολίας στην κατεύθυνση του αισθητήρα του πρώτου αντικειμένου που τέμνει την οπτική ακτίνα του εικονοστοιχείου. Με στατικό φόντο και στατικό φωτισμό, η συγκεκριμένη τιμή θα είναι σχετικά σταθερή. Αν θεωρηθεί ότι εισάγεται ανεξάρτητος θόρυβος Γκάους κατά τη διαδικασία της δειγματοληψίας, η πυκνότητα της τιμής αυτής θα μπορούσε να περιγραφεί από μια και μόνο κατανομή Γκάους κεντραρισμένη στην μέση τιμή των εικονοστοιχείων. Όπως προαναφέρθηκε όμως στις αλληλουχίες των βίντεο υπάρχει αλλαγή του φωτισμού και της σκηνής καθώς και κινούμενα αντικείμενα.

Εαν αλλάξει ο φωτισμός μιας στατικής σκηνής, είναι απαραίτητη η σύλληψη αυτής της αλλαγής από την Γκαουσιανή. Αν ένα στατικό αντικείμενο εισέλθει στη σκηνή, παραμένει για ένα μεγάλο χρονικό διάστημα και δεν ενσωματωθεί στο φόντο, τα αντίστοιχα εικονοστοιχεία θα θεωρηθούν αυθαίρετα προσκήνιο. Αυτό θα οδηγήσει σε συσσωρευμένα σφάλματα στην εκτίμηση του προσκηνίου, με αποτέλεσμα την κακή συμπεριφορά παρακολούθησης. Αυτοί οι παράγοντες υποδεικνύουν ότι πιο πρόσφατες παρατηρήσεις μπορεί να είναι οι πιο σημαντικές για τον προσδιορισμό των εκτιμήσεων των παραμέτρων μιας Γκαουσιανής. Μια επιπρόσθετη μορφή μεταβολής συμβαίνει εάν υπάρχουν κινούμενα αντικείμενα στη σκηνή. Ακόμα και ένα σχετικά σταθερό έγχρωμο κινούμενο αντικείμενο αναμένεται γενικά να παράξει μεγαλύτερη διακύμανση από ένα "στατικό" αντικείμενο. Επίσης, γενικά, θα πρέπει να

υπάρχουν περισσότερα δεδομένα που υποστηρίζουν τις κατανομές φόντου επειδή επαναλαμβάνονται, ενώ οι τιμές των εικονοστοιχείων για διαφορετικά αντικείμενα συχνά δεν έχουν το ίδιο χρώμα. Αυτοί είναι σημαντικοί παράγοντες για την επιλογή του μοντέλου και την εφαρμογή κανόνων ανανέωσης. Οι πρόσφατες ιστορικές τιμές που έλαβε ένα εικονοστοιχείο μοντελοποιούνται από τη μείξη K Γκαουσιανών κατανομών. Η πιθανότητα να παρατηρηθεί η τρέχουσα τιμή του εικονοστοιχείου περιγράφεται από τη σχέση:

$$P(X_T) = \sum_{i=1}^K \omega_{i,t} \eta(X_t, \mu_{i,t}, \Sigma_{i,t}), \quad (3.23)$$

όπου K είναι το πλήθος των κατανομών Γκάους, $\omega_{i,t}$ είναι η εκτίμηση του βάρους (στάθμιση) της i -οστής κατανομής Γκάους της μείξης στο χρόνο t , $\mu_{i,t}$ είναι η μέση τιμή της i -οστής κατανομής Γκάους της μείξης στο χρόνο t , $\Sigma_{i,t}$ είναι ο πίνακας συμμεταβλητότητας της i -οστής κατανομής Γκάους της μείξης στο χρόνο t και η είναι μια συνάρτηση πυκνότητας πιθανότητας Γκάους

$$\eta(X_t, \mu_{i,t}, \Sigma_{i,t}) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(X_t - \mu_t)^T \Sigma^{-1} (X_t - \mu_t)\right). \quad (3.24)$$

Η τιμή του K ορίζεται σύμφωνα με τους διαθέσιμους πόρους ενός συστήματος. Ο πίνακας συμμεταβλητότητας ορίζεται ως:

$$\Sigma_{k,t} = \sigma_K^2 \mathbf{I} \quad (3.25)$$

Αυτό υποθέτει ότι οι τιμές κόκκινου, πράσινου και μπλε του εκάστοτε εικονοστοιχείου είναι ανεξάρτητες και έχουν τις ίδιες διακυμάνσεις. Αν και αυτό δεν συμβαίνει σίγουρα, η υπόθεση μας επιτρέπει να αποφύγουμε δαπανηρές υπολογιστικά διεργασίες. Έτσι, η κατανομή των πρό-

σφατα παρατηρημένων τιμών κάθε εικονοστοιχείου στη σκηνή χαρακτηρίζεται από ένα ΜΜΓ. Μια νέα τιμή εικονοστοιχείου αντιπροσωπεύεται γενικά από μια από τις κύριες συνιστώσες του μοντέλου του μείγματος και χρησιμοποιείται για την ενημέρωση του μοντέλου.

Αν η διεργασία των εικονοστοιχείων θεωρηθεί στάσιμη διεργασία, εφαρμόζεται ο EM αλγόριθμος για τη μεγιστοποίηση της πιθανοφάνειας των παρατηρήσεων. Στην πραγματικότητα όμως οι διεργασίες των εικονοστοιχείων ποικίλουν στο χρόνο όσο οι σκηνές αλλάζουν. Για το λόγο αυτό χρησιμοποιείται μια προσεγγιστική μέθοδος στην οποία κάθε νέα παρατήρηση αντιμετωπίζεται ως δείγμα συνόλου μεγέθους 1 και εφαρμόζονται κανόνες μάθησης για την ενσωμάτωση νέων δεδομένων.

Επειδή ορίζεται ένα ΜΜΓ για κάθε εικονοστοιχείο της εικόνας, η ακριβής εφαρμογή ενός EM αλγόριθμου σε ένα παράθυρο πρόσφατων δεδομένων θα ήταν υπολογιστικά δαπανηρή. Μια λύση στο συγκεκριμένο πρόβλημα παρέχει μια προσέγγιση K-μέσων. Κάθε νέα τιμή εικονοστοιχείου ελέγχεται αν ανήκει σε μία από τις K κατανομές Γκάους, μέχρι να γίνει αντιστοίχιση. Η αντιστοίχιση γίνεται αν διαπιστωθεί ότι η μέση τιμή μιας συνιστώσας είναι εντός του προκαθορισμένου εύρους τιμών του τρέχοντος εικονοστοιχείου. Συνήθως ορίζεται ένα κατώφλι με ευριστικές μεθόδους. Αυτό είναι ιδιαιτέρως χρήσιμο όταν διαφορετικές περιοχές μιας εικόνας φωτίζονται με διαφορετικό τρόπο, επειδή τα αντικείμενα που εμφανίζονται σε σκιασμένες περιοχές δεν εμφανίζουν γενικά πολύ θόρυβο όπως τα αντικείμενα στις φωτισμένες περιοχές. Ένα ομοιόμορφο κατώφλι συχνά οδηγεί σε εξαφάνιση αντικειμένων όταν μπαίνουν στις σκιασμένες περιοχές.

Εάν καμία από τις K κατανομές δεν αντιστοιχίζεται με την τρέχουσα

τιμή του εικονοστοιχείου, η μέση τιμή της λιγότερο πιθανής κατανομής αντικαθίσταται με την τρέχουσα τιμή και έχει αρχικά υψηλή διακύμανση και χαμηλό βάρος.

Τα εκ των προτέρων βάρη των K κατανομών στο χρόνο t , $\omega_{k,t}$ προσαρμόζονται ως εξής:

$$\omega_{k,t} = (1 - a)\omega_{k,t-1} + a(M_{k,t}), \quad (3.26)$$

όπου a είναι ο ρυθμός μάθησης και ο όρος $M_{k,t}$ είναι 1 για το μοντέλο που αντιστοιχήθηκε και 0 για το υπόλοιπο μοντέλο. Μετά από την εν λόγω προσέγγιση, τα βάρη κανονικοποιούνται εκ νέου. Το $1/a$ είναι μια χρονική σταθερά την ταχύτητα αλλαγής των παραμέτρων των κατανομών. Το $\omega_{k,t}$ είναι ένας μέσος όρος της εκ των υστέρων πιθανότητας να έχουν αντιστοιχηθεί οι τιμές των εικονοστοιχείων με το μοντέλο k δοθέντων των παρατηρήσεων από το χρονικό βήμα 1 μέχρι το τρέχον βήμα t .

Οι παράμετροι μ και σ για τις κατανομές που δεν αντιστοιχίστηκαν παραμένουν ίδιες. Αντίθετα οι παράμετροι της κατανομής στην οποία αντιστοιχήθηκε η νέα παρατήρηση ανανεώνονται σύμφωνα με τη σχέση:

$$\begin{aligned} \mu_t &= (1 - \rho)\mu_{t-1} + \rho X_t \\ \sigma_t^2 &= (1 - \rho)\sigma_{t-1}^2 + \rho(X_t - \mu_t)^T(X_t - \mu_t) \end{aligned} \quad (3.27)$$

όπου

$$\rho = \alpha \eta(X_t | \mu_k, \sigma_k) \quad (3.28)$$

το οποίο επιδρά σαν αιτιατό χαμηλοπερατό φίλτρο.

Ένα από τα σημαντικά πλεονεκτήματα αυτής της μεθόδου είναι ότι όταν κάτι επιτρέπεται να γίνει μέρος του φόντου, δεν καταστρέφει το υπάρχον μοντέλο του φόντου. Το αρχικό χρώμα φόντου παραμένει στη μείξη μέχρι να γίνει το K -οστό πιθανότερο και παρατηρηθεί ένα νέο χρώμα. Επομένως, εάν ένα αντικείμενο είναι ακίνητο για τόσο διάστημα ώστε να γίνει μέρος του φόντου και στη συνέχεια κινείται, η κατανομή που περιγράφει το προηγούμενο φόντο εξακολουθεί να υπάρχει με το ίδιο μ και σ^2 , αλλά μικρότερο ω και θα ενσωματωθεί γρήγορα στο φόντο.

Όσο οι παράμετροι του μοντέλου μείξης του κάθε εικονοστοιχείου αλλάζουν, θα πρέπει να προσδιοριστεί ποια από τις κατανομές Γκάους της μείξης είναι πιο πιθανό να παραχθούν από τις διεργασίες φόντου. Από ευριστικής απόψεως, το ενδιαφέρον επικεντρώνεται σε κατανομές Γκάους που έχουν συνήθως τη μικρότερη διακύμανση και τις περισσότερες ενδείξεις. Για να γίνουν σαφή αυτά τα κριτήρια επιλογής μπορεί κανείς να εξετάσει τη συσσώρευση ενδείξεων και τη σχετικά χαμηλή διακύμανση για τις κατανομές φόντου όταν είναι ορατό ένα στατικό αντικείμενο. Αντίθετα, όταν ένα νέο αντικείμενο αποκρύπτει το αντικείμενο φόντου, δεν θα αντιστοιχηθεί με μία από τις υπάρχουσες κατανομές. Αυτό θα έχει ως αποτέλεσμα είτε τη δημιουργία μιας νέας κατανομής είτε την αύξηση της διακύμανσης μιας υπάρχουσας διανομής. Επίσης, η διακύμανση του κινούμενου αντικειμένου αναμένεται να παραμείνει μεγαλύτερη από ένα εικονοστοιχείο φόντου μέχρι να σταματήσει το κινούμενο αντικείμενο. Για να μοντελοποιηθεί αυτό, είναι αναγκαία μια μέθοδος για να αποφασιστεί ποιο τμήμα του μοντέλου μείξης αντιπροσωπεύει καλύτερα τις διεργασίες του φόντου.

Αρχικά οι κατανομές Γκάους διατάσσονται σύμφωνα με την τιμή

του όρου ω/σ . Αυτή η τιμή αυξάνεται τόσο καθώς η κατανομή αποκτά περισσότερες ενδείξεις και όσο μειώνεται η διακύμανση. Μετά την εκ νέου εκτίμηση των παραμέτρων της μείξης, αρκεί να ταξινομηθεί από την αντιστοιχισμένη κατανομή στην πιο πιθανή κατανομή του φόντου, επειδή θα αλλάξει μόνο η σχετική τιμή των μοντέλων που έχουν αντιστοιχηθεί. Αυτή η διάταξη του μοντέλου είναι στην πραγματικότητα μια διατεταγμένη λίστα ανοιχτού τέλους (open-ended). Οι πιο πιθανές κατανομές φόντου παραμένουν στην κορυφή και οι λιγότερο πιθανές παροδικές κατανομές φόντου βυθίζονται προς τα κάτω και τελικά αντικαθίστανται από νέες κατανομές.

Στη συνέχεια επιλέγονται οι πρώτες B κατανομές ως μοντέλο φόντου

$$B = \arg \min_b \left(\sum_{k=1}^b \omega_k > T \right) \quad (3.29)$$

όπου T είναι ένα μέτρο του ελάχιστου τμήματος των δεδομένων που πρέπει να αντιπροσωπεύεται από το φόντο. Αυτό παίρνει τις βέλτιστες κατανομές μέχρι να υπολογιστεί ένα ορισμένο τμήμα, T , των πρόσφατων δεδομένων που έχουν αντιπροσωπευτεί από το φόντο. Αν επιλεγθεί μια μικρή τιμή για το T , το μοντέλο φόντου είναι συνήθως μονοτροπικό. Στη συγκεκριμένη περίπτωση η χρήση μόνο της πιο πιθανής κατανομής θα εξοικονομήσει την επεξεργασία. Αντίθετα, αν το T είναι μεγαλύτερο, προκύπτει μια πολυτροπική κατανομή που προκαλείται από μια επαναλαμβανόμενη κίνηση του φόντου. Θα μπορούσε να οδηγήσει στο να συμπεριληφθούν περισσότερα από ένα χρώματα στο μοντέλο φόντου. Αυτό έχει ως αποτέλεσμα ένα φαινόμενο διαφάνειας που να επιτρέπει στο φόντο να δέχεται δύο ή περισσότερα ξεχωριστά χρώματα.

3.7 Σύνοψη

Το έργο της αξιόπιστης ανίχνευσης και παρακολούθησης κινούμενων αντικειμένων σε βίντεο παρακολούθησης, δεν έχει ακόμη επιλυθεί πλήρως και έχει πολλά ζητήματα που απαιτούν λύσεις. Πολλές συνθήκες οδηγούν σε κακής ποιότητας ανίχνευση φόντου. Για να οδηγηθούμε σε αξιόπιστη ανίχνευση κινούμενου αντικειμένου απαιτείται η ακριβής ανίχνευση του φόντου. Υπάρχουν αρκετές τεχνικές για την εύρεση του φόντου. Οι πιο διαδεδομένες είναι η τεχνική της αφαίρεσης φόντου, η αφαίρεση φόντου μέσω στατιστικών μεθόδων, η χρήση του μοντέλου Γκαουσιανών και η ανίχνευση του προσκηνίου. Στη παρούσα διπλωματική χρησιμοποιείται το μοντέλο μείξης Γκαουσιανών με την ανίχνευση του προσκηνίου. Ο συνδυασμός αυτών των μεθόδων προσφέρει ικανοποιητικά αποτελέσματα αφού διαχωρίζεται το φόντο από το προσκηνίο, επιτυγχάνεται λοιπόν ο πρώτος στόχος που είναι η ανίχνευση του κινούμενου αντικειμένου. Στη συνέχεια θα δοθεί η αναλυτική περιγραφή της μεθοδολογίας για την επίτευξη του δεύτερου στόχου που είναι ο χαρακτηρισμός ενός αντικειμένου ως εγκαταλελειμμένο.

Κεφάλαιο 4

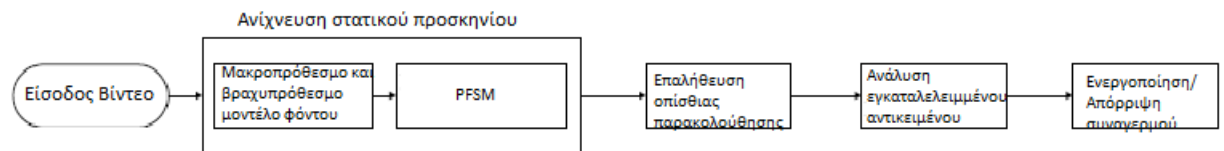
Μεθοδολογία

4.1 Εισαγωγή

Σε αυτή την εργασία, υλοποιήθηκε μέθοδος ενσωμάτωσης διπλού χρονικού ρυθμού ανανέωσης προσκηνίου για την εκτίμηση του στατικού προσκηνίου με χρήση απλών-καμερών εικόνων βίντεο. Στη προσέγγισή μας περιλαμβάνουμε την κατασκευή ενός βραχυπρόθεσμου και ενός μακροπρόθεσμου μοντέλου φόντου, που εκπαιδεύονται συνεχώς καθόλη την διάρκεια της λήψης. Στη συνέχεια εισάγουμε ένα απλό μοντέλο μηχανής πεπερασμένων καταστάσεων βασισμένη στα εικονοστοιχεία (Pixel-Based Finite State Machine, PFSM) [29]. Το συγκεκριμένο μοντέλο χρησιμοποιεί μια χρονική πληροφορία μετάβασης για την αναγνώριση του στατικού προσκηνίου που βασίζεται στο μοτίβο μιας ακολουθίας για κάθε εικονοστοιχείο του αντικειμένου. Επειδή η προτεινόμενη προσέγγιση περιλαμβάνει την πληροφορία χρονικής μετάβασης, μπορούμε να μειώσουμε την επίδραση της ατελούς ανίχνευσης προσκηνίου και στα δυο μοντέλα φόντου, βελτιώνοντας έτσι την ακρίβεια της συμπε-

ρασματολογίας στατικών προσκηνίων. Ένα σύστημα παρακολούθησης του κατόχου χρησιμοποιείται επίσης στην μέθοδο, για να επαληθεύσει σημασιολογικά το εγκαταλελειμμένο αντικείμενο.

Η μέθοδος ανίχνευσης εγκαταλελειμμένου αντικειμένου βασίζεται στη μοντελοποίηση του φόντου και στην αφαίρεσή του. Η ακόλουθη υποενότητα παρέχει μια εννοιολογική ανασκόπηση της αφαίρεσης του φόντου και των σχετικών ρυθμών μάθησης για την ενημέρωση του μοντέλου του φόντου. Στο σχήμα 4.1 παρουσιάζεται ο αλγόριθμος ανίχνευσης στατικών περιοχών προσκηνίου. Στο δεύτερο βήμα του διαγράμματος ροής του σχήματος συνοψίζονται οι δύο βασικές διεργασίες της ανίχνευσης στατικού προσκηνίου. Στη συνέχεια ακολουθεί η διαδικασία παρακολούθησης του ιδιοκτήτη. Μετά, γίνεται σημασιολογική ανάλυση του εγκαταλελειμμένου αντικειμένου. Τέλος δοθέντος του αποτελέσματος της συνολικής διαδικασίας, προάγεται ανάλογη έξοδος.



Σχήμα 4.1: Αλγόριθμος ανίχνευσης προσκηνίου.

4.2 Ανίχνευση προσκηνίου

Αντίστοιχης σημασίας με την ανίχνευση και αφαίρεση φόντου που προαναφέρθηκε, είναι η ανίχνευση του προσκηνίου. Στόχος είναι η ανίχνευση των περιοχών (εμφανές σχήμα) των ανεξάρτητα κινούμενων αντικειμένων σε ένα βίντεο εναξαρτήτως της ταχύτητας, της κατεύθυνσης ή

της υφής τους. Συνήθως για την εξαγωγή των αντικειμένων του προσκηνίου από μια αλληλουχία εικόνων γίνεται μέσω της αφαίρεσης του φόντου. Ωστόσο, αυτό το έργο γίνεται δύσκολο όταν το φόντο περιέχει σκιές και κινούμενα αντικείμενα, π.χ. κλαδιά δέντρων και κινούμενες κυλιόμενες σκάλες, και όταν υπόκειται σε διάφορες αλλαγές, όπως οι αλλαγές φωτισμού. Έχουν προταθεί πολλές μέθοδοι ανίχνευσης προσκηνίου σε πραγματικό χρόνο. Ωστόσο, τα περισσότερα από αυτά αναπτύχθηκαν υπό την προϋπόθεση ότι το φόντο αποτελείται από ακίνητα αντικείμενα των οποίων το χρώμα ή η ένταση ενδέχεται να αλλάζουν σταδιακά με την πάροδο του χρόνου.

Η μέθοδος που περιγράφηκε στο 3.6.1 παρέχει τη δυνατότητα ανίχνευσης εικονοστοιχείων προσκηνίου σε κάθε καρέ. Αυτό επιτυγχάνεται μέσω της ανανέωσης των διεργασιών των εικονοστοιχείων. Τα προσημασμένα εικονοστοιχεία του φόντου ταξινομούνται σε περιοχές από ένα αλγόριθμο συνδεδεμένων συνιστωσών δύο περασμάτων [25]. Επειδή αυτή η διαδικασία είναι αποτελεσματική για τον προσδιορισμό ολόκληρου του κινούμενου αντικειμένου, οι κινούμενες περιοχές μπορούν να χαρακτηριστούν όχι μόνο από τη θέση τους, αλλά από το μέγεθος, τις χρονικές στιγμές και άλλες πληροφορίες σχήματος. Τα χαρακτηριστικά αυτά δεν είναι μόνο χρήσιμα για μεταγενέστερη επεξεργασία και ταξινόμηση, αλλά μπορούν να βοηθήσουν στη διαδικασία παρακολούθησης.

Η ανίχνευση προσκηνίου επικεντρώνεται κυρίως σε αποθηκευμένα βίντεο και εικόνες. Υπάρχει όμως ένας μικρός αριθμός μεθόδων που είναι διαθέσιμες για ανίχνευση προσκηνίου σε πραγματικό χρόνο. Μερικές εξ αυτών είναι [26]:

- Αυτοοργανωμένη Αφαίρεση φόντου [27]. Είναι αλγόριθμος που

χειρίζεται με ακρίβεια σκηνές που περιέχουν κινούμενα φόντα, σταδιακές παραλλαγές φωτισμού και σκιών που μεταδίδονται από κινούμενα αντικείμενα. Είναι ανθεκτική σε ψευδείς ανιχνεύσεις για διαφορετικούς τύπους εικόνων που λαμβάνονται με σταθερές κάμερες. Ακόμη και χωρίς προηγούμενη γνώση, η μέθοδος μπορεί να ανιχνεύσει το κινούμενο αντικείμενο με βάση το μοντέλο φόντου.

- Μοντέλο ακολουθίας εικόνων βασισμένο σε νευρικό δίκτυο. Η μοντελοποίηση γίνεται μέσω της αυτο-οργάνωσης. Οι παραλλαγές στην ακολουθία εικόνων θεωρούνται ως τροχιές των εικονοστοιχείων κατά τη διάρκεια του χρόνου. Η μαθησιακή διαδικασία των ενεργών νευρώνων φαίνεται να περιορίζεται χωρικά και βασίζεται στην τοπική γειτονιά τους. Το μοντέλο του φόντου προσαρμόζεται καλά στις αλλαγές της σκηνής και καταφέρνει να καταγράψει το μεγαλύτερο μέρος της αλλαγών χαρακτηριστικών στην ακολουθία εικόνων.
- Προσαρμοστική κατάτμηση βάσει εικονοστοιχείων (Pixel Based Adaptive Segmenter) [28]. είναι ένα μοντέλο που διατηρεί τις τιμές των εικονοστοιχείων που παρατηρήθηκαν πρόσφατα και εντοπίζει το φόντο. Το μοντέλο περιέχει ένα σύνολο υποδιαιρέσεων. Το μπλοκ αποφάσεων, το οποίο είναι το βασικό συστατικό στοιχείο, αποφασίζει για το εάν ένα τμήμα ενός καρέ είναι προσκήνιο ή όχι, βάσει ενός κατωφλιού που τίθεται ανά εικονοστοιχείο της τρέχουσας εικόνας. Το μοντέλο ενημερώνεται με την πάροδο του χρόνου με μια καθορισμένη διαδικασία για την πραγματοποίηση

των αλλαγών στο παρασκήνιο.

4.3 Συνδυασμός μακροπρόθεσμου και βραχυπρόθεσμου μοντέλου φόντου

Το πρώτο και πιο ζωτικής σημασίας βήμα της μεθόδου περιλαμβάνει τη μοντελοποίηση του φόντου. Αρχικά περιγράφονται τα μοντέλα μακροπρόθεσμου και βραχυπρόθεσμου φόντου. Ο αλγόριθμος ξεκινά από μια γενική μέθοδο μοντελοποίησης φόντου που λειτουργεί σε 2 μαθησιακούς ρυθμούς μάθησης. Δοθείσας μιας ακολουθίας καρέ $I_t (t \in N)$ μεγέθους $m \times n$ και αν $B(x, y)$ το εν δυνάμει μοντέλο φόντου, η γενική φιλοσοφία μοντελοποίησης συνοψίζεται ως εξής:

Για την ανανέωση του φόντου του βήματος 10, χρησιμοποιείται μια παράμετρος ρυθμού μάθησης, $0 \leq \lambda \leq 1$. Ο ρυθμός μάθησης λειτουργεί σαν μέσο εξισορρόπησης μεταξύ του λB και του $(1 - \lambda)I_t$ και παρέχει τη δυνατότητα ρύθμισης (tuning). Η συγκεκριμένη παράμετρος οδηγεί το μοντέλο φόντου να προσαρμόζεται στα νέα εισαχθέντα δεδομένα (π.χ. νέο καρέ). Όταν το λ είναι μικρό, το μοντέλο προσαρμόζεται γρηγορότερα, ενώ όταν λαμβάνει μεγαλύτερες τιμές το μοντέλο φόντου προσαρμόζεται στα νέα δεδομένα εκπαίδευσης με πιο αργό ρυθμό.

Για την μοντελοποίηση του φόντου επιλέχθηκε η μέθοδος MMΓ. Όπως προαναφέρθηκε, ένας μικρός ρυθμός μάθησης δηλαδή ένα μικρό λ ενημερώνει το μοντέλο του φόντου πιο γρήγορα. Σε αυτή την περίπτωση, το μοντέλο φόντου $B(x, y)$ θεωρείται ένα MMΓ στο χρωματικό χώρο RGB. Η παράμετρος ρυθμού μάθησης λ εφαρμόζεται για την κα-

```
Data:  $I_t$   
Result: Background modeling  
1 initialization;  
2 foreach pixel do  
3 | initialize  $B(x, y)$ ;  
4 end  
5 if pixel  $I_t(x, y) \in B(x, y)$  then  
6 |  $(x, y)$  is background pixel;  
7 else  
8 |  $(x, y)$  is foreground pixel;  
9 end  
10 foreach background pixel  $(x, y)$  do  
11 | update  $B(x, y)$ ;  
12 end  
13  $t \leftarrow t + 1$ ;  
14 go to 5;
```


τάλληλη ανανέωση των κατατανομών του ΜΜΓ όταν παρατηρηθεί ένα νέο χρώμα $I_t(x, y)$ και γίνεται ταυτοποίηση του εκάστοτε εικονοστοιχείου για το εάν είναι μέρος του φόντου ή όχι.

Όπως είπαμε, ένας μικρός ρυθμός μάθησης ταυτίζεται με μια ταχύτερη ανανέωση του φόντου. Ένα τέτοιο μοντέλο που μαθαίνει σε γρήγορο χρόνο λόγω του λ_S , ονομάζεται βραχυπρόθεσμο (short-term model) B_S . Η δυαδική εικόνα προσκηνίου που προκύπτει από την εφαρμογή του βραχυπρόθεσμου μοντέλου, συμβολίζεται ως F_S . Αντιθέτως ένα μεγάλο λ_L ταυτίζεται με μια πιο αργή ανανέωση του μοντέλου φόντου. Αυτό το μοντέλο αποτελεί το μακροπρόθεσμο μοντέλο (long term model) B_L . Αντίστοιχα η δυαδική εικόνα προσκηνίου που προκύπτει από την εφαρμογή ενός μακροπρόθεσμου μοντέλου φόντου, συμβολίζεται ως F_L .

Ο συνδυασμός των δυο μοντέλων, μακροπρόθεσμων και βραχυπρόθεσμων, είναι κατάλληλος για την ανίχνευση των στατικών αντικειμένων. Μια αποσκευή η οποία έχει αφεθεί από τον ιδιοκτήτη αποτελεί για το μακροπρόθεσμο μοντέλο αντικείμενο προσκηνίου. Επιπλέον εξαιτίας του γρήγορου ρυθμού μάθησης η αφεθείσα αποσκευή στο βραχυπρόθεσμο μοντέλο αποτελεί αντικείμενο φόντου.

Σύμφωνα με τη μέθοδο, ένα εικονοστοιχείο (x, y) αναπαρίσταται ως ένα δίμπιτος κωδικός S_i που είναι το αποτέλεσμα του συνδυασμού του μακροπρόθεσμου και του βραχυπρόθεσμου μοντέλου φόντου. Συγκεκριμένα,

$$S_i = F_L(i)F_S(i), \quad (4.1)$$

όπου τα $F_L(i)$ και $F_S(i)$ λαμβάνουν τιμές 0 ή 1 και αναπαριστούν τις δυαδικές τιμές του εικονοστοιχείου i που ανήκουν στο προσκήνιο. Δοθέντος ενός τρέχοντος εικονοστοιχείου i , το S_i λαμβάνει τέσσερις πιθανές λογικές καταστάσεις ως εξής:

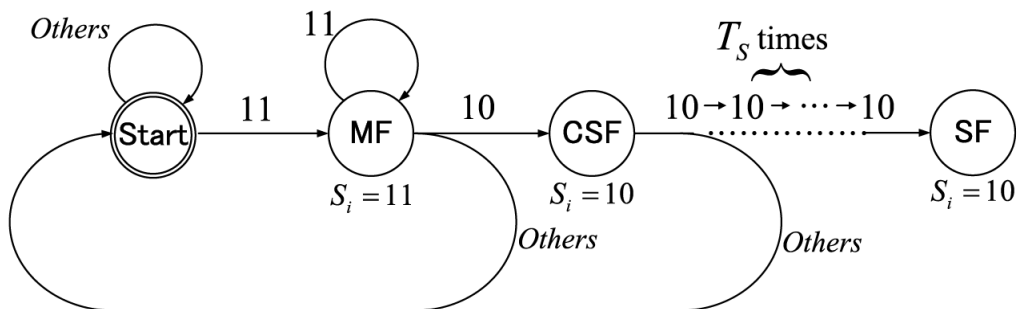
1. $S_i = 00$ υποδηλώνει ότι ένα εικονοστοιχείο i είναι εικονοστοιχεία φόντου επειδή ταυτοποιήθηκε ως φόντο και από το μακροπρόθεσμο μοντέλο B_L , αλλά και από το βραχυπρόθεσμο μοντέλο B_S .
2. $S_i = 01$ υποδηλώνει ότι το εικονοστοιχείο i είναι ένα εικονοστοιχείο φόντου το οποίο προσωρινά κρύφτηκε από ένα αντικείμενο και στη συνέχεια εκτέθηκε σε μια πιο πρόσφατη εικόνα.
3. $S_i = 10$ υποδεικνύει ότι εικονοστοιχείο i είναι ένα πιθανά υποψήφιο στατικό εικονοστοιχείο προσκηνίου.
4. $S_i = 11$ υποδεικνύει ότι το εικονοστοιχείο i αντιστοιχεί σε ένα κινούμενο αντικείμενο της σκηνής.

Όταν ένα εγκαταλελειμμένο αντικείμενο εντοπισθεί, θα πρέπει πρωτίστως να εξετασθεί εάν τα εικονοστοιχεία που το απαρτίζουν στη σκηνή, λαμβάνουν την κατάσταση $S_i = 10$. Διότι,

- Μπορεί να είναι εικονοστοιχεία προσκηνίου, τα οποία υπήρχαν στη σκηνή για μεγάλο χρονικό διάστημα (σε πολλά καρέ), πράγμα που εξάγεται ως απόφαση από το μακροπρόθεσμο μοντέλο.
- Μπορεί να είναι εικονοστοιχεία-μέρος ενός αντικειμένου το οποίο δεν κινήθηκε καθόλου για ένα αρκετό χρονικό διάστημα και για το λόγο αυτό το βραχυπρόθεσμο μοντέλο τα απέρριψε.

Αυτές είναι δύο βασικές ιδιότητες που προσδιορίζουν τα στατικά εικονοστοιχεία προσκηνίου και ορίζουν τον δυαδικό κώδικα κατάστασης ο οποίος με τη σειρά του προσδιορίζει εάν ένα αντικείμενο είναι εγκαταλελειμμένο ή όχι. Έτσι αυτός ο κώδικας S_i ορίζεται για μια μόνο εικόνα. Επειδή ο θόρυβος και η δυναμική φύση των δεδομένων μπορεί να προκαλέσουν προβληματική μοντελοποίηση φόντων, οι συγκεκριμένοι κώδικες πρέπει να είναι προσωρινοί, καθώς δεν γενικεύονται για κάθε επόμενο καρέ ενός βίντεο. Είναι προφανές ότι η ταυτοποίηση των εικονοστοιχείων σε μεμονωμένες εικόνες, είναι αδύνατον να οδηγήσουν σε αξιόπιστα πορίσματα για το εάν ένα αντικείμενο θεωρείται εγκαταλελειμμένο ή όχι.

Στην παρούσα διπλωματική εργασία χρησιμοποιήθηκε χρονική πληροφορία συνέχειας, η οποία αποφέρει ικανοποιητικά αποτελέσματα. Σε μια αλληλουχία εικόνων (ή καρέ), ο κώδικας S_i θα υπόκειται σε αλλαγές, ακολουθώντας ένα χρονικό κανόνα. Ο συγκεκριμένος κανόνας παρέχεται από ένα μοντέλο μηχανής πεπερασμένων καταστάσεων (FSM) (σχήμα 4.2)



Σχήμα 4.2: Μοντέλο Μηχανής Πεπερασμένων Καταστάσεων

4.4 Μοντέλο μηχανής πεπερασμένων καταστάσεων βασισμένη στα εικονοστοιχεία

Σύμφωνα με τη μέθοδο PFSM χρησιμοποιείται χρονική πληροφορία μετάβασης για την ταυτοποίηση ενός αντικείμενου ως στατικό. Αυτό επιτυγχάνεται με τη ανάλυση του πρότυπου αλληλουχίας του κάθε εικονοστοιχείου. Κάθε εικονοστοιχείο θεωρείται ότι σχετίζεται μόνο με μια κατάσταση σε ένα τρέχον βήμα χρόνου. Ανάλογα με την έξοδο του μακροπρόθεσμου και του βραχυπρόθεσμου μοντέλου, η κατάσταση του εικονοστοιχείου i δύναται να αλλάξει από το τρέχον χρονικό βήμα στο επόμενο. Η ανάλυση της συμπεριφοράς των εικονοστοιχείων γίνεται με την εφαρμογή ενός μοντέλου FSM. Εάν για ένα εικονοστοιχείο ταυτοποιηθεί ένα συγκεκριμένο πρότυπο αλληλουχίας μεταβάσεων, τότε ταυτοποιείται ως στατικό προσκήνιο.

Το πρώτο βήμα είναι η εκκίνηση από μία αρχική κατάσταση, όπου το σύστημα ενεργοποιείται από το $S_i = 11$, το οποίο υποδεικνύει ότι το εικονοστοιχείο i καλύφθηκε από ένα μέρος του προσκηνίου όπως προαναφέρθηκε. Όταν ένας άνθρωπος εγκαταλείψει ένα αντικείμενο, το βραχυπρόθεσμο μοντέλο μέσω των ανανεώσεων, σύντομα κατατάσσει το αντικείμενο αυτό στο μοντέλο φόντου του. Παράλληλα το μακροπρόθεσμο μοντέλο δεν το κατατάσσει στο μοντέλο φόντου του. Έτσι, η κατάσταση της συγκεκριμένης περιοχής του καρέ αλλάζει από $S_i = 11$ σε $S_i = 10$, υποδεικνύοντας ότι είναι υποψήφια περιοχή του στατικού προσκηνίου. Εάν η κατάσταση $S_i = 10$ παραμείνει για κάποιο προκαθορισμένο χρόνο T_S , τότε θεωρείται ότι η περιοχή είναι πράγματι μέρος του στατικού προσκηνίου. Αυτή η διαδικασία είναι το απαιτούμενο πρό-

τυπο αλληλουχίας που αποτελεί συνθήκη για το αν τα εικονοστοιχεία της συγκεκριμένης περιοχής είναι εικονοστοιχεία προσκηνίου. Διαφορετικά η κατάσταση του εκάστοτε εικονοστοιχείου i μεταβαίνει στην αρχική κατάσταση και η διαδικασία επαναλαμβάνεται μέχρι η αρχική κατάσταση να γίνει $S_i = 11$.

Data: pixel i with code sequence S_i
Result: Static foreground detection

```

1 initialization;
2  $S_i = 11$ ;
3 if  $S_i = 11 \rightarrow$  long series of  $S_i = 10$  then
4 |    $i$  is static foreground;
5 else
6 |   while  $S_i \neq 11$  for pixel  $i$  do
7 | |   restart;
8 |   end
9 end

```

Σε κάθε καρέ που τίθεται προς ανάλυση, συλλέγονται τα εικονοστοιχεία που ταυτοποιήθηκαν ως στατικό φόντο από τον αλγόριθμο. Στη συνέχεια εφαρμόζεται η μέθοδος συνδεδεμένων συνιστωσών [25] ώστε τα συγκεκριμένα εικονοστοιχεία να ομαδοποιηθούν και να αποκλειστούν οι μικρότερες συνιστώσες. Σε περίπτωση που σε ένα τρέχον καρέ, όλες οι συνιστώσες είναι πολύ μικρές, δεν διενεργείται περαιτέρω επαλήθευση. Στην αντίθετη περίπτωση, οι συνιστώσες που έχουν συγκερατηθεί από τον αλγόριθμο και αντιμετωπίζονται ως μέρος του στατικού προσκηνίου, θεωρούνται υποψήφια εγκαταλελειμμένα αντικείμενα του τρέχοντος καρέ. Αυτή η πληροφορία ανατροφοδοτείται στο

επόμενο στάδιο στο οποίο εφαρμόζεται ο αλγόριθμος οπίσθιας παρακολούθησης (back-tracing)[29] με στόχο την επαλήθευση του τρέχοντος αποτελέσματος.

4.5 Επαλήθευση οπίσθιας παρακολούθησης

Στη συνέχεια ακολουθεί η τελική επαλήθευση για το εάν το αντικείμενο είναι εγκαταλελειμμένο ή όχι. Σημασιολογικά τίθεται το ερώτημα αν το εκάστοτε αντικείμενο έχει πράγματι εγκαταλειφθεί από τον ιδιοκτήτη ή εάν απλώς τοποθετήθηκε σε κάποιο σημείο της σκηνής για ένα σύντομο χρονικό διάστημα. Αυτό το πόρισμα διεξάγεται με την εφαρμογή του αλγορίθμου οπίσθιας παρακολούθησης.

Το πρώτο βήμα είναι ο προσδιορισμός της σχετικής απόστασης του ιδιοκτήτη από το αντικείμενο (αν ο ιδιοκτήτης είναι κοντά του ή έχει απομακρυνθεί). Εφόσον ο κάτοχος του αντικειμένου δεν επιστρέψει στην περιοχή περιμετρικά του αντικειμένου, το τελευταίο αντιμετωπίζεται ως εγκαταλελειμμένο.

Μέσω της μεθόδου PFSM έχει ήδη οριστεί το στατικό προσκήνιο, το οποίο θεωρείται υποψήφιο εγκαταλελειμμένο αντικείμενο. Όταν σε χρόνο t το στατικό προσκήνιο θεωρηθεί υποψήφιο εγκαταλελειμμένο και ταυτόχρονα δεν υπάρχει κάποιο κινούμενο αντικείμενο στην εγγύτερη περιοχή του ακτίνας D , τότε ο αλγόριθμος επιστρέφει από το τρέχον καρέ t , στο προηγούμενο

$$t_0 = t - T_S. \quad (4.2)$$

Το t_0 αντιστοιχεί στη χρονική στιγμή κατά την οποία ο κάτοχος είχε πιθανώς εναποθέσει το αντικείμενο στο συγκεκριμένο σημείο. Αντίστοιχα

το T_S είναι μια σταθερά που υποδηλώνει το πέρασ του χρόνου. Για παράδειγμα, προσδιορίζει τον αριθμό των επαναλήψεων της εμφάνισης $S_i = 10$ για ένα εικονοστοιχείο i σε μια αλληλουχία. Η θέση στην οποία τοποθετήθηκε το υποψήφιο αντικείμενο την χρονική στιγμή t_0 , ορίζεται ως p . Το p θεωρείται κέντρο και με αυτό το σημείο αναφοράς δημιουργείται ένα χωροχρονικό παράθυρο W_0 μεγέθους (r_2, δ) , όπου r είναι η ακτίνα του κύκλου του κέντρου και δ το χρονικό διάστημα $(t_0, t_0 + \delta)$. Στη συνέχεια ακολουθεί η διαδικασία ανίχνευσης ανθρώπου. Για το παράθυρο W_0 θεωρούμε ότι όλες οι οντότητες του προσκηνίου, ταυτοποιήθηκαν μέσω της εφαρμογής της αφαίρεσης φόντου με ΜΜΓ. Από αυτές τις οντότητες επιλέγονται εκείνες που προσεγγίζουν το ανθρώπινο σχήμα χρησιμοποιώντας το ύψος και το πλάτος τους. Η διαδικασία ανίχνευσης ανθρώπου στηρίζεται στη μέθοδο ανίχνευσης deformable part-based model (DPM)[30]. Σύμφωνα με τη συγκεκριμένη μέθοδο εφαρμόζεται η τεχνική του ολισθαίνοντος παραθύρου με φίλτρα πολλαπλών πυρήνων (kernels). Το υποψήφιο αντικείμενο αναπαρίσταται με τη χρήση ενός βασικού φίλτρου και διάφορων μερικών φίλτρων. Το βασικό φίλτρο περιγράφει την συνολική εμφάνιση του αντικειμένου, ενώ τα μερικά φίλτρα είναι υπεύθυνα για τα μερικά τμήματά του. Ένα αντικείμενο εντοπίζεται όταν η περιοχή που καταλαμβάνει έχει κριθεί ως υποψήφια τόσο από το βασικό φίλτρο όσο και από τα μερικά. Για να αποφευχθεί το υπολογιστικό κόστος της χρήσης πολλών φίλτρων εφαρμόζεται μια τεχνική ολισθαίνοντος παραθύρου που στηρίζεται στη συνέλιξη [31]. Σύμφωνα με τη συγκεκριμένη μέθοδο, εφαρμόζεται γρήγορος μετασχηματισμός Fourier, δοθείσας της παραδοχής ότι η συνέλιξη στο πεδίο του χρόνου ταυτίζεται με τον πολλαπλασιασμό στο πεδίο των συχνοτήτων.

Η περιοχή του προσκηνίου που περιέχει άνθρωπο αντιμετωπίζεται ως η οντότητα του ιδιοκτήτη που είναι μια άμορφη μάζα. Η συγκεκριμένη περιοχή θα τεθεί προς παρακολούθηση. Η θέση της στο καρέ συμβολίζεται με p_1 . Αν έχουν εντοπιστεί περισσότερες από μια τέτοιες μάζες επιλέγουμε εκείνη που βρίσκεται πλησιέστερα του σημείου p .

Αρχικά εξάγεται η κατανομή του χρώματος ως χαρακτηριστικό των περιοχών προσκηνίου. Έπειτα έχοντας ως κέντρο το p_1 δημιουργείται ένα νέο χρονοχωρικό παράθυρο W_1 , το οποίο έχει μέγεθος (r^2, δ) . Για την εύρεση της περιοχής που παρουσιάζει περισσότερη ομοιότητα με εκείνη του κατόχου του W_1 , εφαρμόζεται ο συντελεστής Bhattacharyya. Η απόσταση Bhattacharyya χρησιμοποιείται για τη μέτρηση της ομοιότητας μεταξύ δύο κατανομών πιθανοτήτων. Ο συντελεστής Bhattacharyya είναι ένα μέτρο της ποσότητας αλληλεπικάλυψης μεταξύ δύο κατανομών. Χρησιμοποιείται συνήθως για τον προσδιορισμό της σχετικής εγγύτητας των δύο υπό εξέταση δειγμάτων. Αν p και q δύο κατανομές πιθανοτήτων τότε η απόσταση Bhattacharyya ορίζεται ως:

$$D_B(p, q) = -\ln(BC(p, q)) \quad (4.3)$$

όπου

$$BC(p, q) = \sum_{x \in X} \sqrt{p(x)q(x)} \quad (4.4)$$

ο συντελεστής Bhattacharyya για διακριτές κατανομές πιθανοτήτων.

Αντίστοιχα για συνεχείς κατανομές πιθανοτήτων ο συντελεστής Bhattacharyya είναι:

$$BC(p, q) = \int \sqrt{p(x)q(x)} dx \quad (4.5)$$

Στη συνέχεια δημιουργείται ένα νέο παράθυρο W_2 με κέντρο τη νέα περιοχή που ταυτοποιήθηκε. Με τη συγκεκριμένη διαδικασία παρέχε-

ται η δυνατότητα παρακολούθησης της περιοχής που αντιπροσωπεύει τον ιδιοκτήτη του αντικειμένου, μέχρι να παρέλθει η προκαθορισμένη χρονική στιγμή t ή μέχρις ότου η παρακολουθούμενη περιοχή να μετακινηθεί πέραν της ακτίνας D .

4.6 Σημαιολογική ανάλυση εγκαταλελειμμένου αντικειμένου

Μόλις ληφθεί η τροχιά του ιδιοκτήτη, εμφανίζεται μία προειδοποίηση ότι τα αντικείμενα έχουν εγκαταλειφθεί, σύμφωνα με τους παρακάτω δύο κανόνες:

- Χρονικός κανόνας: Το αντικείμενο δηλώνεται ως αφύλακτο όταν έχει εγκαταλειφθεί από τον ιδιοκτήτη του, και παραμένει χωρίς επίβλεψη για χρόνο = 30 δευτερόλεπτα.
- Χωρικός κανόνας: Το αφύλακτο αντικείμενο ορίζεται ως εγκαταλελειμμένο όταν αφήνεται από τον ιδιοκτήτη του. Όταν η απόσταση μεταξύ του ιδιοκτήτη και του αντικειμένου είναι μεγαλύτερη από μια προκαθορισμένη απόσταση $D = 3$ μέτρα, τότε ενεργοποιείται ένας συναγερμός.

Σύμφωνα με το PFSM ο χρονικός κανόνας ικανοποιείται για $T_S = 30f$ καρέ, όπου f το πλήθος των καρέ ανά δευτερόλεπτο που είναι η ταχύτητα καταγραφής του βίντεο. Ο χωρικός κανόνας πιστοποιείται εξετά-

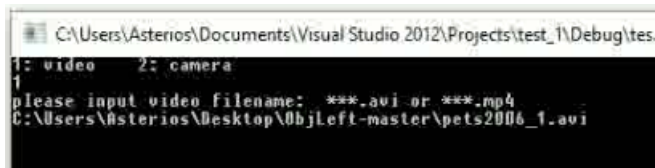
ζώντας την τροχιά του ιδιοκτήτη. Φτιάχνουμε μια περιοχή με κέντρο το υποψήφιο εγκαταλελειμμένο αντικείμενο με μια ακτίνα ίση με $D = 3\mu$ εικονοστοιχεία, όπου μ ο παράγοντας κλιμάκωσης ο οποίος μετατρέπει τα εικονοστοιχεία σε αποστάσεις του πραγματικού κόσμου. Στη συνέχεια ελέγχεται αν ο κάτοχος απομακρύνεται από την περιοχή ενδιαφέροντος. Ένας συναγερμός χτυπά όταν ικανοποιούνται ταυτόχρονα και ο χωρικός και ο χρονικός κανόνας.

4.7 Εκτέλεση Μεθοδολογίας

Πριν την εκτέλεση του προγράμματος ο χρήστης καλείται να ορίσει μια σειρά παραμέτρων που είναι: το πλήθος των καρέ που επιθυμεί ο χρήστης να συλλέξει για την εκπαίδευση του μοντέλου φόντου, το μέγιστο MIN SFG και ελάχιστο πλήθος εικονοστοιχείων MAX SFG εντοπισμού του στατικού φόντου, το μέγιστο MAX FG και ελάχιστο πλήθος MIN FG εικονοστοιχείων που αποτελούν προσκήνιο, την τιμή του ρυθμού μάθησης μακροπρόθεσμου και βραχυπρόθεσμου GMM LONG LEARN RATE, GMM SHORT LEARN RATE αντίστοιχα μοντέλου φόντου και το μέγεθος ανάλυσης του βίντεου. Το μέγιστο και ελάχιστο πλήθος εικονοστοιχείων στατικού φόντου είναι ιδιαίτερης σημασίας γιατί δίνεται η δυνατότητα στο χρήστη να ορίζει το μέγεθος των αντικειμένων που τον απασχολούν και μπορεί να προσαρμόζει το σύστημα ανάλογα με τις απαιτήσεις του. Επίσης ο ορισμός του ρυθμού μάθησης είναι σημαντικός για το χρόνο που θα προσαρμόζονται αντίστοιχα το μακροπρόθεσμο και βραχυπρόθεσμο μοντέλο.

4.7.1 Με είσοδο βίντεο

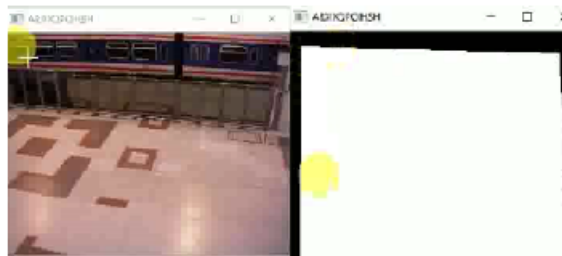
Στην εκτέλεση του προγράμματος με πηγή εισόδου ένα βίντεο, ο χρήστης πληκτρολογεί την τιμή ένα και την αντίστοιχη θέση του αρχείου που είναι αποθηκευμένο το βίντεο (εικόνα 4.3)



```
C:\Users\Asterios\Documents\Visual Studio 2012\Projects\test_1\Debug\tes..
1: video 2: camera
1
please input video filename: ***.avi or ***.mp4
C:\Users\Asterios\Desktop\ObjLeft-master\pets2006_1.avi
```

Σχήμα 4.3: Εισαγωγή Πηγής Εισόδου Βίντεο

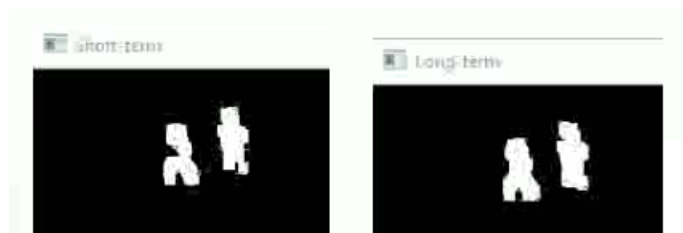
Στη συνέχεια ο χρήστης ορίζει την περιοχή του ενδιαφέροντος, δηλαδή την έκταση της εικόνας που επιθυμεί να εξεταστεί (εικόνα 4.4).



Σχήμα 4.4: Επιλογή Περιοχής ενδιαφέροντος

Έπειτα ακολουθεί μια χρονική περίοδος αποθήκευσης του αριθμού των καρτέ που επιθυμεί ο χρήστης να χρησιμοποιηθούν για την εκπαίδευση του μοντέλου φόντου. Αφού ολοκληρωθεί η συλλογή των καρτέ ξεκινά το βίντεο. Στα αντικείμενα που κινούνται και αποτελούν προσκήνιο, τα αποτελέσματα σε n χρονικές τιμές ήταν σχεδόν μηδαμινής διαφοράς, εξαιτίας ότι ένα αντικείμενο προσκήνιου με το ρυθμό μάθη-

σης που ορίστηκε δεν προλαβαίνει να αφομοιωθεί ως αντικείμενο φόντου ούτε στο μακροπρόθεσμο ούτε στο βραχυπρόθεσμο μοντέλο (σχήμα 4.5-4.6).



Σχήμα 4.5: Αριστερά: Βραχυπρόθεσμο Μοντέλο $n+5$, Δεξιά: Μακροπρόθεσμο $n+5$



Σχήμα 4.6: Αριστερά: Βραχυπρόθεσμο Μοντέλο $n+10$, Δεξιά: Μακροπρόθεσμο $n+10$

Στη συνέχεια εμφανίζεται ο κάτοχος ενός σακιδίου εγκαταλείποντας στη περιοχή που καταγράφεται στο βίντεο. Παρατηρούμε την στιγμή που αφήνει το σακίδιο στο έδαφος το βραχυπρόθεσμο μοντέλο έχει προλάβει να χαρακτηρίζει το αντικείμενο φόντου σε αντίθεση με το μακροπρόθεσμο που αποτελεί αντικείμενο προσκηνίου για μια σειρά χρονικών τιμών.(σχήμα 4.7-4.9)



Σχήμα 4.7: Αριστερά: Βραχυπρόθεσμο Μοντέλο $n+5$, Δεξιά: Μακροπρόθεσμο $n+5$

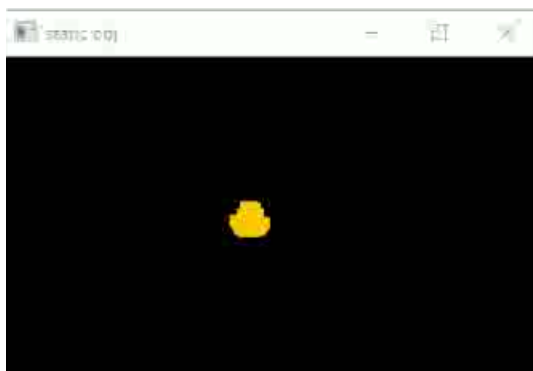


Σχήμα 4.8: Αριστερά: Βραχυπρόθεσμο Μοντέλο $n+10$, Δεξιά: Μακροπρόθεσμο $n+10$



Σχήμα 4.9: Αριστερά: Βραχυπρόθεσμο Μοντέλο $n+20$, Δεξιά: Μακροπρόθεσμο $n+20$

Ο συνδυασμός των δυο μοντέλων μακροπρόθεσμου και βραχυπρόθεσμου είναι κατάλληλος για την ανίχνευση ενός στατικού αντικειμένου. Όπως έχει αναφερθεί ένα εικονοστοιχείο μπορεί να αναπαρασταθεί ως ένα δίμπιτος κωδικός S_i . Όταν λάβει την τιμή $S_i=11$ υποδεικνύει ότι το εικονοστοιχείο αντιστοιχεί σε ένα κινούμενο αντικείμενο σκληρής (εικόνα 4.10)



Σχήμα 4.10: Ορισμός στατικού αντικειμένου

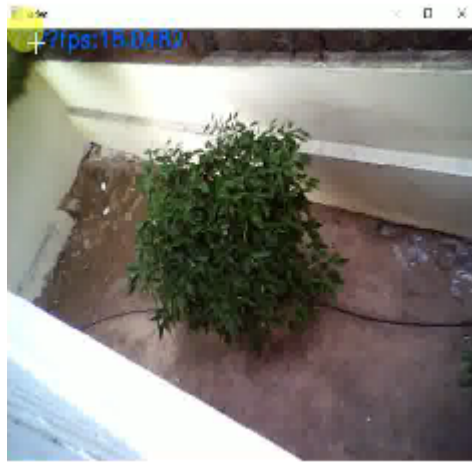
Μέσω του μοντέλου μηχανής πεπερασμένων καταστάσεων και της επαλήθευσης της οπίσθιας παρακολούθησης αν ικανοποιείται και η παράμετρος του μέγιστου και ελάχιστου πλήθους εικονοστοιχείων, προκύπτει και ο αντίστοιχος συναγερμός με την αντίστοιχη εικόνα σύνοψης της στιγμής που εγκαταλείφθηκε το αντικείμενο (σχήμα 4.11)



Σχήμα 4.11: Αριστερά:Εμφάνιση Μηνύματος Συναγεραμού Δεξιά: Σύνοψη

4.7.2 Με είσοδο βίντεο πραγματικού χρόνου

Σε αυτή την περίπτωση ο χρήστης επιλέγει την είσοδο βίντεο πραγματικού χρόνου πληκτρολογώντας την τιμή 2. Αρχικά συλλέγονται 500 καρέ για την εκπαίδευση του φόντου όπως και στη προηγούμενη περίπτωση. Το αρχικό φόντο της εικόνας φαίνεται στη παρακάτω εικόνα (σχήμα 4.12).

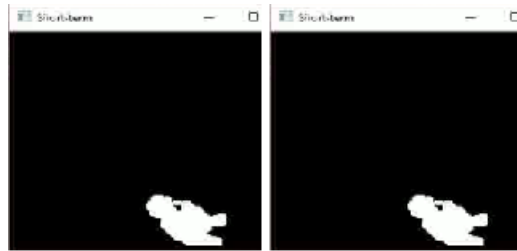


Σχήμα 4.12: Αρχικό Φόντο

Στη συνέχεια εμφανίζεται ένας άνθρωπος κρατώντας μια βαλίτσα. Όπως έχει προαναφερθεί ο ιδιοκτήτης της βαλίτσας αποτελεί το προσκήνιο στη σκηνή, παρατηρώντας ότι τόσο στο βραχυπρόθεσμο όσο και στο μακροπρόθεσμο μοντέλο παραμένει ενεργός για όσο αυτός κινείται (σχήμα 4.13-4.14).

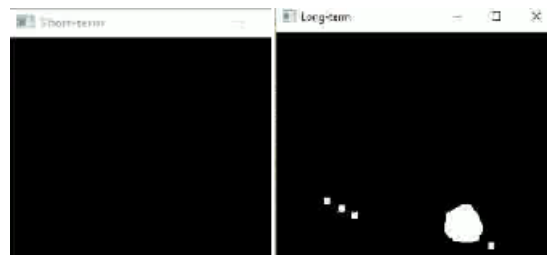


Σχήμα 4.13: Αριστερά: RGB Απεικόνιση, Κέντρο: Βραχυπρόθεσμο Μοντέλο $n+5$ Δεξιά: Μακροπρόθεσμο $n+5$

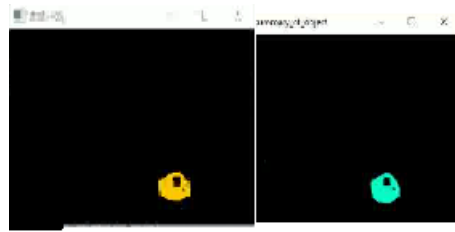


Σχήμα 4.14: Αριστερά: Βραχυπρόθεσμο Μοντέλο $n+5$, Δεξιά: Μακροπρόθεσμο $n+5$

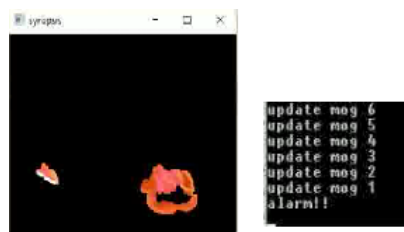
Έπειτα ο ιδιοκτήτης της βαλίτσας αφήνει την βαλίτσα και την εγκαταλείπει αποχωρώντας από το μοντέλο του φόντου. Με την πάροδο 30 δευτερολέπτων η βαλίτσα θεωρείται στατικό αντικείμενο στο μακροπρόθεσμο μοντέλο ενώ στο βραχυπρόθεσμο όχι, (σχήμα 4.15) και ξεκινά η διαδικασία του PSFM μοντέλου και της επαλήθευσης οπίσθιας παρακολούθησης. Εφόσον ικανοποιηθούν τα δυο παραπάνω δημιουργείται ένα παράθυρο επιβεβαίωσης και ένα παράθυρο σύνοψης του αποτελέσματος, ότι το στατικό αντικείμενο αποτελεί εγκαταλελειμμένο αντικείμενο χτυπώντας και ο σχετικός συναγερμός (σχήμα 4.15-4.17).



Σχήμα 4.15: Αριστερά: Βραχυπρόθεσμο Μοντέλο $n+5$, Δεξιά: Μακροπρόθεσμο $n+5$



Σχήμα 4.16: Αριστερά: Στατικό Αντικείμενο, Δεξιά: Σύνοψη αντικειμένου

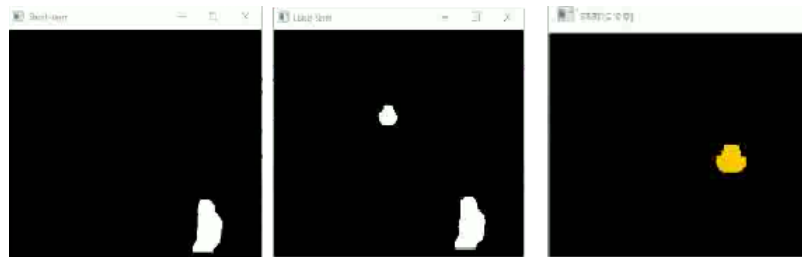


Σχήμα 4.17: Αριστερά: Καρέ Εγκατάλειψης, Δεξιά: Χτύπημα Συναγερμού

4.7.3 Με είσοδο βίντεο και ορισμό διαφορετικών MAXSFG-MINSFG

Εκτελέστηκαν μια σειρά πειραμάτων με διαφορετικά MAXSFG-MINSFG. Στη παρούσα υποενότητα παραθέτονται 2 πειράματα με παραμετροποιημένα MAXSFG-MINSFG.

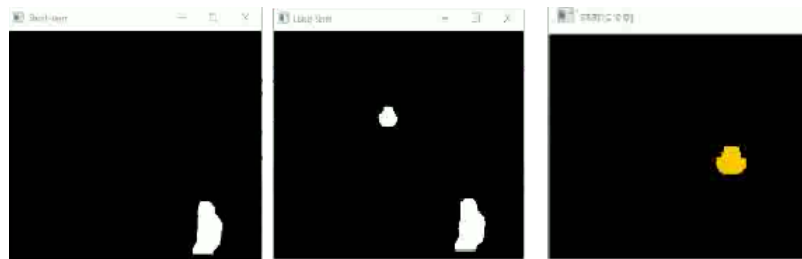
Στη πρώτη περίπτωση ορίστηκε το MAXSFG=10 και το MINSFG=5. Τα αποτελέσματα ήταν τα ακόλουθα (Σχήμα 4.18):



Σχήμα 4.18: Αριστερά: Βραχυπρόθεσμο Μοντέλο, Κέντρο: Μακροπρόθεσμο Μοντέλο, Δεξιά: Στατικό αντικείμενο

Στο συγκεκριμένο πείραμα δεν υπήρξε συναγερομός παρόλο που ο ιδιοκτήτης του αντικειμένου εγκατέλειψε το αντικείμενο επειδή η έκταση των εικονοστοιχείων του αντικειμένου δεν πληρούσαι τις προϋποθέσεις.

Στο επόμενο πείραμα ορίστηκε το $MAXSFG=120$ και το $MINSFG=10$. Τα αποτελέσματα ήταν τα ακόλουθα (Σχήμα 4.19):



Σχήμα 4.19: Αριστερά: Βραχυπρόθεσμο Μοντέλο, Κέντρο: Μακροπρόθεσμο Μοντέλο, Δεξιά: Στατικό αντικείμενο

Στο συγκεκριμένο πείραμα υπήρξε συναγερομός επειδή ο ιδιοκτήτης του αντικειμένου εγκατέλειψε το αντικείμενο και η έκταση των εικονοστοιχείων του αντικειμένου πληρούσαι τις προϋποθέσεις.

4.8 Σύνοψη

Με το συνδυασμό του μακροπρόθεσμου και βραχυπρόθεσμου μοντέλου επιτυγχάνεται ο σωστός διαχωρισμός του προσκηνίου από το φόντο. Στη συνέχεια ακολουθεί το μοντέλο μηχανής πεπερασμένων καταστάσεων μέσω του οποίου επιτυγχάνεται ο χαρακτηρισμός ενός αντικειμένου ως στατικό. Ακολουθεί η διαδικασία της οπίσθιας παρακολούθησης όπου οριστικοποιείται και ο χαρακτηρισμός ενός αντικείμενου ως εγκαταλελειμμένου.

Κεφάλαιο 5

Πειραματική Διαδικασία

Η εκτέλεση των πειραμάτων διενεργήθηκε σε λειτουργικό σύστημα Windows 10. Το προγραμματιστικό περιβάλλον που χρησιμοποιήθηκε ήταν το Visual Studio version 2012 με την χρήση της βιβλιοθήκης του OpenCV. Η βιβλιοθήκη αυτή αποτελεί λογισμικό ανοικτού κώδικα, με τεχνικές ψηφιακής επεξεργασίας εικόνας και αλγορίθμων μηχανικής μάθησης που αποτελεί πολύ σημαντικό εργαλείο του ευρύτερου επιστημονικού πεδίου της μηχανικής όρασης.

Στη πειραματική διαδικασία που υλοποιήθηκε, εξετάστηκε η ακρίβεια της μεθοδολογίας για τρεις διαφορετικές περιπτώσεις. Στη πρώτη περίπτωση εφαρμόστηκε μόνο ο συνδυασμός μακροπρόθεσμου και βραχυπρόθεσμου μοντέλου φόντου DualBG. Η εφαρμογή της συγκεκριμένης μεθόδου δεν απέφερε ικανοποιητικά αποτελέσματα, εξαιτίας του θορύβου της εικόνας και της ατελούς διαδικασίας αφαίρεσης του φόντου. Στη δεύτερη περίπτωση, εκτελέστηκε πείραμα με χρήση του DualBG και PFSM χωρίς την επαλήθευση οπίσθιας παρακολούθησης. Με αυτό το μοντέλο όπως έχει προαναφερθεί, επιτυγχάνεται ο χαρακτηρισμός

της κατάστασης ενός εικονοστοιχείου, μέσω της μεθόδου PFSM για το αν αυτό αποτελεί εικονοστοιχείο στατικού προσκηνίου. Όπως είναι αναμενόμενο, η εφαρμογή αυτής της μεθόδου, προσφέρει καλύτερα αποτελέσματα από αυτά της προηγούμενης, αφού εξασφαλίζει τον χαρακτηρισμό ενός εικονοστοιχείου ως στατικό παρόλα αυτά η ακρίβεια παραμένει μη ικανοποιητική. Η αιτία είναι ότι ο κάτοχος ενός αντικειμένου, μπορεί να το αφήσει στη περιοχή ενδιαφέροντος για κάποιο χρονικό διάστημα, αυτό να οριστεί στατικό αντικείμενο προσκηνίου αλλά ταυτόχρονα να χαρακτηριστεί λανθασμένα εγκαταλελειμμένο. Η τρίτη μέθοδος έρχεται για να επιλύσει το συγκεκριμένο ζήτημα, όπου συνδυάζει την προηγούμενη μέθοδο, με την επαλήθευση οπίσθιας παρακολούθησης. Οι μετρικές αξιολόγησης που χρησιμοποιήθηκαν για την σύγκριση των τριών περιπτώσεων ήταν το precision και το recall. Στη σειρά των πειραμάτων που υλοποιήθηκαν, σκοπός ήταν ο εντοπισμός εγκαταλελειμμένων (έγκυρο χτύπημα συναγερμού) αντικειμένων εντός μιας περιοχής.

Ως precision ορίζεται ο λόγος των αληθώς ενεργοποιημένων συναγερμών προς τον αριθμό των συναγερμών που ορθώς ή ψευδώς χαρακτήρισαν το αντικείμενο ως εγκαταλελειμμένο. Το recall αποτελεί τον λόγο των αληθώς ενεργοποιημένων συναγερμών προς το σύνολο των συναγερμών που θα έπρεπε να λάβουμε.

Όπως παρατηρείται και στο πίνακα η καλύτερη μέθοδος μεταξύ των τριών είναι ο συνδυασμός του PFSM-BackTracing όπου το precision που παρουσιάζεται είναι της τάξεως του 96%. Στη μέθοδο του PFSM το precision είναι 46% ενώ στη λιγότερο αποδοτική μέθοδο το precision είναι μόλις στο 4%.

Και στις τρεις περιπτώσεις το recall είναι ίσο με 1 γιατί δεν υπάρχει ποτέ περίπτωση, που να μην χτυπήσει ο συναγερμός όταν ένα αντικείμενο σταματάει να κινείται. Αντιθέτως το precision μεταβάλλεται γιατί ο αριθμός των λανθασμένων συναγερμών κυμαίνεται σε διαφορετικές τιμές σε κάθε περίπτωση. Στη πρώτη έχουμε μεγάλο αριθμό λανθασμένων συναγερμών, στη δεύτερη εμφανώς λιγότερα, ενώ στη τρίτη καθόλου εξαιτίας της επαλήθευσης της οπίσθιας παρακολούθησης.

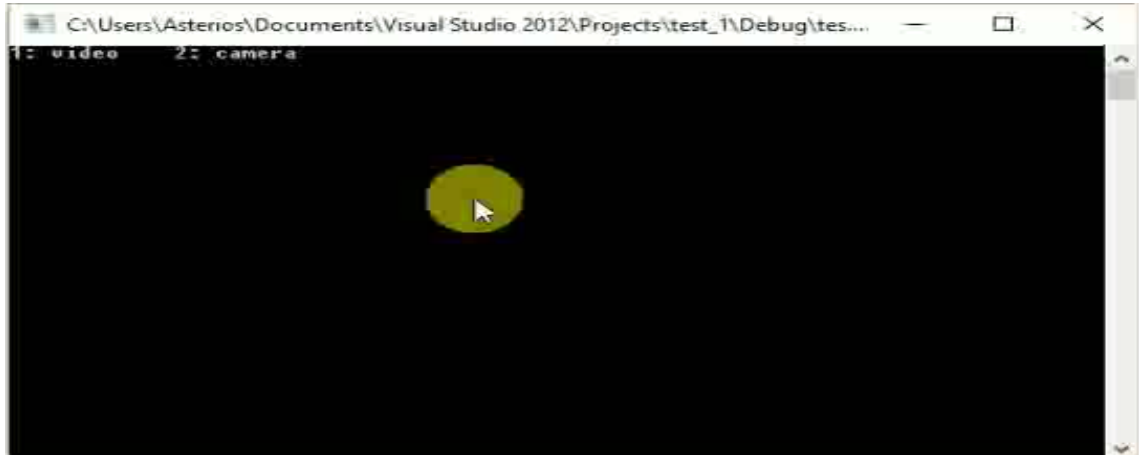
Ο καθορισμός των παραμέτρων του ρυθμού μάθησης λ_S , λ_L έγινε με ευριστικές διαδικασίες. Ο ρυθμός μάθησης του μακροπρόθεσμου μοντέλου ορίστηκε 10 φορές μεγαλύτερος από τον ρυθμό μάθησης του βραχυπρόθεσμου ώστε να εξασφαλιστεί σαφής διάκριση των μοντέλων. Το μέγιστο μέγεθος στατικού προσκηνίου σε πλήθος εικονοστοιχείων MAXSFG ορίστηκε έτσι ώστε να είναι μικρότερο από το μέσο μέγεθος των ανθρώπων που παρουσιάζονται μέσα στα καρέ. Το ελάχιστο μέγεθος στατικού προσκηνίου σε πλήθος εικονοστοιχείων MINSFG ορίστηκε πολύ χαμηλά ώστε να συλλαμβάνει και μικρά αντικείμενα αλλά και μεγάλα τα οποία βρίσκονται μακριά από την κάμερα. Χρησιμοποιήθηκαν τα πρώτα 500 καρέ για την εκμάθηση του μοντέλου φόντου.

Πίνακας Πειραμάτων 1

	<i>method</i>	<i>Precision</i>	<i>Recall</i>
1	<i>DualBG</i>	0.04	1.00
2	<i>PFSM</i>	0.46	1.00
3	<i>PFSM – BackTracing</i>	0.96	1.00

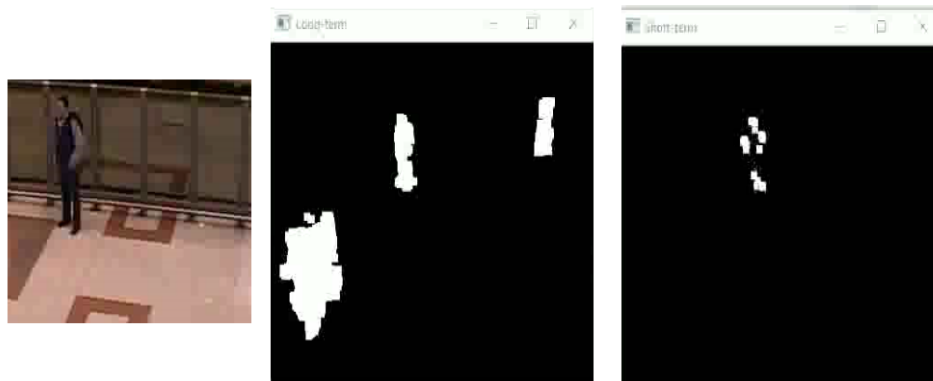
Στη συνέχεια υπάρχει μια ανασκόπηση ενός πειράματος που υλοποιήθηκε με την χρήση της τρίτης και πιο αποδοτικής μεθόδου. Αρχικά

στο πρόγραμμα εισάγουμε το βίντεο που αποτέλεσε και την βάση των πειραμάτων μας.



Σχήμα 5.1: Είσοδος Βίντεο

Στη συνέχεια έχουμε την εμφάνιση ενός ατόμου στη σκηνή παρατηρώντας την διαμόρφωση του μακροπρόθεσμου και βραχυπρόθεσμου μοντέλου.



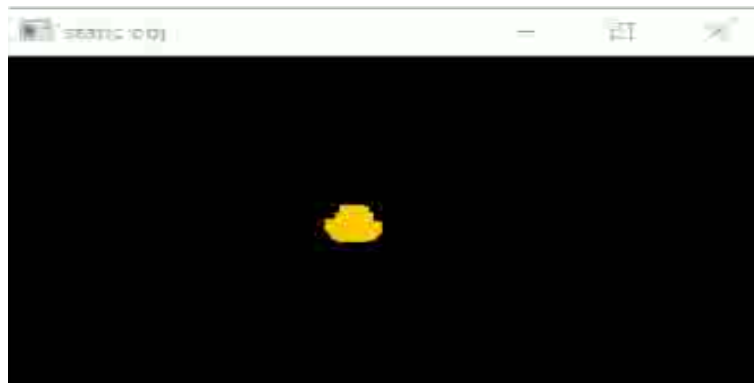
Σχήμα 5.2: Αριστερά: RGB μοντέλο Κέντρο: Μακροπρόθεσμο Μοντέλο
Δεξιά: Βραχυπρόθεσμο Μοντέλο

Στη συνέχεια του βίντεο ένα άτομο αφήνει μια τσάντα στην περιοχή καταγραφής και απομακρύνεται.



Σχήμα 5.3: Αριστερά: RGB μοντέλο Κέντρο: Μακροπρόθεσμο Μοντέλο Δεξιά: Βραχυπρόθεσμο Μοντέλο

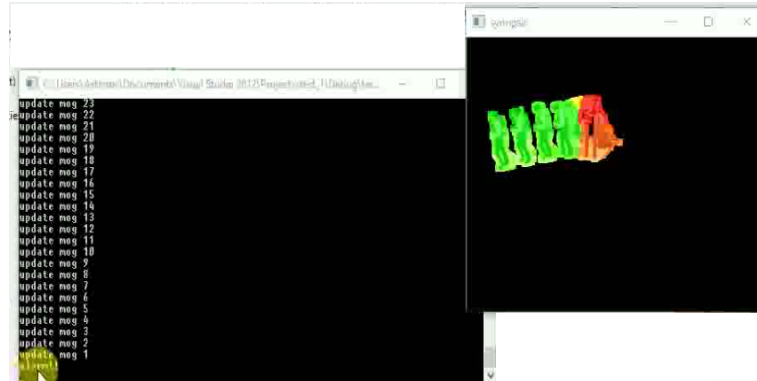
Με την υλοποίηση της τρίτης μεθόδου αφού οριστεί το αντικείμενο ως στατικό αντικείμενο προσκηνίου ελέγχουμε το παρακάτω αποτέλεσμα.



Σχήμα 5.4: Χαρακτηρισμός περιοχής ως στατικό αντικείμενο

Αφού ικανοποιηθεί και η συνθήκη του μοντέλου επαλήθευσης οπί-

σθιας παρακολούθησης το αντικείμενο χαρακτηρίζεται εγκαταλελειμμένο και παίρνουμε το μήνυμα συναγερμού στο τερματικό.



Σχήμα 5.5: Ορισμός αντικειμένου ως εγκαταλελειμμένο

5.1 Σύνοψη

Από την προαναφερθείσα πειραματική διαδικασία, διαπιστώθηκε ότι η καλύτερη μέθοδος με τα πιο ικανοποιητικά αποτελέσματα ήταν ο συνδυασμός του μοντέλου μηχανής πεπερασμένων καταστάσεων με την επαλήθευση οπίσθιας παρακολούθησης. Στις άλλες δυο μεθόδους χρησιμοποίησης μόνο του διπλού μοντέλου φόντου και στη χρησιμοποίηση μόνο του μοντέλου μηχανής πεπερασμένων καταστάσεων, στα αποτελέσματα των πειραμάτων προέκυπταν αρκετοί λανθασμένοι συναγερμοί για υπάρχοντα εγκαταλελειμμένα αντικείμενα.

Κεφάλαιο 6

Σύνοψη - Συμπεράσματα

Στη παρούσα διπλωματική εργασία αναπτύχθηκε αλγόριθμος εντοπισμού εγκαταλελειμμένου αντικειμένου. Χρησιμοποιήθηκαν τρία μοντέλα μεθοδολογίας το DualBG, PFSM και το Back Tracing Verification. Από την υλοποίηση των παραπάνω πειραμάτων διαπιστώθηκε ότι μεταξύ των τριών περιπτώσεων του DualBG του PFSM και του συνδυασμού PFSM με Back Tracing Verification η πιο συνεπής και ακριβής στο στόχο περίπτωση είναι η τρίτη. Στο DualBG λήφθηκαν οι περισσότεροι λανθασμένοι συναγερμοί εξαιτίας της δυσκολίας του διαχωρισμού και αφαίρεσης του φόντου από τα καταγεγραμμένα καρέ. Όπως έχει αναφερθεί σε αυτή τη διαδικασία ο αλγόριθμος ξεκινά από μια γενική μέθοδο όπου το φόντο μοντελοποιείται με την χρήση δυο μαθησιακών ρυθμών μάθησης το μακροπρόθεσμο και το βραχυπρόθεσμο. Ο συνδυασμός τους είναι ικανός για να εντοπίσει ένα στατικό αντικείμενο προσκηνίου και να το χαρακτηρίσει ως ακίνητο, άλλα δεν δύναται να πιστοποιήσει αν είναι εγκαταλελειμμένο αφού δεν ορίζεται ένα χρονικό διάστημα που αν ικανοποιείται να το πιστοποιεί ως τέτοιο. Στην επό-

μενη περίπτωση μεθοδολογίας PFSM ικανοποιείται μια τέτοια συνθήκη για να χαρακτηριστεί ένα αντικείμενο ως εγκαταλελειμμένο. Στη τελευταία μεθοδολογία, πέρα της ικανοποίησης του χρονικού κατωφλίου, του χρόνου δηλαδή που απαιτείται για να χαρακτηριστεί ένα αντικείμενο στατικό, απαιτείται και η ικανοποίηση ενός χωρικού κανόνα, να έχει το αντικείμενο αφηθεί από τον ιδιοκτήτη του, έχοντας απομακρυνθεί μιας απόστασης που αν ικανοποιείται να ενεργοποιείται ο συναγερμός. Η ικανοποίηση και των δυο συνθηκών, εξασφαλίζει καλύτερα αποτελέσματα από τις δυο προηγούμενες μεθόδους.

Στη περίπτωση του PFSM τα αποτελέσματα ήταν κατά 42% πιο ακριβή σε σχέση με την πρώτη περίπτωση εξαιτίας της ακρίβειας που προσδίδει ο χαρακτηρισμός του κάθε εικονοστοιχείου αν αποτελεί εικονοστοιχείο στατικού προσκηνίου. Το μειονέκτημα για να πιστοποιηθεί αν ένα αντικείμενο είναι εγκαταλελειμμένο είναι ότι ο χαρακτηρισμός μόνο των εικονοστοιχείων αν αποτελούν στατικά εικονοστοιχεία δεν αρκεί για να χτυπήσει αληθής συναγερμός. Ο συνδυασμός του PFSM με την επαλήθευση οπίσθιας παρακολούθησης προσφέρει καλύτερα αποτελέσματα με διαφορά precision 50%. Επιπρόσθετα συμπεράσματα που εξάχθηκαν από την πειραματική διαδικασία αφορούν την τιμή των απαιτούμενων καρτέ που ορίζει ο χρήστης για εκπαιδευτεί το μοντέλο φόντου. Έπειτα από ευριστικές μεθόδους, η βέλτιστη τιμή καρτέ είναι τα 500 καρτέ. Σε ορισμό λιγότερων καρτέ το αποτέλεσμα του εντοπισμού έπεφτε σε απόδοση και σε περισσότερα καρτέ υπήρχε το πρόβλημα των απαιτούμενων υπολογιστικών πόρων.

Κεφάλαιο 7

Μελλοντική Εργασία

Αντικείμενο μιας Μελλοντικής εργασίας μπορεί να αποτελέσει μια άμεση αλληλεπίδραση του χρήστη του προγράμματος με το runtime περιβάλλον του. Συγκεκριμένα όταν ενεργοποιηθεί κάποιος συναγεργμός για ένα αντικείμενο, ο χρήστης που εποπτεύει τα αποτελέσματα του βίντεο να μπορεί αφότου εξετάσει το εντοπισμένο εγκαταλελειμμένο αντικείμενο ως προς την επικινδυνότητά του, να επιστρέφει στο περιβάλλον του προγράμματος και να χαρακτηρίζει αν το αντικείμενο είναι ύποπτο. Εφόσον η απάντηση είναι θετική, ο αλγόριθμος θα αποθηκεύει την αποτύπωση ενός καρέ του αντικειμένου σε RGB μοντέλο σε μια βάση δεδομένων. Το μοντέλο χρησιμοποιώντας τα καρέ από παρελθοντικά ύποπτα αντικείμενα θα μπορεί να κάνει μια σύγκριση με μελλοντικά εντοπισμένα ως εγκαταλελειμμένα αντικείμενα και ανάλογα την ομοιότητά τους να ενεργοποιείται συναγεργμός υψηλότερου επιπέδου ασφαλείας.

Βιβλιογραφία

- [1] Smith, Kevin C and Quelhas, Pedro and Gatica-Perez, Daniel. Detecting abandoned luggage items in a public space. 2006.
- [2] Porikli, Fatih and Ivanov, Yuri and Haga, Tetsuji. Robust abandoned object detection using dual foregrounds. *EURASIP Journal on Advances in Signal Processing*, 2008(1):197875, 2007.
- [3] Tian, YingLi and Feris, Rogerio Schmidt and Liu, Haowei and Hampapur, Arun and Sun, Ming-Ting. Robust detection of abandoned and removed objects in complex surveillance videos. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 41(5):565–576, 2011.
- [4] Fan, Quanfu and Pankanti, Sharath. Modeling of temporarily static objects for robust abandoned object detection in urban surveillance. In *Advanced Video and Signal-Based Surveillance (AVSS), 2011 8th IEEE International Conference on*, pages 36–41. IEEE, 2011.
- [5] Evangelio, Rubén Heras and Senst, Tobias and Sikora, Thomas. Detection of static objects for the task of video surveillance. In

-
- Applications of Computer Vision (WACV), 2011 IEEE Workshop on*, pages 534–540. IEEE, 2011.
- [6] Amaleswarao, K and Vijayadeep, G and Shivaji, U. Improved Background Matching Framework For Motion Detection. *the International Journal of Computer Trends and Technology*, 4, 2013.
- [7] Guezouli, Larbi and Belhani, Hanane. Automatic detection of moving objects in video surveillance. *arXiv preprint arXiv:1608.03617*, 2016.
- [8] Brutzer, Sebastian and Höferlin, Benjamin and Heidemann, Gunther. Evaluation of background subtraction techniques for video surveillance. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*.
- [9] Martínez-del-Rincón, Jesús and Herrero-Jaraba, J Elías and Gómez, J Raúl and Orrite-Urunuela, Carlos. Automatic left luggage detection and tracking using multi-camera ukf. In *Proceedings of the 9th IEEE International Workshop on Performance Evaluation in Tracking and Surveillance (PETS'06)*, 2006.
- [10] Bouwmans, Thierry. Recent advanced statistical background modeling for foreground detection-a systematic survey. *Recent Patents on Computer Science*, (3), 2011.
- [11] Liao, Huei-Hung and Chang, Jing-Ying and Chen, Liang-Gee. A localized approach to abandoned luggage detection with foreground-mask sampling. In *Advanced Video and Signal*

Based Surveillance, 2008. AVSS'08. IEEE Fifth International Conference on. IEEE, 2008.

- [12] Tambe, Sagar B and Kulhare, Deepak and Nirmal, MD and Prajapati, Gopal. Image Processing (IP) Through Erosion and Dilation Methods. 2013.
- [13] Heikkilä, Janne and Silvén, Olli. A real-time system for monitoring of cyclists and pedestrians. *Image and Vision Computing*, 22(7): 563–570, 2004.
- [14] Koskinen, Lasse and Astola, Jaakko and Neuvo, Yrjö. Soft morphological filters. In *Proc. SPIE*, 1991.
- [15] Friedman, Nir and Russell, Stuart. Image segmentation in video sequences: A probabilistic approach. In *Proceedings of the Thirteenth conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc., 1997.
- [16] Stauffer, Chris and Grimson, W Eric L. Adaptive background mixture models for real-time tracking. In *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.* IEEE, 1999.
- [17] Neri, Alessandro and Colonnese, Stefania and Russo, Giuseppe and Talone, Paolo. Automatic moving object and background separation. *Signal processing*, 1998.

- [18] Horprasert, Thanarat and Harwood, David and Davis, Larry S. A statistical approach for real-time robust background subtraction and shadow detection. Citeseer, 1999.
- [19] Haritaoglu, Ismail and Harwood, David and Davis, Larry S. W/sup 4: real-time surveillance of people and their activities. *IEEE Transactions on pattern analysis and machine intelligence*, 2000.
- [20] Wren, Christopher Richard and Azarbayejani, Ali and Darrell, Trevor and Pentland, Alex Paul. Pfunder: Real-time tracking of the human body. *IEEE Transactions on pattern analysis and machine intelligence*, 1997.
- [21] Elgammal, Ahmed and Harwood, David and Davis, Larry. Non-parametric model for background subtraction. *Computer Vision, ECCV 2000*, 2000.
- [22] Sheikh, Yaser and Shah, Mubarak. Bayesian modeling of dynamic scenes for object detection. *IEEE transactions on pattern analysis and machine intelligence*, 2005.
- [23] Oliver, Nuria M and Rosario, Barbara and Pentland, Alex P. A Bayesian computer vision system for modeling human interactions. *IEEE transactions on pattern analysis and machine intelligence*, 2000.
- [24] Lee, Dar-Shyang. Effective Gaussian mixture learning for video background subtraction. *IEEE transactions on pattern analysis and machine intelligence*, 2005.

-
- [25] Horn, Berthold. *Robot vision*. MIT press, 1986.
- [26] Jeeva, S and Sivabalakrishnan, M. Survey on background modeling and foreground detection for real time video surveillance. *Procedia Computer Science*, 2015.
- [27] Maddalena, Lucia and Petrosino, Alfredo. A self-organizing approach to background subtraction for visual surveillance applications. *IEEE Transactions on Image Processing*, 2008.
- [28] Hofmann, Martin and Tiefenbacher, Philipp and Rigoll, Gerhard. Background segmentation with feedback: The pixel-based adaptive segmenter. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on*. IEEE, 2012.
- [29] Lin, Kevin and Chen, Shen-Chi and Chen, Chu-Song and Lin, Daw-Tung and Hung, Yi-Ping. Abandoned object detection via temporal consistency modeling and back-tracing verification for visual surveillance. *IEEE Transactions on Information Forensics and Security*, 2015.
- [30] Felzenszwalb, Pedro F and Girshick, Ross B and McAllester, David and Ramanan, Deva. Object detection with discriminatively trained part-based models. *IEEE transactions on pattern analysis and machine intelligence*, 2010.
- [31] Dubout, Charles and Fleuret, François. Exact acceleration of linear object detectors. *Computer Vision–ECCV 2012*, pages 301–311, 2012.