



ΠΟΛΥΤΕΧΝΕΙΟ ΚΡΗΤΗΣ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ
ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Εικονική Μορφοποίηση Βίντεο για Εξατομικευμένη Εκγύμναση

Φίλιππος Μαρκολέφας

ΕΞΕΤΑΣΤΙΚΗ ΕΠΙΤΡΟΠΗ

Καθ. Μιχάλης Ζερβάκης (Επιβλέπων)

Αν. Καθ. Αικατερίνη Μανιά

Καθ. Ευριπίδης Πετράκης

ΧΑΝΙΑ, ΜΑΙΟΣ 2018

Ευχαριστίες

Πρώτα, θα ήθελα να ευχαριστήσω τους Γονείς μου για την αμέριστη υποστήριξη τους όλα αυτά τα χρόνια που σπούδαζα στο Πολυτεχνείο Κρήτης.

Θα ήθελα επίσης να ευχαριστήσω τους Άγγελο, Αποστόλη για όλες τις ωραίες στιγμές που περάσαμε στα Χανιά. Την Έφη για την συνολική υποστήριξη της, αλλά και για την συμμετοχή της σε ένα από τα βίντεο του πειραματικού μέρους της διπλωματικής.

Τέλος, θα ήθελα να ευχαριστήσω την βοηθό του εργαστηρίου Ντίνα Μοιρογιώργου για την βοήθεια της συνολικά. Τον επιβλέποντα καθηγητή της διπλωματικής Μιχάλη Ζερβάκη, για την συνολική συνεργασία που είχαμε. Επίσης, τους καθηγητές της εξεταστικής επιτροπής Αικ. Μανιά και Ευρ. Πετράκη για την συμμετοχή τους στην αξιολόγηση της εργασίας μου.

Περιεχόμενα

Περίληψη	7
Abstract	9
1. ΚΕΦΑΛΑΙΟ – ΕΙΣΑΓΩΓΗ.....	11
1.1 Γενική Περιγραφή	11
1.2 Σκοπός της Εργασίας	11
1.3 Ορισμός του Προβλήματος.....	11
1.4 Δεδομένα προς Επεξεργασία.....	13
1.5 Προτεινόμενη μέθοδος.....	14
2. ΚΕΦΑΛΑΙΟ – ΠΑΡΑΔΟΣΙΑΚΕΣ ΤΕΧΝΙΚΕΣ ΕΞΑΓΩΓΗΣ ΤΟΥ FOREGROUND.....	16
2.1 Βασικό Σχήμα Εξαγωγής του Foreground	16
2.2 Στατιστική Μοντελοποίηση του Background και Εξαγωγή του Foreground.....	19
2.2.1 Μονή Gaussian Συνάρτηση (Single Gaussian - SG)	21
2.2.2 Μίξη Gaussian Συναρτήσεων (Mixture Of Gaussians - MOG)	22
2.2.3 Συνάρτηση Πυκνότητας Πιθανότητας με Χρήση Πυρήνα (Kernel Density Estimation - KDE)	26
2.2.4 Οικογένεια Γενικευμένων Gaussian Συναρτήσεων (Generalized Gaussian Family - GGF)	29
2.2.5 Μηχανές Διανυσμάτων Υποστήριξης (Support Vector Machines - SVM).....	32
2.2.5 Εκμάθηση ΥΠΟΧΩΡΟΥ και Χρήση Βασικών Συστατικών Στοιχείων (Subspace learning with PCA , SL - PCA)	35
2.2.6 Αποτελέσματα & Συμπεράσματα	37
2.3 Αλγόριθμοι Ομαδοποίησης για Εξαγωγή του Foreground.....	40
2.3.1 K-mean Μοντέλο	40
2.3.2 Χρήση Codebook.....	42
2.3.3 Αποτελέσματα & Συμπεράσματα	44
2.4 Χρήση Νευρωνικών Δικτύων για Εξαγωγή του Foreground (Neural Networks).....	45
2.4.1 Αυτό-Οργανωτικό Νευρωνικό Δίκτυο (Self-Organizing Background Subtraction–SOBS)	46
2.5 Μοντελοποίηση του Background με Θεωρία Εκτίμησης.....	49

3.	ΚΕΦΑΛΑΙΟ – ΠΡΟΤΕΙΝΟΜΕΝΗ ΜΕΘΟΔΟΣ ΑΝΙΧΝΕΥΣΗΣ FOREGROUND ΑΝΤΙΚΕΙΜΕΝΩΝ	50
3.1	Εισαγωγή.....	50
3.2	Θεωρητικό Υπόβαθρο	51
3.2.1	Χρωματικό Μοντέλο $L^*a^*b^*$	51
3.2.2	Ανίχνευση Περιοχών Κίνησης	52
3.2.3	Τμηματοποίηση Εικόνας.....	56
3.3	Προτεινόμενη Μεθοδολογία	60
3.3.1	Απλή Συσχέτιση Τμηματοποιημένης Εικόνας και Περιοχών Κίνησης	60
3.3.2	Προτεινόμενη Μέθοδος Ανακατασκευής του Background (Background Reconstruction).....	61
3.3.3	Ασαφή Λογική Ανανέωσης της Περιοχής Κίνησης.....	68
3.3.4	Γραφικό Σχήμα Προτεινόμενης Μεθόδου (Flow Chart)	72
4.	ΚΕΦΑΛΑΙΟ – ΑΠΟΤΕΛΕΣΜΑΤΑ ΥΛΟΠΟΙΗΣΗΣ	73
4.1	Μέτρηση Αποτελεσματικότητας – Απόδοσης Αλγορίθμου.....	73
4.2	Δεδομένα Προς Επεξεργασία	74
4.2.1	Βίντεο με Ασκήσεις Yoga (60 frames, resolution 854×480, 25 fps)	74
4.2.2	Βίντεο από Κάμερα Παρακολούθησης σε Σταθμό του Μετρό (i-LIDS dataset [130]) (190 frames, resolution 720×576, 30 fps)	75
4.2.3	Βίντεο από Κάμερα Παρακολούθησης σε Δρόμο 1 ([131]) (50 frames, resolution 854×480, 24 fps)	77
4.2.4	Βίντεο από Κάμερα Παρακολούθησης σε Δρόμο 2 ([132]) (60 frames, resolution 1020×580, 30 fps)	78
4.2.5	Βίντεο από Εσωτερικό Περιβάλλον με Ανομοιόμορφο Background.....	80
4.3	Χρόνοι Εκτέλεσης Αλγορίθμων.....	80
5.	ΚΕΦΑΛΑΙΟ - ΣΥΜΠΕΡΑΣΜΑΤΑ.....	81
6.	ΒΙΒΛΙΟΓΡΑΦΙΑ.....	83

Λίστα Εικόνων

Εικόνα 1.1 Στιγμιότυπο από το βίντεο με ασκήσεις yoga (Αριστερά). Το αποτέλεσμα της προτεινόμενης μεθόδου (Δεξιά).	15
Εικόνα 2.1 Ασπρόμαυρο βίντεο με yoga, frame (Αριστερά). Χωρίς την εφαρμογή IIR φίλτρου (Δεξιά).	18
Εικόνα 2.2 Ασπρόμαυρο βίντεο με yoga, frame (Αριστερά). Με την εφαρμογή IIR φίλτρου (Δεξιά).	18
Εικόνα 2.3 Ασπρόμαυρο βίντεο με το περπάτημα ενός ατόμου, frame (Αριστερά) [134]. Το αποτέλεσμα (με ίδιο όριο και ίδια frames κατά την διαδικασία της αρχικοποίησης) με ή χωρίς το selectivity για διαφορετικό παράγοντα εκμάθησης.	18
Εικόνα 2.4 Διάγραμμα PDF. Κανονική ή Gaussian Κατανομή παράδειγμα Unimodal κατανομής.....	19
Εικόνα 2.5 Στην εικόνα φαίνεται η θέση του pixel καθώς και τα χρώματα που αναπαριστά κατά την διάρκεια των n frames.	20
Εικόνα 2.6 Παράδειγμα ιστογράμματος στο χρόνο. Αριθμός frames n=50. Κάθετος άξονας, πλήθος τιμών για κάθε χρώμα στο χρόνο. Οριζόντιος άξονας, διακριτά χρώματα (0~255). ..	20
Εικόνα 2.7 Γεωμετρικά χαρακτηριστικά κατανομών.....	20
Εικόνα 2.8 Μίξη Gaussian συναρτήσεων για την μοντελοποίηση του pixel.	22
Εικόνα 2.9 Το αποτέλεσμα της διαδικασίας KDE για διαφορετικά h.....	26
Εικόνα 2.10 Διαγράμματα Συναρτήσεων Πυρήνα (Kernel).....	27
Εικόνα 2.11 Γενικευμένη Gaussian Συνάρτηση (GGF).	29
Εικόνα 2.12 Παραδείγματα Κατανομών μιας Γενικευμένης Gaussian Συνάρτησης.	30
Εικόνα 2.13 Παράδειγμα διαχωρισμού κλάσεων από διαφορετικά υπερπίεδα. H3 το βέλτιστο υπερπίεδο.	32
Εικόνα 2.14 PCA από μια Πολυδιάστατη Gaussian κατανομή με κέντρο το (1,3) και τυπική απόκλιση 3 στην κατεύθυνση (0.866, 0.5) και 1 στην ορθογώνια σε αυτήν. Τα διανύσματα που φαίνονται είναι τα ιδιοδιανύσματα του πίνακα συνδιακύμανσης κανονικοποιημένα από την αντίστοιχη ιδιοτιμή και μετατοπισμένα έτσι ώστε τα βελάκια να είναι ίδια με το μέσο. ..	35
Εικόνα 2.15 Κανονικοποιημένο αποτελέσματα μοντελοποίησης του ιστογράμματος του pixel με Gaussian κατανομή (Μπλε πλαίσιο) και με Laplace κατανομή (Πράσινο πλαίσιο), για διαφορετικές τιμές της υπερβάλλουσας κύρτωσης g2.	38
Εικόνα 2.16 Στην πρώτη σειρά βλέπουμε το πρώτο frame και στην δεύτερη το τελευταίο frame που δέχεται σαν είσοδο η μέθοδος MOG. Στην τελευταία σειρά φαίνεται το αποτέλεσμα της μεθόδου MOG για το τελευταίο frame για RGB βίντεο και παραμέτρους: Εκπ. περίοδος N=50, K=5, T=0.8, παράγοντας εκμάθησης a=0.005.....	39
Εικόνα 2.17 Χρωματική παραμόρφωση (Color distortion) μεταξύ εισερχόμενου pixel και στοιχείο του codebook.....	43
Εικόνα 2.18 Στην πρώτη σειρά βλέπουμε το πρώτο frame και στην δεύτερη το τελευταίο frame που δέχεται σαν είσοδο η μέθοδος με την χρήση Codebook. Στην τελευταία σειρά φαίνεται το αποτέλεσμα της μεθόδου για το τελευταίο frame για RGB βίντεο και παραμέτρους: Εκπαιδευτική περίοδος N=50, e1=3, a=0.55, b=1.3.	44
Εικόνα 2.19 Παράδειγμα νευρωνικού δικτύου.....	45
Εικόνα 2.20 Παράδειγμα εικόνας (Αριστερά), ο νευρωνικός χάρτης (Δεξιά). Nbg είναι η περιοχή όπου γίνεται η ανανέωση του background για $cmpt = b9$	48
Εικόνα 2.21 Στην πρώτη σειρά βλέπουμε το πρώτο frame και στην δεύτερη το τελευταίο frame που δέχεται σαν είσοδο η μέθοδος με την χρήση Νευρωνικού Δικτύου. Στην τελευταία σειρά φαίνεται το αποτέλεσμα της μεθόδου για το τελευταίο frame για RGB βίντεο και παραμέτρους: Εκπαιδευτική περίοδος K=50, n=3, e1=0.5, e2=0.005, c2=0.1.....	48
Εικόνα 2.22 Wiener Φίλτρο με είσοδο διακριτό σήμα w[n].	49

Εικόνα 3.1 Παράδειγμα εικόνας (πάνω) και το ιστόγραμμα της (κάτω).	50
Εικόνα 3.2 Χρωματικό Μοντέλο $L^*a^*b^*$	51
Εικόνα 3.3 Το αποτέλεσμα της μοντελοποίησης του background με χρήση μέσης τιμής και εφαρμογή ενός ορίου στη χρωματική απόσταση μεταξύ της τρέχουσας εικόνας και της background εικόνας για το χρωματικό μοντέλο $L^*a^*b^*$.	52
Εικόνα 3.4 Αποτέλεσμα αφαίρεσης του background META την χρήση μορφολογικών αλγορίθμων.	53
Εικόνα 3.5 Παράδειγμα αφαίρεσης του background για τα χρωματικά μοντέλα RGB και $L^*a^*b^*$.	53
Εικόνα 3.6 Μορφολογική πράξη διαστολής μεταξύ δυαδικής εικόνας και ενός τετραγώνου 5×5 και στην συνέχεια 'γέμισμα' των εσωτερικών χώρων.	55
Εικόνα 3.7 Παράδειγμα τμηματοποίησης της εικόνας με τον αλγόριθμο K-means και απεικόνιση με χρήση διαγραμμάτων Voronoi, για τα χρωματικά μοντέλα L^*a^*b και RGB και αριθμό συστάδων $K=6$.	57
Εικόνα 3.8 Τμηματοποίηση της εικόνας και αναπαράσταση με τιμή για κάθε pixel την μέση τιμή της συστάδα, $K=6$.	58
Εικόνα 3.9 Παράδειγμα εξαγωγής foreground αντικειμένου στο βίντεο με τις ασκήσεις yoga. Το ποσοστιαίο όριο επικάλυψης για την ανίχνευση είναι $Throp = 0.7$ και αριθμός συστάδων $K=8$. Σε κόκκινο πλαίσιο δεξιά της εικόνας φαίνονται οι δυο περιορισμοί.	61
Εικόνα 3.10 Παράδειγμα τμηματοποίησης της εικόνας σε $K=6$ αριθμό συστάδων. Με λευκό χρώμα φαίνονται οι περιοχές για κάθε συστάδα K .	63
Εικόνα 3.11 Αποτέλεσμα προσδιορισμού εν δυνάμει foreground περιοχών (λευκές περιοχές) σύμφωνα με τον αλγόριθμό παραπάνω και $Throp = 0.1$, $K = 6$.	64
Εικόνα 3.12 Παράδειγμα υπολογισμού γειτονικών περιοχών μιας foreground περιοχής (λευκή περιοχή).	65
Εικόνα 3.13 Τελικό αποτέλεσμα ανακατασκευής του background από τα χρώματα του codebook χρησιμοποιώντας τον παραπάνω αλγόριθμο. Για καλύτερα αποτελέσματα έχει χρησιμοποιηθεί και ένα τετραγωνικό median φίλτρο μεγάλης διάστασης στο τελικό αποτέλεσμα.	66
Εικόνα 3.14 Παράδειγμα της εικόνας Upd a) την χρονική στιγμή t και την χρονική στιγμή $t+1$ b) μια περιοχή της εξαφανίζεται ξαφνικά λόγω αργής κίνησης του αντικειμένου (κόκκινο πλαίσιο).	67
Εικόνα 3.15 Παράδειγμα τμηματοποίησης της εικόνας σε blocks μεγέθους 61×60 .	68
Εικόνα 3.16 Παράδειγμα χρησιμότητας του Bounding Box και τελικώς τα blocks τα οποία είναι ικανά να ανανεωθούν ή όχι. Τα υπόλοιπα blocks παραμένουν ως έχει.	69
Εικόνα 3.17 Δεξιά φαίνονται τα pixels του block που αλλάζουν εξαιτίας της κίνησης του αντικειμένου (με μπλε περίγραμμα), ενώ σε κόκκινο πλαίσιο τα pixels του block του οποίου εξαφανίστηκαν λόγω αστοχίας της μεθόδου της Παραγράφου 1.	71
Εικόνα 4.1 Πίνακας Confusion για τις τιμές TP, TN, FP, FN και FPR TPR.	73
Εικόνα 4.2 Ενδεικτικά frames από το βίντεο.	74
Εικόνα 4.3 Αποτελέσματα χρήσης μεθόδου και μέτρηση αποτελεσματικότητας.	75
Εικόνα 4.4 Ενδεικτικά frames από το βίντεο.	76
Εικόνα 4.5 Αποτελέσματα χρήσης μεθόδου και μέτρηση αποτελεσματικότητας.	76
Εικόνα 4.6 Ενδεικτικά frames από το βίντεο.	77
Εικόνα 4.7 Αποτελέσματα χρήσης μεθόδου και μέτρηση αποτελεσματικότητας.	78
Εικόνα 4.8 Ενδεικτικά frames από το βίντεο.	78
Εικόνα 4.9 Αποτελέσματα χρήσης μεθόδου και μέτρηση αποτελεσματικότητας.	79
Εικόνα 4.10 Ενδεικτικά frames βίντεο και αποτελέσματα χρήσης μεθόδου	80

Περίληψη

Η εξατομικευμένη εκγύμναση μέσω υπολογιστή δίνει τη δυνατότητα σε χρηστές διαφόρων διαδικτυακών εφαρμογών να χρησιμοποιούν την κάμερα τους και ειδικά διαμορφωμένα βίντεο από επαγγελματίες γυμναστές, έτσι ώστε να γυμναστούν από την άνεση του σπιτιού τους. Προς αυτή την κατεύθυνση και με κύριο σκοπό ο χρήστης να μπορεί εύκολα να αναπαράγει τις ασκήσεις αυτές καλό είναι να υπάρχει έξτρα λειτουργικότητα στην εξατομικευμένη εκγύμναση μέσω υπολογιστή πέραν της απλής θέασης των βίντεο των ασκήσεων. Ένας τρόπος που θα μπορούσε να βοηθήσει προς αυτή την κατεύθυνση είναι η εξαγωγή της σιλουέτας του ατόμου που γυμνάζεται έτσι ώστε το άτομο να μπορεί εύκολα καταρχάς να παρακολουθεί τις κινήσεις του και, στη συνέχεια, να τις συγχρονίζει σύμφωνα με αυτές του επαγγελματία γυμναστή. Στην παρούσα διπλωματική, που στόχο έχει την εξαγωγή της σιλουέτας του κινούμενου αντικειμένου (ανθρώπου) από βίντεο, αρχικά ασχοληθήκαμε με παραδοσιακές μεθόδους που αφορούν στην αφαίρεση του παρασκήνιου (background subtraction) σε ένα βίντεο. Δυστυχώς, οι παραδοσιακές τεχνικές δεν μπορούν να αντιμετωπίσουν όλες τις προκλήσεις που μπορεί να υφίστανται σε ένα βίντεο και που εντοπίζονται και αναλύονται στην παρούσα εργασία. Γι' αυτό το λόγο, προτείνεται ένας συνδυασμός τεχνικών για την επιτυχή αφαίρεση της σιλουέτας ενός ατόμου σε ένα βίντεο με ασκήσεις γυμναστικής (επιλέχθηκε η άσκηση yoga). Η προτεινόμενη μέθοδος περιλαμβάνει, εκτός των άλλων, την αρχική ανακατασκευή του background και, στη συνέχεια, την επιλεκτική ανανέωση περιοχών για την προσαρμοστικότητα του background στα πραγματικά δεδομένα. Για τον προσδιορισμό των περιοχών απαιτείται γνώση που αφορά στις περιοχές κινήσεις σε ένα βίντεο καθώς και στις ομοιογενείς περιοχές σε ένα frame, γεγονός που χρησιμοποιήθηκε με κατάλληλο τρόπο για τον επιτυχή σχεδιασμό της μεθοδολογίας. Τέλος, εφαρμόστηκε η προτεινόμενη μέθοδος και σε βίντεο από διαφορετικά περιβάλλοντα, πέραν αυτών της εξατομικευμένης εκγύμνασης, με σκοπό να μετρηθεί η αποτελεσματικότητα της μεθόδου για διαφορετικές περιπτώσεις.

Abstract

Online personal training allow users to work out from the comfort of their own homes, using workout videos from fitness instructors. Moreover, users of such applications can also use their PC camera and work out with others in a group setting, which carries a plethora of intertwined benefits. Towards training efficiency enhancement and consistency, it could be helpful to extract human silhouette from a video which could be used afterwards on exposure to fitness instructor's video and make the fitness instructions easily followed by the user. In this thesis, initially we involved with traditional background estimation approaches as well as foreground extraction techniques for videos from static cameras. Unfortunately, none of the methods was able to face efficiently all the possible challenges, including slow moving foreground object and presence of foreground object during background initialization, which are dominant problems in our main video (yoga). Thus, we propose a series of techniques that consist of an initial background reconstruction method followed by a selective update scheme. The background image adaptively converges to ground truth data using the above scheme which combines information of detected moving regions (temporal) and color-based regions (spatial) of a video frame. Finally, we apply the proposed method in different environmental conditions such as video from surveillance cameras and we measure the efficiency of the proposed method.

1. Κεφάλαιο – Εισαγωγή

1.1 Γενική Περιγραφή

Με τη συνεχή άνοδο των διαδικτυακών εφαρμογών και την εισχώρησή τους σε διάφορους τομείς της ζωής μας, η ιδέα της εξατομικευμένης εκγύμνασης (personal training) από την άνεση του σπιτιού μας αποτελεί πραγματικότητα. Πρακτικά ο χρήστης, χρησιμοποιώντας την κάμερα του υπολογιστή ή του κινητού τηλεφώνου, μπορεί με τη βοήθεια κάποιων ασκήσεων γυμναστικής από επαγγελματίες γυμναστές (fitness instructors) να καταφέρει να συγχρονίσει τις κινήσεις του με αυτές του επαγγελματία γυμναστή έτσι ώστε η εκτέλεση των ασκήσεων να γίνεται σωστά. Το πρόβλημα που προκύπτει έγκειται στο ότι, ο χρήστης πρέπει να παρακολουθεί ταυτόχρονα τις κινήσεις του και τις κινήσεις του επαγγελματία γυμναστή, κάτι το οποίο του αποσπά την προσοχή και δεν διευκολύνει τη σωστή εκτέλεση των ασκήσεων.

Για το λόγο αυτό, ο τομέας της εικονικής μορφοποίησης βίντεο μπορεί να βοηθήσει το χρήστη τέτοιων εφαρμογών έτσι ώστε η παρακολούθηση και η εκτέλεση των ασκήσεων γυμναστικής να γίνεται πιο εύκολα. Αυτό που θα βοηθούσε στην παραπάνω περίπτωση είναι η σύμπτυξη των δύο εικόνων, του επαγγελματία γυμναστή και του χρήστη, σε μία, ώστε να φαίνονται εύκολα οι διαφοροποιήσεις. Για να γίνει αυτό θα πρέπει, είτε στο βίντεο με τον χρήστη είτε στο βίντεο με τον επαγγελματία γυμναστή προτιμότερα, να γίνει εξαγωγή της σιλουέτας του ατόμου με κάποιο τρόπο και στη συνέχεια να αποτυπωθεί η σιλουέτα πάνω στο άλλο βίντεο.

1.2 Σκοπός της Εργασίας

Στην παρούσα διπλωματική ασχοληθήκαμε μόνο με το κομμάτι της εξαγωγής της σιλουέτας του ατόμου σε ένα βίντεο. Σαν δείγμα βίντεο γυμναστικής μας δόθηκε ένα βίντεο με ασκήσεις γυμναστικής yoga όπου η κάμερα είναι στατική. Το συγκεκριμένο βίντεο, σε σχέση με άλλα, όπως από κλειστά κυκλώματα παρακολούθησης, έχει το χαρακτηριστικό ότι το αντικείμενο προς αναγνώριση βρίσκεται πάντα σε μια καθορισμένη θέση και κινείται γύρω από αυτήν.

Εκτός από το βίντεο της yoga, χρησιμοποιήθηκαν και άλλα βίντεο από διαφορετικά περιβάλλοντα με διαφορετικές συνθήκες για τον έλεγχο της απόδοσης των μεθόδων που εφαρμόστηκαν.

1.3 Ορισμός του Προβλήματος

Η ανάλυση και η εξαγωγή συμπερασμάτων από ένα βίντεο αποτελεί έναν ενεργό και σχετικά πρόσφατο επιστημονικό τομέα. Πολλές εφαρμογές στην καθημερινή μας ζωή (όπως συστήματα παρακολούθησης, εφαρμογές πολυμέσων), απαιτούν την ανίχνευση των κινούμενων αντικειμένων σε ένα πλάνο. Οπότε, η βασική λειτουργία έγκειται στο διαχωρισμό των κινούμενων αντικειμένων (foreground) από το στατικό περιβάλλον, αλλιώς από το φόντο (background). Οι παραδοσιακές τεχνικές που εφαρμόζονται αφορούν στην αφαίρεση του στατικού περιβάλλοντος, φόντου (background subtraction) από ένα βίντεο. Ο πιο απλός τρόπος μοντελοποίησης του

background είναι η καταγραφή του σκηνικού χωρίς να περιέχει κανένα κινούμενο αντικείμενο. Τις περισσότερες φορές αυτό δεν είναι εφικτό, επειδή το background μπορεί να αλλάξει εξαιτίας αλλαγών στις συνθήκες φωτισμού ή εξαιτίας των αντικειμένων του background που μετακινήθηκαν κατά την διάρκεια του βίντεο. Γι' αυτό το λόγο η μοντελοποίηση του background πρέπει να γίνει με τέτοιο τρόπο ώστε να αντιμετωπίζονται επαρκώς τέτοιες προκλήσεις.

Η αφαίρεση του στατικού περιβάλλοντος, του φόντου (background subtraction) βασίζεται στην παραδοχή ότι η διαφορά σε μια εικόνα που περιέχει μόνο το φόντο (background) και σε μια τωρινή εικόνα, προκαλείται από την παρουσία κινούμενων αντικειμένων. Εικονοστοιχειά (pixels) της εικόνας που δεν έχουν αλλάξει θεωρούνται 'background', ενώ pixels που έχουν αλλάξει χαρακτηρίζονται ως 'κινούμενα αντικείμενα' ή αλλιώς 'foreground'.

Στην πραγματικότητα ο ορισμός αυτός δεν καλύπτει πλήρως την συμπεριφορά των pixels, καθώς τα pixels μπορούν να ανήκουν στα κινούμενα αντικείμενα αλλά να μην καταγραφούν σαν foreground εξαιτίας του ότι τα pixels αυτά έχουν ίδιο χρώμα με το background. Ομοίως, pixels μπορούν να καταγραφούν λανθασμένα σαν κινούμενα αντικείμενα εξαιτίας της αλλαγής στις συνθήκες φωτισμού ή λόγω του θορύβου του αισθητήρα της κάμερας. Γι' αυτό το λόγο είναι απαραίτητο να οριστούν κάποιες βασικές παραδοχές που θα διέπουν τις παρούσες μεθόδους και τα υπό διερεύνηση βίντεο μας.

- **Παραδοχή 1:** Η κάμερα όπως και οι παράμετροι της πρέπει να είναι σταθερά κατά την διάρκεια καταγραφής του βίντεο.
- **Παραδοχή 2:** Για την αναγνώριση ενός pixel ως foreground απαιτείται το χρώμα του pixel να είναι διαφορετικό από εκείνο του background. Η διαφορετικότητα αφορά την κατηγοριοποίηση των χρωματικών περιοχών της εικόνας σε έναν αριθμό από βασικές. Εάν υπάρχει επικάλυψη περιοχών που μοιράζονται την ίδια βασική κατηγορία, όμως ανήκουν σε περιοχές του background και foreground αντίστοιχα τότε δεν μπορεί να γίνει σωστή ανίχνευση (πρόβλημα καμουφλάζ).
- **Παραδοχή 3:** Το background πρέπει να είναι σε θέση να προσαρμόζεται στις αλλαγές της φωτεινότητας αλλά και στις αλλαγές που προέρχονται από μετακινήσεις αντικειμένων από και προς το σκηνικό καταγραφής.
- **Παραδοχή 4:** Το background πρέπει να μπορεί να δημιουργηθεί σωστά ακόμα και εάν κατά τη διάρκεια της εκπαιδευτικής περιόδου (training period) υπάρχουν κινούμενα (foreground) αντικείμενα.
- **Παραδοχή 5:** Ο διαχωρισμός των pixels σε background και foreground γίνεται με την εφαρμογή ορίου, μεταξύ αυτού που έχει οριστεί ως background εικόνα και της τωρινής εικόνας. Η επιλογή του χαρακτηριστικού στο οποίο θα εφαρμοστεί το όριο αποτελεί κλειδί για την αποτελεσματικότητα της μεθόδου.

Πρακτικά, οι παραπάνω παραδοχές δεν ικανοποιούνται πλήρως από τις παρούσες μεθόδους για background subtraction, καθώς στην πραγματικότητα τα σενάρια σε ένα

βίντεο είναι πολλά. Πολλές τεχνικές έχουν αναπτυχθεί και έχουν προταθεί τα τελευταία χρόνια σχετικά με το background subtraction σε ένα βίντεο, καμία όμως δεν έχει δώσει μια καθολική λύση για την αντιμετώπιση όλων των πιθανών προκλήσεων σε ένα βίντεο [1].

1.4 Δεδομένα προς Επεξεργασία

Για τον έλεγχο της αποτελεσματικότητας των κλασσικών μεθόδων για background subtraction αλλά και της προτεινόμενης μεθόδου, χρησιμοποιήθηκαν διαφορετικά σύνολα δεδομένων (datasets) από διαφορετικά περιβάλλοντα. Ένα σύνολο δεδομένων αποτελείται από ακολουθίες στιγμιότυπων (sequence of frames). Πιο απλά, οι ακολουθίες αυτές είναι εικόνες που προέρχονται από τη δειγματοληψία ενός βίντεο και είναι εύκολα επεξεργάσιμες από κάποιον υπολογιστή. Ο ρυθμός δειγματοληψίας που χρησιμοποιείται για την επεξεργασία είναι συνήθως όσο και ο ρυθμός προβολής των στιγμιότυπων ανά δευτερόλεπτο (frames rate).

Κάθε βίντεο έχει κάποιες βασικές παραμέτρους, όπως:

- Τη χρονική διάρκεια σε δευτερόλεπτα.
- Την κωδικοποίηση του αρχείου βίντεο (video file format).
- Το ρυθμό προβολής στιγμιότυπων ανά δευτερόλεπτο (frames rate).
- Την κωδικοποίηση που απαιτείται για την ψηφιακή αναπαράσταση του κάθε pixel (video format), όπως επίσης και τον αριθμό των δυαδικών ψηφίων (bits per pixel).
- Τις διαστάσεις ύψος και πλάτος του βίντεο και κατ' επέκταση του κάθε frame.

1.5 Προγενέστερη Έρευνα

Πολλές μέθοδοι έχουν διατυπωθεί τα τελευταία χρόνια και αφορούν την μοντελοποίηση του background. Αρχικά, οι πιο απλές τεχνικές αναφέρονταν στη μοντελοποίηση του background χρησιμοποιώντας τον μέσο όρο (Mean) ή τον αριθμητικό μέσο (Median) των προηγούμενων N χρονικά frames.

Στην συνέχεια διατυπώθηκαν στατιστικές μέθοδοι που προσπαθούσαν να μοντελοποιήσουν την συνάρτηση Πυκνότητας Πιθανότητας (PDF) της έντασης ενός pixel στο χρόνο με χρήση μονής Gaussian συνάρτησης, γενικευμένων Gaussian συναρτήσεων (Generalized normal distribution - GGD), μίξη Gaussian συναρτήσεων (Mixture of Gaussian - MOG) καθώς και εκτίμηση της PDF ενός pixel με χρήση Πυρήνα (Kernel Density Estimation - KDE). Επίσης, έχουν προταθεί αλγόριθμοι όπου γίνεται χρήση Μηχανών Διανυσμάτων Υποστήριξης (Support Vector Machines) αλλά και εκμάθησης υποχώρου (Subspace learning).

Άλλες τεχνικές υποθέτουν ότι κάθε pixel του frame μπορεί να αναπαρασταθεί από συστοιχίες (clusters). Εισερχόμενα pixels συγκρίνονται με τις συστοιχίες του κάθε pixel και ανάλογα την κατηγοριοποίηση της συστοιχίας σε background ή foreground γίνεται και η κατηγοριοποίηση του pixel. Τέτοιες τεχνικές χρησιμοποιούν τον K-mean αλγόριθμο, καθώς και την δημιουργία – συντήρηση ενός Codebook. Η χρήση

νευρωνικών δικτύων (Neural Network) έχει επεκταθεί και στον τομέα της εικονικής αναγνώρισης foreground αντικειμένων. Το δίκτυο μέσα από μια διαδικασία εκμάθησης, κατηγοριοποιεί τα pixels σε background ή foreground αντίστοιχα.

Η χρήση επίσης μετασχηματισμών και συγκεκριμένα του Διακριτού Μετασχηματισμού Κυματιδίων (Discrete Wavelet Transform - DWT), καθώς και η χρήση φίλτρων όπως Wiener, Kalman, Tchebychev έχουν χρησιμοποιηθεί για την μοντελοποίηση του background.

Τέλος, τα τελευταία χρόνια έχουν αναπτυχθεί πιο εξελιγμένοι αλγόριθμοι που αφορούν την ανακατασκευή του background σε σημεία που το foreground αντικείμενο καλύπτει πλήρως το background. Τέτοιοι αλγόριθμοι απαιτούν τον υπολογισμό της οπτικής ροής (optical flow), της υφής (texture) και άλλων χαρακτηριστικών της εικόνας.

1.6 Προτεινόμενη μέθοδος

Στη συνέχεια της εργασίας, αφού διατυπωθούν οι κλασσικές τεχνικές μοντελοποίησης του background κι ελεγχθούν ως προς την απόδοση τους, παρουσιάζεται η πρότασή μας για μια νέα συνδυαστική μέθοδο η οποία αντιμετωπίζει προβλήματα στα οποία οι κλασσικές τεχνικές αποτυγχάνουν να αντιμετωπίσουν επαρκώς.

Κάποια από τα βασικά προβλήματα στα οποία καλείται να ανταποκριθεί η προτεινόμενη μέθοδος είναι τα εξής:

- Foreground αντικείμενα τα οποία είναι παρόντα κατά τη διάρκεια αρχικοποίησης του background.
- Αντικείμενα του foreground που κινούνται σχετικά αργά στο σκηνικό καταγραφής.
- Νέα αντικείμενα που εισέρχονται στο πλάνο, καθώς και υπάρχοντα αντικείμενα που εξέρχονται από αυτό.

Η προτεινόμενη μέθοδος σε σχέση με τις παραδοσιακές τεχνικές αντιμετωπίζει σε ικανοποιητικό βαθμό τους παραπάνω περιορισμούς, πράγμα που κάνει την προαναφερθείσα μέθοδο κατάλληλη για εφαρμογές όπου απαιτείται η αφαίρεση της σιλουέτας ενός ατόμου από ένα βίντεο. Μια τέτοια πιθανή χρήση της προτεινόμενης τεχνικής μπορεί να εφαρμοστεί στο πρόβλημα που περιγράφεται στην Παράγραφο 1.1

Στην προτεινόμενη μέθοδο γίνεται χρήση του χρωματικού μοντέλου $L^*a^*b^*$ και δημιουργούνται δυο εικόνες. Η μια εικόνα αφορά στην ανίχνευση περιοχών κίνησης χρησιμοποιώντας ένα μοντέλο μέσης τιμής για την background εικόνα, ενισχυμένο με μορφολογικούς αλγορίθμους και μορφολογικά φίλτρα. Η δεύτερη εικόνα αφορά στην τμηματοποίηση των frames σε ομοιογενείς περιοχές ανάλογα με το χρώμα τους. Η αρχική τμηματοποίηση της εικόνας πραγματοποιείται με τον αλγόριθμό K-means και αποφέρει εν τέλει τη δημιουργία ενός codebook με τα βασικά χρώματα της εικόνας. Το codebook είναι υπεύθυνο για την αντιστοίχιση των βασικών χρωμάτων στα pixels και για το λόγο αυτό είναι αναγκαία η ανανέωση του σε κάθε frame έτσι ώστε να προλαμβάνει τυχόν αλλαγές στο σκηνικό καταγραφής.

Ο συνδυασμός των δυο εικόνων πραγματοποιείται με μια αρχική ανακατασκευή του background, όπου οι εν δυνάμει περιοχές των foreground αντικειμένων αντικαθίστανται με το επικρατέστερο χρώμα (από το codebook) των γειτονικών τους περιοχών, οι οποίες ανήκουν στο background. Οι εν δυνάμει περιοχές προσδιορίζονται από τον βαθμό επικάλυψης των περιοχών κίνησης και της τμηματοποιημένης εικόνας που προκύπτει μετά την αντιστοίχιση των τιμών του codebook στα pixels. Στη συνέχεια, με ένα σχήμα επιλεκτικής ανανέωσης περιοχών αποφασίζεται η ανανέωση ή όχι της δεύτερης background εικόνας, η οποία προέρχεται από την τμηματοποίηση του frame. Η τελική εξαγωγή των foreground αντικειμένων προκύπτει με την αφαίρεση της τμηματοποιημένης εικόνας από τη δεύτερη background εικόνα.

Τέλος, για την περίπτωση που το foreground αντικείμενο μετακινείται αργά, έχει επιληφθεί ένας ασαφής αλγόριθμος ανανέωσης της εικόνας με τις περιοχές κίνησης (πρώτο background μοντέλο), στον οποίο η εικόνα με τις περιοχές κίνησης χωρίζεται σε blocks και συγκρίνεται με το προηγούμενο χρονικά frame και τα αντίστοιχα blocks. Το παραπάνω πλαίσιο τεχνικών αφαίρεσης του background έχει εμφανώς καλύτερα αποτελέσματα για το βίντεο με τις ασκήσεις γυμναστικής (Εικόνα 1.1) αλλά ταυτόχρονα μπορεί να αντιμετωπίσει και άλλα περιβάλλοντα όπως βίντεο απο κυκλώματα παρακολούθησης κυκλοφορίας.



Εικόνα 1.1 Στιγμιότυπο από το βίντεο με ασκήσεις yoga (Αριστερά). Το αποτέλεσμα της προτεινόμενης μεθόδου (Δεξιά).

2. Κεφάλαιο – Παραδοσιακές Τεχνικές Εξαγωγής του Foreground

Σε αυτό το κεφάλαιο αναλύονται οι βασικές τεχνικές που εφαρμόζονται για Background Subtraction μέσα από δημοσιεύσεις και παραδείγματα, καθώς και προβλήματα που προκύπτουν μετά την εφαρμογή αυτών σε δεδομένα. Εάν και είναι αδύνατον να καλυφθούν όλες οι κατηγορίες, δίνεται βάση σε βασικές κατηγορίες μοντελοποίησης του Background.

2.1 Βασικό Σχήμα Εξαγωγής του Foreground

Το βασικό σχήμα υλοποίησης που έχει προταθεί για τη μοντελοποίηση του Background προκύπτει από την απλή μαθηματική πράξη της αφαίρεσης. Η κοινή αγγλική ορολογία για την περιγραφή τέτοιων μεθόδων είναι: Frame Difference. Όπως περιγράφεται και στην Παράγραφο 1.3, εάν πάρουμε την εικόνα που περιέχει μόνο το background και την αφαιρέσουμε από την τωρινή εικόνα τότε μπορούμε να εξάγουμε τα κινούμενα αντικείμενα. Η αφαίρεση γίνεται σε επίπεδο pixels και η εικόνα μπορεί να είναι είτε ασπρόμαυρη (grayscale) είτε έγχρωμη. Τα pixels στα οποία η διαφορά της αφαίρεσης είναι πάνω από ένα όριο θεωρούνται ως foreground.

$$Pixel(x, y) = \begin{cases} foreground & |I_t(x, y) - B(x, y)| > Th, \\ background & otherwise \end{cases} \quad 2.1$$

Όπου (x, y) η θέση του pixel στην εικόνα, B η εικόνα του background, I_t το frame την χρονική στιγμή t και Th το όριο.

Σαν εικόνα για το B μπορεί να χρησιμοποιηθεί κάποιο προγενέστερο χρονικά frame [2]. Προτιμότερα, το B προκύπτει με τον υπολογισμό του αριθμητικού μέσου (mean) των προηγούμενων frames [3] [4], όπως επίσης μπορεί να χρησιμοποιηθεί και ο διάμεσος (median) των προηγούμενων frames [5]. Για την αρχικοποίηση των τιμών παραπάνω και κατ' επέκταση του B πρέπει να υπάρχει ένα στάδιο αρχικοποίησης (background initialization) όπου ένας αριθμός από frames θα λαμβάνεται υπόψιν στο υπολογισμό.

Στο ίδιο πλαίσιο είναι και η εξής μοντελοποίηση background [6], όπου το σχήμα μοντελοποίησης του background φαίνεται παρακάτω.

$$Pixel(x, y) = \begin{cases} foreground & \begin{cases} I_t(x, y) - m(x, y) < kd(x, y) \\ I_t(x, y) - n(x, y) < kd(x, y) \end{cases} \\ background & otherwise \end{cases} \quad 2.2$$

Όπου m, n μέγιστη και ελάχιστη τιμή αντίστοιχα του pixel κατά της διαδικασία της αρχικοποίησης, d η μέγιστη διαφορά μεταξύ δύο διαδοχικών frames κατά της διαδικασία της αρχικοποίησης και k ένας παράγοντας με τιμή 3~5 ο οποίος προκύπτει από την Gaussian κατανομή.

Το πρόβλημα στα παραπάνω σχήματα υλοποίησης είναι ότι δεν υπάρχει ανάδραση και το B δεν είναι ικανό να προσαρμοστεί σε τυχόν αλλαγές που μπορεί να προκύψουν, όπως για παράδειγμα σε αλλαγές στις συνθήκες φωτισμού ή αντικείμενα του background που μετακινούνται. Πρέπει λοιπόν να γίνει η εφαρμογή ενός φίλτρου άπειρης κρουστικής απόκρισης (IIR) το οποίο θα έχει την εξής μορφή:

$$B_t(x, y) = (1 - \alpha)B_t(x, y) + \alpha I_t(x, y) \quad 2.3$$

Όπου α παράγοντας εκμάθησης (learning rate), B_t και I_t η εικόνα του background και η τωρινή εικόνα την χρονική στιγμή t , αντίστοιχα.

Το βασικό μειονέκτημα του παραπάνω σχήματος υλοποίησης, είναι ότι η τιμή των pixels που έχουν χαρακτηριστεί σαν foreground λαμβάνεται υπόψιν στον υπολογισμό του νέου background και ως εκ τούτου μολύνει την background εικόνα. Για την αντιμετώπιση των παραπάνω χρησιμοποιείται διαφορετικός παράγοντας εκμάθησης ή και διαφορετικό όριο ανάλογα τα pixels [3]. Στην επιστημονική ορολογία κάτι τέτοιο περιγράφεται ως Επιλεκτικότητα (Selectivity).

$$\begin{aligned} B_{t+1}(x, y) &= (1 - \alpha)B_t(x, y) + \alpha I_t(x, y) \\ &\quad \text{if } (x, y) \text{ is background} \\ B_{t+1}(x, y) &= (1 - \beta)B_t(x, y) + \beta I_t(x, y) \\ &\quad \text{if } (x, y) \text{ is foreground} \end{aligned} \quad 2.4$$

Ο παράγοντας εκμάθησης β πρέπει να είναι πολύ μικρότερος του α , συνήθως 0, καθώς οι αλλαγές στα pixels του background πρέπει να υιοθετούνται πιο γρήγορα από αυτές των pixels του foreground. Ο κύριος λόγος για αυτό είναι ότι όταν ένα pixel χαρακτηριστεί ως foreground, μπορεί να χαρακτηριστεί πάλι ως background μόνο όταν η τιμή του γίνει ίδια με αυτή που ήταν πριν γίνει foreground.

Ο παράγοντας εκμάθησης εκφράζει τον ρυθμό προσαρμογής στις αλλαγές του σκηνικού του βίντεο και παίρνει τιμές από 0 έως 1. Ο παράγοντας εκμάθησης μπορεί να είναι σταθερός για όλα τα pixels και η τιμή του να προκύπτει είτε με προσεκτική προσέγγιση [7], είτε μέσα από μια αυτόματη διαδικασία ενός αλγόριθμου βελτιστοποίησης [8]. Επίσης, ο παράγοντας εκμάθησης μπορεί να αλλάζει στο χρόνο και η τιμή του να προκύπτει από κάποια στατιστική μέθοδο όπου εξαρτάται από την δραστηριότητα των pixels [9] ή από κάποια ασαφή λογική (fuzzy logic) [10].

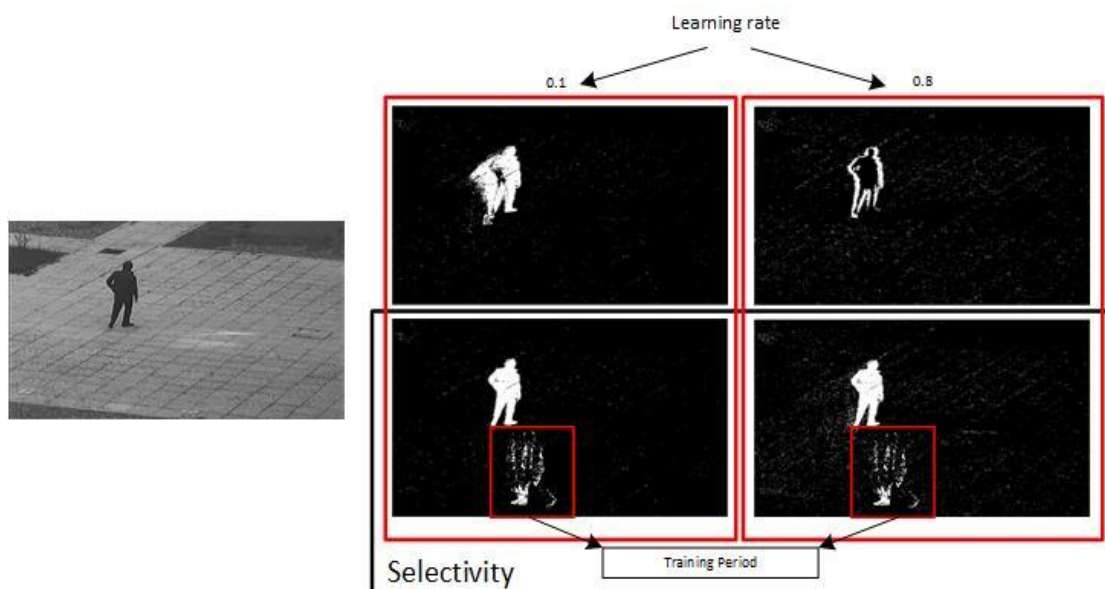
Τα αποτελέσματα της Εξίσωσης 2.1, με background εικόνα το μέσο όρο των N τελευταίων frames, φαίνονται παρακάτω. Η Εικόνα 2.1 έχει σαν κύριο μειονέκτημα ότι το κινούμενο αντικείμενο βρίσκεται κατά την διαδικασία της αρχικοποίησης μέσα στο πλάνο καταγραφής, οπότε κατά την μετακίνηση από την αρχική του θέση αφήνει πίσω του ένα λανθασμένο αποτύπωμα που ο αλγόριθμος υπολογίζει σαν foreground. Με την εφαρμογή του IIR φίλτρου η κατάσταση βελτιώνεται ως προς το ότι η παλιά θέση του αντικειμένου ενσωματώνεται στο background αλλά δημιουργεί το πρόβλημα ότι η σιλουέτα του ατόμου δεν εξάγεται σωστά αλλά μόνο η ακμή αυτής (Εικόνα 2.2). Με την χρήση του selectivity δημιουργείται πάλι το αρχικό πρόβλημα που φαίνεται στην Εικόνα 2.1 αλλά η μέθοδος παύει να είναι ευαίσθητη στο παράγοντα εκμάθησης (Εικόνα 2.3).



Εικόνα 2.1 Ασπρόμαυρο βίντεο με yoga, frame (Αριστερά). Χωρίς την εφαρμογή IIR φίλτρου (Δεξιά).



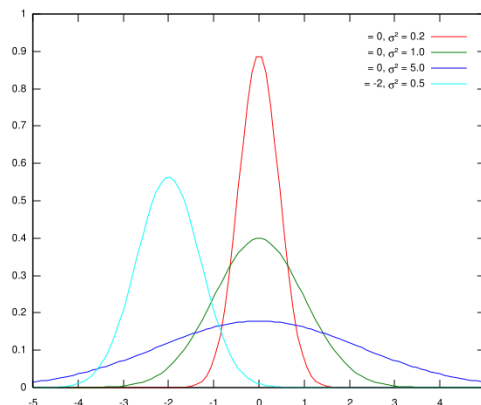
Εικόνα 2.2 Ασπρόμαυρο βίντεο με yoga, frame (Αριστερά). Με την εφαρμογή IIR φίλτρου (Δεξιά).



Εικόνα 2.3 Ασπρόμαυρο βίντεο με το περπάτημα ενός ατόμου, frame (Αριστερά) [134]. Το αποτέλεσμα (με ίδιο όριο και ίδια frames κατά την διαδικασία της αρχικοποίησης) με ή χωρίς το selectivity για διαφορετικό παράγοντα εκμάθησης.

2.2 Στατιστική Μοντελοποίηση του Background και Εξαγωγή του Foreground

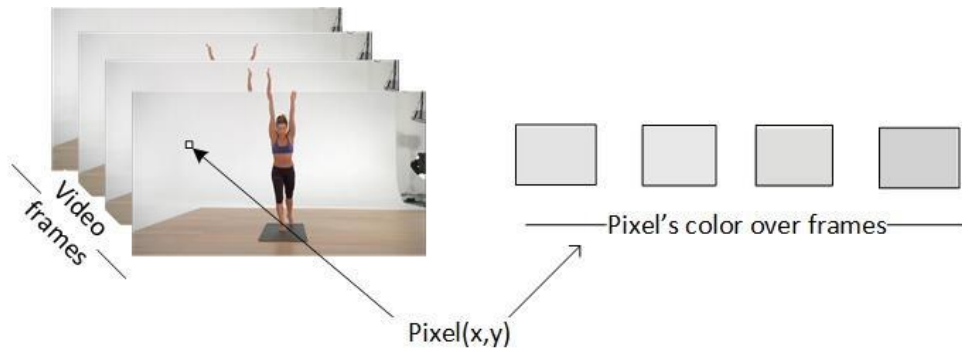
Το μοντέλο που χρησιμοποιείται για τη μοντελοποίηση του background πρέπει να είναι ικανό να αντιμετωπίζει τόσο Μονότροπα (Unimodal) όσο και Πολύτροπα (Multimodal) background. Unimodal κατανομές στην Επιστήμη της Στατιστικής είναι κατανομές οι οποίες έχουν μόνο μία κύρια κορυφή στο διάγραμμα της Συνάρτησης Πυκνότητας Πιθανότητας (PDF), ενώ Multimodal είναι κατανομές οι οποίες έχουν περισσότερες από μία κορυφές [11].



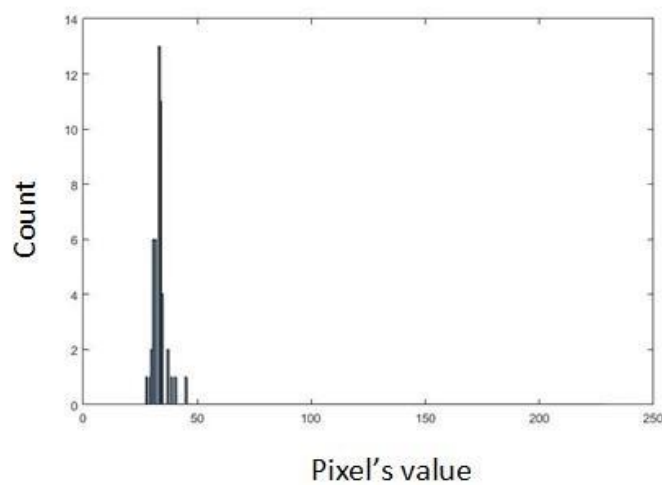
Εικόνα 2.4 Διάγραμμα PDF. Κανονική ή Gaussian Κατανομή παράδειγμα Unimodal κατανομής.

Η μοντελοποίηση του background στις περισσότερες μεθόδους επιτυγχάνεται με τη μοντελοποίηση της συμπεριφοράς του κάθε pixel της εικόνας. Κάθε pixel προσδιορίζεται από τη θέση του στην εικόνα (x,y) και από την τιμή του. Για παράδειγμα, η τιμή που παίρνει το pixel για μια ασπρόμαυρη εικόνα (grayscale) με 8 δυαδικά ψηφία για την ψηφιακή αναπαράσταση του (bits per pixel) είναι από 0 έως 255. Σε ένα βίντεο όπου υπάρχει και η συνιστώσα του χρόνου, ένα pixel σε κάθε frame παίρνει και διαφορετικές τιμές. Σε περιπτώσεις που η κωδικοποίηση της εικόνα σε ένα βίντεο δεν είναι ασπρόμαυρη αλλά RGB πρότυπο χρώματος, ένα pixel παίρνει τρεις τιμές μία για κάθε συνιστώσα χρώματος (κόκκινο, πράσινο, μπλε). Ο συνδυασμός των τριών συνιστωσών δίνει το χρώμα του pixel.

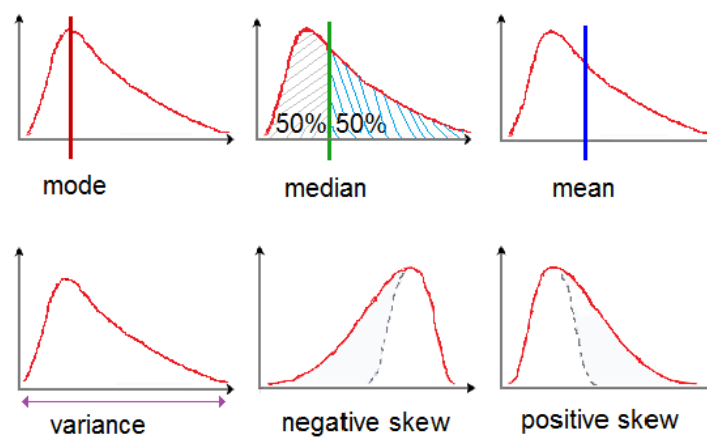
Το ιστόγραμμα ενός pixel (για έγχρωμη ή ασπρόμαυρη εικόνα) στο χρόνο μπορεί να μας δώσει σημαντική πληροφορία [12]. Τέτοιες πληροφορίες αφορούν κυρίως την Μέση Τιμή (mean), την Διακύμανση (Variance), την Λοξότητα (Skewness) και την Κύρτωση (Kurtosis). Αυτές οι τιμές περιγράφουν το σχήμα και την μορφή του συνόλου των δεδομένων και είναι γνωστά στην Στατιστική ως Moments [13] (Εικόνα 2.7). Στην περίπτωση μας, το σύνολο των δεδομένων προκύπτει από την τιμή του pixel στο χρόνο. Ένα pixel ανάλογα με την θέση του στην εικόνα μπορεί να καταγράψει είτε το background (Εικόνα 2.6), είτε κάποιο κινούμενο αντικείμενο (foreground), είτε και τα δύο. Το ιστόγραμμα ενός pixel για μια ακολουθία από n frames, μετρά τον αριθμό των διακριτών τιμών-εντάσεων του pixel στο χρόνο.



Εικόνα 2.5 Στην εικόνα φαίνεται η θέση του pixel καθώς και τα χρώματα που αναπαριστά κατά την διάρκεια των n frames.



Εικόνα 2.6 Παράδειγμα ιστογράμματος στο χρόνο. Αριθμός frames $n=50$. Κάθετος άξονας, πλήθος τιμών για κάθε χρώμα στο χρόνο. Οριζόντιος άξονας, διακριτά χρώματα (0~255).



Εικόνα 2.7 Γεωμετρικά χαρακτηριστικά κατανομών.

2.2.1 Μονή Gaussian Συνάρτηση (Single Gaussian - SG)

Η μοντελοποίηση της δραστηριότητας του κάθε pixel με Gaussian κατανομή υλοποιήθηκε πρώτα από τους Wren *et al.* [14]. Η ιδέα της συγκεκριμένης μεθόδου είναι η προσαρμογή μιας Gaussian PDF στις τελευταίες n τιμές του pixel. Η κοινή αγγλική ορολογία για την μέθοδο που περιγράφεται παρακάτω είναι: Running Gaussian Average. Για την αποφυγή του υπολογισμού της μέσης τιμής και της διασποράς κάθε φορά που έρχεται ένα καινούργιο frame, προτείνεται η ενημέρωση των τιμών αυτών συμφωνά με τα παρακάτω:

$$\begin{aligned}\mu_{t+1} &= (1 - \alpha)\mu_t + \alpha I_{t+1} \\ \sigma_{t+1}^2 &= (1 - \alpha)\sigma_t^2 + \alpha (I_{t+1} - \mu_{t+1})^2\end{aligned}\quad 2.5$$

Οπού I_{t+1} το τωρινό frame, μ_t η μέση τιμή των προηγούμενων frames, σ_t^2 η τιμή της διασποράς για τα προηγούμενα frames και α ο παράγοντας εκμάθησης. Η συνάρτηση παραπάνω εφαρμόζεται σε κάθε pixel του frame όποτε οι μεταβλητές μ_t , σ_t^2 είναι στην ουσία πίνακες με διαστάσεις όσο και το frame.

Ο διαχωρισμός των αντικειμένων σε background γίνεται πάλι με την εφαρμογή ενός ορίου:

$$pixel(x, y) = \begin{cases} background & \text{if } |I_{t+1} - \mu_{t+1}| < Th, \\ foreground & \text{otherwise} \end{cases}\quad 2.6$$

Το όριο Th μπορεί να επιλεγεί έτσι ώστε να έχει την μορφή σταθερής τιμής συναρτήσει της διασποράς, δηλαδή $Th = k * \sigma_t$, όπου k μια σταθερά συνήθως γύρω στο 3. Το k έχει φυσική σημασία εάν σκεφτούμε την Gaussian κατανομή η οποία χαρακτηρίζεται από μια μέση τιμή και μια διασπορά. Για την Gaussian κατανομή ισχύει ότι το 68% των τιμών του δείγματος βρίσκεται σε απόσταση $\pm\sigma$ από την μέση τιμή, ενώ για απόσταση $\pm 3\sigma$ από την μέση τιμή βρίσκεται το 99.7% των τιμών του δείγματος. Συνήθως το k παίρνει τιμές μεγαλύτερες του 3 όταν το περιβάλλον είναι εξωτερικό με αρκετό θόρυβο.

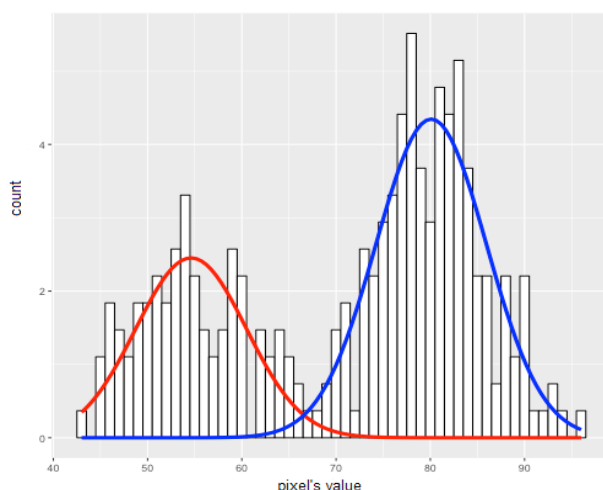
Βελτιστοποιήσεις

Στη βιβλιογραφία υπάρχουν αρκετές βελτιώσεις του παραπάνω σχήματος, μια από αυτές είναι η εφαρμογή του όρου της Επιλεκτικότητας (Selectivity) που αναφέρεται στην Εξίσωση 2.4. Η Επιλεκτικότητα αφορά τον παράγοντα εκμάθησης (α) και ως εκ τούτου όλες οι δημοσιεύσεις που αφορούν τον παράγοντα εκμάθησης μπορούν να εφαρμοστούν και στην σχέση παραπάνω. Στη δημοσίευση Tang *et al* [15] προτείνεται ένας απλός συνδυασμός της μεθόδου Running Gaussian Average και του Frame Difference. Άλλες βελτιώσεις αφορούν την εφαρμογή της μεθόδου σε άλλα χρωματικά μοντέλα όπως το RGB και το HSV. Στην αναφορά Cerman *et al* [16] συγκρίνεται η μέθοδος του Running Gaussian Average σε δύο χρωματικά μοντέλα: RGB και Grayscale, όπου το συμπέρασμα ήταν ότι τα δύο χρωματικά μοντέλα έχουν πολύ μικρές διαφορές στην απόδοσή τους. Στην δημοσίευση Singh *et al.* [17] δοκιμάζονται τρεις βασικές μέθοδοι του Background Subtraction, όπως Frame Difference και Running

Gaussian Average και χρησιμοποιείται το HSV χρωματικό μοντέλο για την ψηφιακή αναπαράσταση του pixel. Το συμπέρασμα που προκύπτει είναι ότι το χρωματικό μοντέλο HSV είναι καταλληλότερο για την αντιμετώπιση των βαθμιαίων αλλαγών στις συνθήκες φωτισμού και στην αντιμετώπιση των σκιών [18], καθώς σαν μοντέλο διαχωρίζει την πληροφορία του χρώματός από την πληροφορία της έντασης.

2.2.2 Μίξη Gaussian Συναρτήσεων (Mixture Of Gaussians - MOG)

Η χρήση μονής Gaussian συνάρτησης για τη μοντελοποίηση του background είναι ικανοποιητική μόνο όταν το pixel καταγράφει μια μόνο επιφάνεια κάτω από ίδιες συνθήκες φωτισμού. Επίσης, η χρήση μονής Gaussian συνάρτησης χρησιμοποιείται για τη μοντελοποίηση τυχόν θορύβου του αισθητήρα της κάμερας, γνωστού και ως Gaussian Noise [19]. Τις περισσότερες φορές το pixel καταγράφει διαφορετικές επιφάνειες και οι συνθήκες φωτισμού σε ένα βίντεο μπορούν να αλλάξουν. Το ιστόγραμμα στο χρόνο ενός τέτοιου pixel θα έχει πάνω από μια κορυφή στον κάθετο άξονα.



Εικόνα 2.8 Μίξη Gaussian συναρτήσεων για την μοντελοποίηση του pixel.

Η λύση στο πρόβλημα είναι η χρήση k Gaussian συναρτήσεων. Η αρχική ιδέα πρωτοεμφανίζεται σε μια αναφορά των Friedman και Russel [20] για κάμερες παρακολούθησης κυκλοφορίας. Σε αυτήν, κάθε pixel του background μοντελοποιείται χρησιμοποιώντας τρεις Gaussian συναρτήσεις, μία για το δρόμο, μία για τα αμάξια και άλλη μία για τις σκιές. Τα pixels λοιπόν κατηγοριοποιούνται στις τρεις αυτές Gaussian συναρτήσεις ανάλογα με την δικιά τους κατανομή. Η διατήρηση του μοντέλου γίνεται χρησιμοποιώντας έναν Expectation Maximization (EM) αλγόριθμο έτσι ώστε να χρησιμοποιείται για real time εφαρμογές. Το βασικό πρόβλημα είναι ότι το μοντέλο που χρησιμοποιείται δεν είναι προσαρμόσιμο σε αλλαγές του σκηνικού.

Γενικεύοντας την ιδέα παραπάνω, οι Stauffer και Grimson [21] πρότειναν την μοντελοποίηση της πρόσφατης ιστορίας του χρώματος του κάθε pixel $\{X_1, \dots, X_t\}$ με μίξη από K Gaussian συναρτήσεις.

Βασική Ιδέα

Αρχικά, κάθε pixel χαρακτηρίζεται από την ένταση του στο RGB χρωματικό μοντέλο. Έπειτα, η πιθανότητα να πάρει το συγκεκριμένο pixel τη συγκεκριμένη ένταση δίνεται από την κάτωθι πολύ-διανυσματική συνάρτηση:

$$P(X_t) = \sum_{i=1}^K \omega_{i,t} \cdot \eta(X_t, \mu_{i,t}, \Sigma_{i,t}) \quad 2.7$$

Όπου K ο αριθμός των Gaussian συναρτήσεων, $\omega_{i,t}$ είναι τα βάρη της i^{th} Gaussian συνάρτησης την χρονική στιγμή t με μέση τιμή $\mu_{i,t}$ και τυπική απόκλιση $\Sigma_{i,t}$.

η είναι η PDF της Gaussian συνάρτησης:

$$\eta(X_t, m, \Sigma) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} e^{-\frac{1}{2}(X_t - \mu)^T \Sigma^{-1} (X_t - \mu)}$$

Για υπολογιστικούς λόγους, οι Stauffer και Grimson [21] υποθέτουν ότι οι τρεις συνιστώσες του RGB χρωματικού μοντέλου είναι ανεξάρτητες και έχουν την ίδια διακύμανση (variance). Οπότε, ο πίνακας συνδιακύμανσης είναι της μορφής:

$$\Sigma_{i,t} = \sigma_{i,t}^2 I$$

Όπου I ο μοναδιαίος πίνακας.

Οι παραπάνω συναρτήσεις είναι αυτές που χρησιμοποιούνται για τη μοντελοποίηση κάθε pixel. Οι παράμετροι του Mixture of Gaussians (MOG) μοντέλου που χρειάζονται αρχικοποίηση είναι: ο αριθμός των K Gaussian συναρτήσεων, τα βάρη $\omega_{i,t}$ που σχετίζονται με την i^{th} Gaussian συνάρτηση την χρονική στιγμή t , η μέση τιμή $\mu_{i,t}$ και ο πίνακας συνδιακύμανσης $\Sigma_{i,t}$.

Η παράμετρος K εκφράζει την πολυτροπικότητα (Multimodality) του background. Στην δημοσίευση των Stauffer και Grimson η τιμή του K παίρνει τιμές από 3 έως 5. Η αρχικοποίηση των βαρών $\omega_{i,t}$, της μέσης τιμής $\mu_{i,t}$ όπως και του πίνακα συνδιακύμανσης $\Sigma_{i,t}$ γίνεται χρησιμοποιώντας έναν EM αλγόριθμο και συγκεκριμένα τον K-mean αλγόριθμο.

Ανίχνευση Κινούμενων Αντικειμένων

Αφού γίνει η αρχικοποίηση των παραμέτρων, το επόμενο στάδιο είναι να γίνει μια πρώτη ανίχνευση των κινούμενων αντικειμένων και μετά να γίνει ανανέωση των παραμέτρων αυτών. Αρχικά οι Stauffer και Grimson χρησιμοποιούν τον λόγο $r_j = \omega_j / \sigma_j$ έτσι ώστε να ταξινομήσουν βάση αυτού τις K Gaussian κατανομές.

Αυτή η ταξινόμηση υποθέτει ότι ένα pixel του background έχει μεγάλο βάρος αλλά μικρή διακύμανση εξαιτίας του ότι το background θεωρείται γενικά πιο σταθερό στην διακύμανση των τιμών από τα κινούμενα αντικείμενα. Οι πρώτες B Gaussian συναρτήσεις οι οποίες υπερβαίνουν ένα συγκεκριμένο όριο T θεωρούνται κατανομές του background:

$$B = \operatorname{argmin}_b \left(\sum_{i=1}^b \omega_{i,t} > T \right) \quad 2.8$$

Οι υπόλοιπες κατανομές θεωρούνται ότι ανήκουν στο foreground. Στη συνέχεια όταν έρθει ένα καινούργιο frame την χρονική στιγμή $t+1$, εφαρμόζεται ένα όριο για κάθε pixel. Το pixel ταιριάζει με μια Gaussian κατανομή εάν:

$$((X_{t+1} - \mu_{i,t})^2 \Sigma_{i,t}^{-1})^{1/2} < k \sigma_{i,t} \quad 2.9$$

Όπου k είναι μια σταθερά όπως έχει αναφερθεί παραπάνω με τιμή ίση 2.5. Μετά την εφαρμογή του ορίου δύο περιπτώσεις μπορούν να συμβούν:

- Περίπτωση 1: Το pixel όντως ταιριάζει με μια από τις K Gaussian κατανομές. Σε αυτή την περίπτωση, εάν η Gaussian κατανομή αναφέρεται για το background τότε το pixel χαρακτηρίζεται ως background αλλιώς το pixel χαρακτηρίζεται ως foreground.
- Περίπτωση 2: Το pixel δεν ταιριάζει με καμία από τις K Gaussian κατανομές. Σε αυτή τη περίπτωση το pixel χαρακτηρίζεται ως foreground.

Ανανέωση Παραμέτρων

Μετά την εφαρμογή του ορίου αυτού μια δυαδική εικόνα έχει σχηματιστεί. Το επόμενο βήμα είναι η ανανέωση των παραμέτρων ανάλογα με το εάν ένα pixel ταιριάζει ή όχι με κάποια από τις K Gaussian κατανομές

Περίπτωση 1: Το Pixel ταιριάζει με κάποια από τις K Gaussian κατανομές.

- Για τα pixels που έγινε η ταύτιση με κάποια Gaussian κατανομή που ανήκει στο background, οι παράμετροι ανανεώνονται ως εξής:

$$\begin{aligned} \omega_{i,t+1} &= (1 - a)\omega_{i,t} + a \\ \mu_{i,t+1} &= (1 - \rho)\mu_{i,t} + \rho X_{t+1} \\ \sigma_{i,t+1}^2 &= (1 - \rho)\sigma_{i,t}^2 + \rho(X_{t+1} - \mu_{i,t+1})^2 \end{aligned} \quad 2.10$$

Όπου a παράγοντας εκμάθησης, $\rho = a\eta(X_{t+1}, \mu_i, \Sigma_i)$

- Για τα pixels που δεν έγινε ταύτιση με κάποια από τις b Gaussian κατανομές, το μ , Σ παραμένουν όπως έχουν ενώ τα βάρη ανανεώνονται ως εξής:

$$\omega_{j,t+1} = (1 - a)\omega_{j,t}$$

Περίπτωση 2: Το Pixel δεν ταιριάζει με κάποια από τις K Gaussian κατανομές. Σε αυτή την περίπτωση η λιγότερο πιθανή Gaussian κατανομή K αντικαθίσταται από μια νέα με νέες παραμέτρους:

$$\begin{aligned}
\omega_{k,t+1} &= \text{Low Prior Weight} \\
\mu_{k,t+1} &= X_{t+1} \\
\sigma_{k,t+1}^2 &= \text{Large Initial Variance}
\end{aligned}
\tag{2.11}$$

Αφού γίνει η ανανέωση των τιμών η εξαγωγή των κινούμενων αντικειμένων μπορεί να γίνει πάλι και ούτω καθεξής. Μια πιο ολοκληρωμένη μελέτη πάνω στις παραμέτρους και στην απόδοση τιμών σε αυτές βρίσκεται εδώ [22] [23].

Βελτιστοποιήσεις

Δεκάδες δημοσιεύσεις έχουν προταθεί τα τελευταία χρόνια. Στη δημοσίευση Bouwmans *et al* [24] γίνεται αναφορά σε σχεδόν 100 δημοσιεύσεις που αφορούν βελτιστοποιήσεις πάνω στη μέθοδο MOG, την περίοδο 1999-2007. Στον παρακάτω πίνακα φαίνονται ενδεικτικά κάποιες από αυτές:

Βήματα Αλγορίθμου	Παράμετροι	Αναφορές
Αρχικοποίηση Background	K	Zivkovic [25], Cheng <i>et al.</i> [26], Shimada <i>et al.</i> [27], Tan <i>et al.</i> [28], Carminati <i>et al.</i> [29]
	μ, σ, ω	Morellas <i>et al.</i> [30], Lee [31], Zhang <i>et al.</i> [32], Amintoosi <i>et al.</i> [33], Lepisk [34]
Ανανέωση Background	μ, σ, ω	Zhang <i>et al.</i> [32], Han and Li [35], Wang and Suter [36], Lindstrom <i>et al.</i> [37], Stauffer and Grimson [38], Landabaso and Pardas [39], Park <i>et al.</i> [40] Mittal and Huttenlocher [41]
	Παράγοντας εκμάθησης α, ρ	Wang and Suter [36], Lindstrom <i>et al.</i> [37], Stauffer and Grimson [38], KaewTraKulPong and Bowden [42] [43], [44], Lee [45], Harville <i>et al.</i> [46], Porikli [47], Liu <i>et al.</i> [48], Pnevmatikakis <i>et al.</i> [49] [50], Power <i>et al.</i> [51], Leotta <i>et al.</i> [52]

Πίνακας 2-1 Ενδεικτικές δημοσιεύσεις πάνω στην μέθοδο MOG.

Εκτός από τις παραπάνω βελτιώσεις που αφορούν την εσωτερική (Intrinsic) βελτιστοποίηση του αλγορίθμου έχουν προταθεί και αλγόριθμοι για την μείωση του υπολογιστικού κόστους, όπως επίσης και αλγόριθμοι που δεν ασχολούνται με την δομή του αλγορίθμου αλλά με τα χαρακτηριστικά της εικόνα στα οποία εφαρμόζεται η μέθοδος. Οι Stauffer και Grimson [21] χρησιμοποίησαν τις εντάσεις στο RGB

χρωματικό μοντέλο. Σε άλλες δημοσιεύσεις έχει προταθεί η χρησιμοποίηση άλλων χρωματικών μοντέλων (όπως HSV, HIS, YUV) καθώς και χαρακτηριστικά ακμών (edge), υφών (texture), κίνησης (motion), φάσης (phase, HOG) μιας εικόνας για την εφαρμογή της μεθόδου MOG [62].

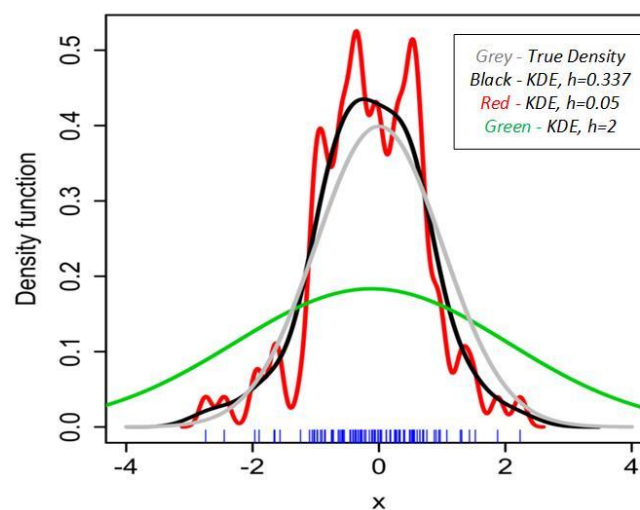
2.2.3 Συνάρτηση Πυκνότητας Πιθανότητας με Χρήση Πυρήνα (Kernel Density Estimation - KDE)

Στο προηγούμενο κεφάλαιο είδαμε πως μπορούμε με τη χρήση Gaussian συναρτήσεων να μοντελοποιήσουμε την PDF ενός pixel. Σε αυτήν την παράγραφο αναλύεται η χρήση Πυρήνα (Kernel) για την μοντελοποίηση της PDF, σε αντίθεση με την μέθοδο MOG που είναι παραμετρική μέθοδος η μέθοδος KDE είναι μη παραμετρική μέθοδος. Για το λόγο αυτό η χρήση της μεθόδου μπορεί να βοηθήσει στην μοντελοποίηση κατανομών για τις οποίες δεν μπορούν να γίνουν υποθέσεις για την μορφή της κατανομής. Μια συνάρτηση κατανομής με χρήση πυρήνα χαρακτηρίζεται από μια συνάρτηση Πυρήνα K και ένα εύρος ζώνης.

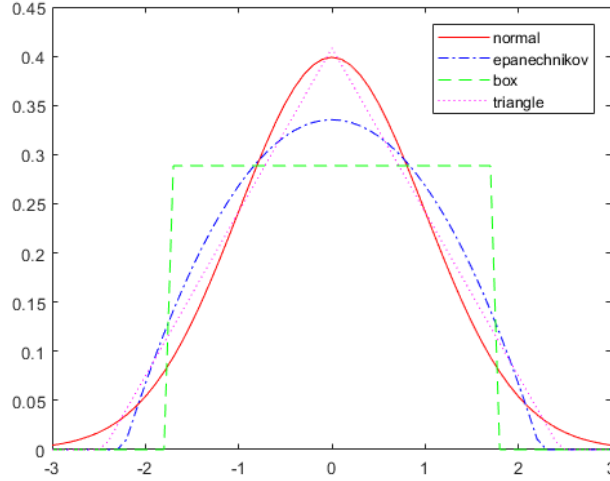
Η συνάρτηση πυκνότητας πιθανότητας με χρήση Πυρήνα είναι η εξής:

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right) \quad 2.12$$

Όπου x_1, x_2, \dots, x_n είναι τα δείγματα από μια άγνωστη κατανομή, K η συνάρτηση Πυρήνα και h το εύρος ζώνης (bandwidth). Η συνάρτηση Πυρήνα K είναι μια μη-αρνητική συνάρτηση η οποία ολοκληρώνεται στο ένα. Σαν συνάρτηση Πυρήνα συνήθως χρησιμοποιούνται: η Κανονική συνάρτηση, η τριγωνική συνάρτηση, η ομοιόμορφη συνάρτηση, η Συνάρτηση Epanechnikov καθώς και άλλες (Εικόνα 2.10). Η παράμετρος h αποτελεί κάτι σαν κλιμάκωση για την συνάρτηση Πυρήνα, η τιμή του επιλέγεται έτσι ώστε να είναι όσο πιο μικρή γίνεται για τα δεδομένα. Ο πιο σύνηθες υπολογισμός του h γίνεται μέσω του υπολογισμού του Mean Integrated Squared Error (MISE) [63].



Εικόνα 2.9 Το αποτέλεσμα της διαδικασίας KDE για διαφορετικά h .



Εικόνα 2.10 Διαγράμματα Συναρτήσεων Πυρήνα (Kernel).

Βασική Ιδέα

Στην δημοσίευση των Elgammal *et al.* [64] προτείνεται η μοντελοποίηση της PDF κάθε pixel χρησιμοποιώντας Πυρήνα K για τις τελευταίες N τιμές της έντασης του pixel $\{X_1, X_2, \dots, X_n\}$. Οι τιμές του pixel ανανεώνονται μέσα από ένα χρονικό παράθυρο W κρατώντας πάντα τις N τελευταίες. Η PDF του pixel υπολογίζεται από την σχέση:

$$P(X_t) = \frac{1}{n} \sum_{i=1}^N K(X_t - X_i) \quad 2.13$$

Όπου K είναι η συνάρτηση του Πυρήνα, η οποία για τη Κανονική ή Gaussian κατανομή $N(0, \Sigma)$ έχει την εξής μορφή:

$$P(X_t) = \frac{1}{N} \sum_{i=1}^N \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} e^{-1/2 * (X_t - X_i)^T \Sigma^{-1} (X_t - X_i)} \quad 2.14$$

Στη δημοσίευση των Elgammal *et al.* [64] υποθέτουν ότι τα διαφορετικά χρώματα κόκκινο, πράσινο, μπλε είναι ανεξάρτητα με διαφορετικά εύρη ζωνών (bandwidths) Πυρήνα, όπου η συνάρτηση του εύρους ζώνης του Πυρήνα είναι η εξής:

$$\Sigma = \begin{pmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{pmatrix}$$

Οπότε η PDF μπορεί να γράφει ως εξής:

$$P(X_t) = \frac{1}{N} \sum_{i=1}^N \prod_{j=1}^d \frac{1}{\sqrt{2\pi\sigma_j^2}} e^{-1/2 * (x_{t_j} - x_{i_j})^T / \sigma_j^2} \quad 2.15$$

Ανίχνευση Κινούμενων Αντικειμένων

Η ανίχνευση των κινούμενων αντικειμένων γίνεται χρησιμοποιώντας τη σχέση παραπάνω και ένα όριο. Εάν $P(X_t) < T$ τότε το pixel χαρακτηρίζεται ως foreground αλλιώς ως background. Μετά την εφαρμογή του ορίου μια δυαδική εικόνα που είναι το αποτέλεσμα της διαδικασίας έχει προκύψει. Το επόμενο στάδιο είναι η ανανέωση των παραμέτρων.

Ανανέωση Παραμέτρων

Για την περίπτωση αυτή οι Elgammal *et al* [64] χρησιμοποίησαν δύο μοντέλα για την μοντελοποίηση του background: ένα μικρής περιόδου και ένα μεγάλης περιόδου. Τα δύο αυτά μοντέλα πετυχαίνουν τα εξής:

Το μοντέλο μικρής περιόδου προσαρμόζεται πολύ γρήγορα στις αλλαγές του σκηνικού. Το μοντέλο αποτελείται από τις N τελευταίες τιμές μόνο των pixels που έχουν χαρακτηριστεί ως background, χρησιμοποιείται δηλαδή ο μηχανισμός του Selectivity που έχει αναφερθεί στις προηγούμενες παραγράφους.

Το μεγάλης περιόδου μοντέλο από την άλλη προσαρμόζεται στις αλλαγές πιο αργά. Το μοντέλο αποτελείται από τις N τελευταίες τιμές ενός χρονικού παραθύρου το οποίο είναι πιο μεγάλο από το μοντέλο μικρής περιόδου. Στην ανανέωση των τιμών δεν χρησιμοποιείται ο μηχανισμός του Selectivity.

Αφού γίνει η ένωση των δυο μοντέλων, τα pixel που έχουν χαρακτηριστεί λανθασμένα ως foreground στο πρώτο μοντέλο και συνάμα στο δεύτερο μοντέλο εξαφανίζονται. Τα pixel που λανθασμένα έχουν χαρακτηριστεί ως foreground και παραμένουν στο νέο μοντέλο είναι εκείνα τα οποία δεν υπάρχουν και στα δυο μοντέλα ταυτόχρονα. Τα pixel αυτά που έχουν μείνει και συνεχίζουν και υπάρχουν για μεγάλο χρονικό διάστημα αφομοιώνονται από το μεγάλης περιόδου μοντέλο και εξαφανίζονται τελικά. Εκτός από την παραπάνω περίπτωση, pixels τα οποία έχουν χαρακτηριστεί σωστά σαν foreground στο πρώτο μοντέλο και συνάμα σωστά σαν background στο δεύτερο μοντέλο (καθώς το δεύτερο μοντέλο είναι πιο αργό στις προσαρμογές), δημιουργούν ένα πρόβλημα στην μέθοδο. Η λύση είναι ότι όλα τα pixels που έχουν χαρακτηριστεί ως foreground στο πρώτο μοντέλο και είναι γειτονικά στα pixels που έχουν χαρακτηριστεί ως foreground στην ένωση των δυο μοντέλων συμπεριλαμβάνονται στο τελικό αποτέλεσμα.

Βελτιστοποιήσεις

Όπως και στο μοντέλο MOG έτσι και σε αυτό έχουν προταθεί πολλές βελτιστοποιήσεις, τόσο στην δομή του αλγορίθμου όσο και στα χαρακτηριστικά της εικόνας στα οποία εφαρμόζεται η μέθοδος.

Στο παρακάτω πίνακα αναφέρονται ενδεικτικά κάποιες από αυτές:

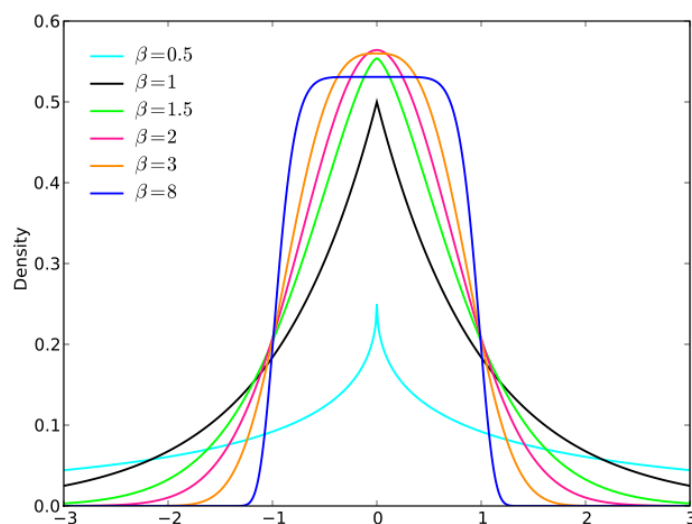
Βήματα Αλγορίθμου	Βελτιώσεις	Αναφορές
Μοντελοποίηση Background	Gaussian Πυρήνας	Tavakkoli <i>et al.</i> [65] [66]
	Rectangular Πυρήνας	Ianasi <i>et al.</i> [67], Tanaka <i>et al.</i> [68] [69], Zivkovic [70]

	Derivative Πυρήνας	Cvetkovic <i>et al.</i> [71]
	Negative Coefficient polynomial Πυρήνας	Witherspoon and Zhang [72]
	Cauchy Πυρήνας	Ramezani <i>et al.</i> [73]
Αρχικοποίηση Background	Μείωση στον αριθμό των δειγμάτων	Ianasi <i>et al.</i> [67], Mao and Shi [74] [75], Tang <i>et al.</i> [76]
Μηχανισμός Διατήρησης Background	Background Εικόνα	Ianasi <i>et al.</i> [67]
	Αναδρομική Ανανέωση του background	Ianasi <i>et al.</i> [67], Tanaka <i>et al.</i> [68] [69], Ramezani <i>et al.</i> [73], Tavakkoli <i>et al.</i> [77] [78]
	Αριθμός Δειγμάτων	Zivkovic [70]
	Μηχανισμός Selectivity	Cvetkovic <i>et al.</i> [71]
Ανίχνευση Κινούμενων Αντικειμένων	Μέτρο Ανομοιότητας	Ianasi <i>et al.</i> [67]
	Πιθανότητα	Zivkovic [70]
	Foreground Μοντέλο	Tavakkoli <i>et al.</i> [77] [78]
	Δυο Όρια	Cvetkovic <i>et al.</i> [71]

Πίνακας 2-2 Ενδεικτικές δημοσιεύσεις πάνω στην μέθοδο KDE.

2.2.4 Οικογένεια Γενικευμένων Gaussian Συναρτήσεων (Generalized Gaussian Family - GGF)

Η Γενικευμένη Gaussian Συνάρτηση (Generalized Gaussian Family–GGF) περιλαμβάνει μια οικογένεια παραμετρικών και συμμετρικών κατανομών. Περιλαμβάνει όλες τις Gaussian και Laplace κατανομές, όπως επίσης και τις συνεχής ομοιόμορφες κατανομές.



Εικόνα 2.11 Γενικευμένη Gaussian Συνάρτηση (GGF).

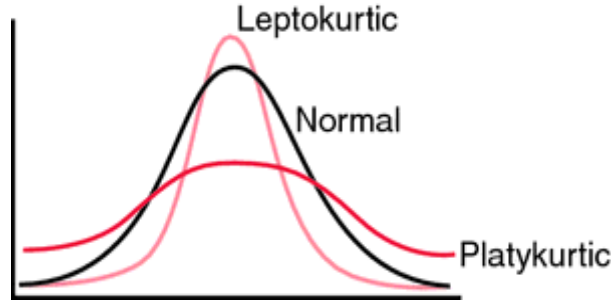
Η PDF μιας γενικευμένης Gaussian συνάρτησης είναι η εξής:

$$f(x; \beta) = \frac{\beta}{2\alpha\Gamma(1/\beta)} e^{-(|x-\mu|/\alpha)^\beta} \quad 2.16$$

Όπου $\Gamma(\cdot)$ είναι μια συνάρτηση Γάμμα, μ η μέση τιμή η οποία συμπίπτει και με τον διάμεσο (median) αλλά και με την κορυφή της καμπύλης (mode). Η διακύμανση της κατανομής προκύπτει από την σχέση $\frac{\alpha^2\Gamma(3/\beta)}{\Gamma(1/\beta)}$. β συντελεστής σχήματος και α συντελεστής κλίμακας της συνάρτησης.

Ο συντελεστής β για Gaussian κατανομές παίρνει τιμή $\beta = 2$ (με μέση τιμή μ και διακύμανση $\alpha^2/2$), ενώ $\beta = 1$ όταν η κατανομή είναι Laplace. Καθώς το β μεγαλώνει και τείνει στο άπειρο η κατανομή συγκλίνει στην ομοιόμορφη κατανομή με όρια $(\mu - \alpha, \mu + \alpha)$. Η γενικευμένη Gaussian συνάρτηση για διαφορετικές τιμές του β μπορεί να προσεγγίσει ένα εύρος από συμμετρικές κατανομές, όπως *Λεπτό κυρτές* και τιμές β από 1 έως 2 και *Πλάτυ κυρτές* με τιμές β μεγαλύτερες του 2 έως το άπειρο.

Για τον ακριβή προσδιορισμό των παραμέτρων της παραπάνω κατανομής και συγκεκριμένα του α, β έχει προταθεί η χρήση της μεθόδου Μέγιστης Πιθανοφάνειας (Maximum Likelihood Estimation-MLE) [79] όπως περιγράφεται εδώ [80] [81].



Εικόνα 2.12 Παραδείγματα Κατανομών μιας Γενικευμένης Gaussian Συνάρτησης.

Βασική Ιδέα

Στις δημοσιεύσεις των Kim *et al.* [82] [83] προτείνουν την μοντελοποίηση του background χρησιμοποιώντας γενικευμένη Gaussian κατανομή έτσι ώστε να αντιμετωπίσουν backgrounds από εξωτερικά περιβάλλοντα καθώς και τις σκιές σε αυτό. Η ιδέα είναι ότι η διακύμανση του pixel μερικές φορές ταιριάζει σε Gaussian κατανομή και άλλες φορές σε Laplace. Πράγματι σε εσωτερικούς χώρους με στατικό περιβάλλον, η διακύμανση του pixel είναι πιο κοντά σε Laplace κατανομή από ότι σε Gaussian. Η διακύμανση του pixel σε ένα στατικό περιβάλλον στο χρόνο δίνεται από την σχέση

$$P(X_t) = \frac{\rho\gamma}{2\Gamma(1/\rho)} e^{-(\gamma^\rho|x-\mu|^\rho)} \quad , \text{ με } \gamma = \frac{1}{\sigma} \left(\frac{\Gamma(3/\rho)}{\Gamma(1/\rho)} \right) \quad 2.17$$

Όπου $\Gamma(\cdot)$ συνάρτηση Γάμμα, σ^2 τυπική απόκλιση. Στο συγκεκριμένο μοντέλο το ρ παίρνει μόνο δύο τιμές: για $\rho = 1$ η συνάρτηση αντιπροσωπεύει Laplace κατανομή

και για $\rho = 2$ η συνάρτηση αντιπροσωπεύει Gaussian κατανομή. Στο συγκεκριμένο μοντέλο η επιλογή του ρ γίνεται μέσα από τον υπολογισμό της υπερβάλλουσας κύρτωσης (excess kurtosis). Η υπερβάλλουσα κύρτωση είναι η τιμή της κύρτωσης μείον 3.

$$g_2 = \frac{N \sum_{i=1}^N (x_i - \mu)^4}{(\sum_{i=1}^N (x_i - \mu)^2)^2} - 3 \quad 2.18$$

Η υπερβάλλουσα κύρτωση μετρά την πιθανότητα να υπάρχει κάποια ακραία τιμή σε σχέση με κάποια δοσμένη κατανομή, στην περίπτωση παραπάνω την Gaussian κατανομή για αυτό το λόγο και το μείον τρία. Η υπερβάλλουσα κύρτωση παίρνει τιμές 3 ή 0 για Laplace και Gaussian κατανομή αντίστοιχα.

Ανανέωση Παραμέτρων

Πρακτικά οι Kim *et al.* [82] [83] προτείνουν τη μοντελοποίηση του background σε δυο μέρη για διαφορετικά μοντέλα χρώματος: στο πρώτο χρησιμοποιείται μια συνιστώσα φωτεινότητας όπου είναι απλά ένας σταθμισμένος μέσος όρος των τριών συνιστωσών των βασικών χρωμάτων (RGB) και στο δεύτερο χρησιμοποιείται η H συνιστώσα του HSI χρωματικού μοντέλου. Η ανανέωση των παραμέτρων γίνεται χρησιμοποιώντας ένα σχήμα Running Gaussian Average με εφαρμογή και του μηχανισμού Selectivity όπως αυτό περιγράφεται εδώ [14].

Ανίχνευση Κινούμενων Αντικειμένων

Η ανίχνευση κινούμενων αντικειμένων γίνεται εάν αφαιρέσουμε πρώτα την συνιστώσα της φωτεινότητας του τωρινού frame από την background εικόνα:

$$D(x, y) = |I(x, y) - B(x, y)| \quad 2.19$$

Έπειτα, τα pixel μπορούν να ταξινομηθούν σε τρεις κατηγορίες χρησιμοποιώντας τρία όρια ως εξής:

- Background pixel: $D(x, y) < T_1 k(x, y)$ (a)
- Αμφιλεγόμενα background pixel: $T_1 k(x, y) \leq D(x, y) \leq T_2 k(x, y)$ (b)
- Αμφιλεγόμενα foreground pixel: $T_2 k(x, y) \leq D(x, y) \leq T_3 k(x, y)$ (c)
- Foreground pixel: $T_3 k(x, y) < D(x, y)$ (d)

2.20

Όπου $k(x, y)$ είναι ένας παράγοντας ίσος με σ/β , με σ τυπική απόκλιση της background εικόνας και β ίσο με τρία. Τα όρια T_1, T_2, T_3 προκύπτουν από μια εκπαιδευτική περίοδο. Τέλος, τα αμφιλεγόμενα pixel και στις δύο κατηγορίες που προέρχονται από σκιές μπορεί να συνενωθούν χρησιμοποιώντας την H χρωματική συνιστώσα από το HSI χρωματικό μοντέλο.

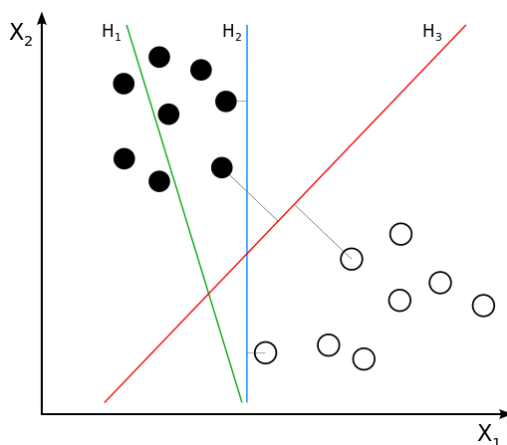
$$if(p \in region(c) \& |H_I - H_B| < T_1 \sigma_H) then p \Rightarrow region(b) \quad 2.21$$

Όπου H_I , H_B , η H χρωματική συνιστώσα από το τωρινό frame και την background εικόνα αντίστοιχα και σ_H η τυπική απόκλιση της H χρωματικής συνιστώσας του background μοντέλου. Τα pixel στην (b), (c) κατηγορία παραπάνω μέσα από μια διαδικασία που περιλαμβάνει τιτλοφόρηση (labeling) και εφαρμογή ορίων, αποφασίζεται ένα θα συμπεριληφθούν ή όχι στο τελικό αποτέλεσμα.

Κάποιες βελτιστοποιήσεις του αλγορίθμου βρίσκονται εδώ [84] [85].

2.2.5 Μηχανές Διανυσμάτων Υποστήριξης (Support Vector Machines - SVM)

Η ταξινόμηση δεδομένων (Classification) είναι μια διαδικασία που ανήκει στον ευρύ τομέα της Μηχανικής Μάθησης (Machine Learning). Έστω ότι υπάρχουν κάποια γνωστά δεδομένα τα οποία ανήκουν σε δύο κλάσες, ο στόχος είναι να αποφασιστεί σωστά σε ποια κλάση ανήκει ένα καινούργιο δεδομένο που θα έρθει ως είσοδο στον αλγόριθμο. Ένας από τους αλγορίθμους που έχει προταθεί για την ταξινόμηση των δεδομένων είναι και ο αλγόριθμος Support Vector Machine (SVM). Η ιδέα ανήκει στους Vapnik *et al.* [86], και σε αυτήν: ο διαχωρισμός κάποιων n -διάστατων διανυσμάτων που αντιπροσωπεύουν δεδομένα σε δύο κλάσεων, γίνεται με την χρήση n -διάστατου υπερεπίπεδου. Το καλύτερο υπερεπίπεδο είναι εκείνο που θα έχει την μεγαλύτερη απόσταση από τις δύο κλάσεις. Δηλαδή επιλέγεται εκείνο το υπερεπίπεδο έτσι ώστε το πιο κοντινό σημείο μεταξύ του υπερεπίπεδου και των δύο κλάσεων να είναι το μέγιστο.



Εικόνα 2.13 Παράδειγμα διαχωρισμού κλάσεων από διαφορετικά υπερεπίπεδα. H_3 το βέλτιστο υπερεπίπεδο.

Βασική Ιδέα

Χρησιμοποιώντας αυτή την βασική αρχή οι Lin *et al.* [87] προτείνουν τη χρήση του SVM αλγορίθμου για τη μοντελοποίηση του Background. Συγκεκριμένα ανέπτυξαν ένα παράλληλο SVM (PSVM) αλγόριθμό με πιθανολογική έξοδο καθώς ο κλασικός SVM αλγόριθμος παράγει μόνο δυαδική έξοδο. Ένα στιγμοειδές μοντέλο [88] χρησιμοποιείται για να μετατρέψει το αποτέλεσμα του SVM αλγορίθμου σε μεταγενέστερη (posterior) πιθανότητα [89].

$$p(y = 1|f) = \frac{1}{1 + \exp(Af + B)} \quad 2.22$$

Όπου y είναι το αναγνωριστικό (label) μιας δυαδικής κλάσης και f είναι το σκορ της συνάρτησης απόφασης του SVM αλγορίθμου. Οι δύο παράμετροι A και B εκτιμώνται με την χρήση της μεθόδου Μέγιστης Πιθανοφάνειας από ένα δείγμα δεδομένων (f_i, y_i) που είναι προς εκπαίδευση για τον αλγόριθμο και η τιμή τους προέρχεται από την ελαχιστοποίηση του αρνητικού φυσικού λογαρίθμου της συνάρτησης Πιθανοφάνειας:

$$\underset{A,B}{\operatorname{argmin}} \left(- \sum_t t_i \log(p_i) + (1 - t_i) \log(1 - p_i) \right) \quad 2.23$$

Όπου

$$t_i = \frac{y_i + 1}{2} \text{ και } p_i = \frac{1}{1 + \exp(Af_i + B)} \quad 2.24$$

Εκπαιδευτική Περίοδος

Για τον αμερόληπτο προσδιορισμό των τιμών των Εξισώσεων 2.22 και 2.23 από τα δεδομένα που είναι προς εκπαίδευση, το 80% των δειγμάτων χρησιμοποιείται για την εκπαίδευση του αλγορίθμου SVM και το άλλο 20% για τον προσδιορισμό των παραμέτρων A και B . Οι Lin *et al.* [87] χρησιμοποιούν σαν δεδομένα προς εκπαίδευση 100 εικόνες διαστάσεων 160*120 όπου το background είναι γνωστό. Κάθε εικόνα χωρίζεται σε μπλοκ διαστάσεων 4*4 και λαμβάνονται υπόψιν δυο χαρακτηριστικά του μπλοκ: το ένα είναι η τιμή της οπτική ροής (optical flow) [90] και το άλλο είναι η διαφορά δύο συνεχόμενων εικόνων.

Κάθε μπλοκ έχει ένα αναγνωριστικό: +1 εάν πρόκειται για background, -1 εάν πρόκειται για foreground. Η αρχικοποίηση του background ξεκινά με την πρώτη εικόνα και κάθε μπλοκ χρησιμοποιείται για είσοδο στον PSVM αλγόριθμο. Ένα μπλοκ κατηγοριοποιείται ως background εάν το αποτέλεσμα της πιθανότητας της Εξίσωσης 2.22 είναι μεγαλύτερη ενός ορίου T :

$$p(b_i) > T \quad 2.25$$

Εάν ένα μπλοκ χαρακτηρίζεται ως background για M συνεχόμενα frames, τότε η απόσταση Fisher υπολογίζεται:

$$d(b_i, b_{back}) = \frac{(\mu_i - \mu_{back})^2}{(\sigma_i^2 - \sigma_{back}^2)} \quad 2.26$$

Όπου μ και σ^2 είναι η μέση τιμή και η διασπορά της έντασης του συγκεκριμένου μπλοκ.

Όταν η απόσταση των δύο μπλοκ είναι μεγάλη δυο πιθανές εκδοχές μπορούν να συμβούν. Το συγκεκριμένο μπλοκ ανήκει σε ομοιόμορφη περιοχή ενός κινούμενου

αντικειμένου ή ένα νέο background αντικείμενο έχει εμφανιστεί. Η μέση τιμή (για τα M συνεχόμενα frames) της πιθανότητας της μεθόδου PSVM για το συγκεκριμένο μπλοκ συγκρίνεται με την πιθανότητα της PSVM μεθόδου του background. Εάν η τιμή είναι μεγαλύτερη από αυτή του background, τότε το μπλοκ εκείνο του background αντικαθίσταται με το καινούργιο μπλοκ.

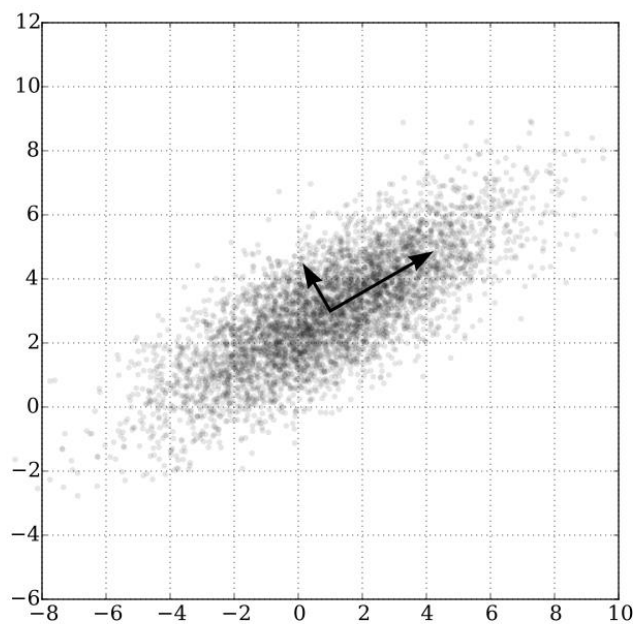
Ανίχνευση Κινούμενων Αντικειμένων

Συνεχίζοντας έτσι, η διαδικασία αρχικοποίησης θα τερματίσει όταν δεν υπάρχουν καθόλου μπλοκ που χρειάζονται αντικατάσταση μέσα σε M συνεχόμενα frames. Όταν η διαδικασία αρχικοποίησης τελειώσει η ανίχνευση κινούμενων αντικειμένων γίνεται απλώς εφαρμόζοντας ένα όριο στην διαφορά της τωρινής εικόνας με την background εικόνα.

Κάποιες διαφοροποιήσεις του αλγορίθμου βρίσκονται εδώ [91] [92] [93] [94] (Support Vector Regression) και εδώ [95] [96] [97] [98] (Support Vector Data Description). Εκτός από τα χαρακτηριστικά της οπτικής ροής και της διαφοράς των frames στην παρακάτω δημοσίευση, μπορούν να χρησιμοποιηθούν και άλλα χαρακτηριστικά εικόνας όπως τα χαρακτηριστικά HOG [99] αλλά και άλλα [100].

2.2.5 Εκμάθηση Υποχώρου και Χρήση Βασικών Συστατικών Στοιχείων (Subspace learning with PCA , SL - PCA)

Η μέθοδος του Subspace learning [101] είναι μια τεχνική για μείωση των διαστάσεων σε ένα σύνολο δεδομένων. Η μείωση των διαστάσεων στον τομέα της μηχανικής όρασης είναι η διαδικασία μείωσης των τυχαίων μεταβλητών έπειτα από παρατηρήσεις. Η εξαγωγή χαρακτηριστικών (Feature extraction) σε μια εικόνα είναι ένα παράδειγμα όπου από n -διάστατα δεδομένα μετασχηματίζονται σε γραμμικά, όπως στον αλγόριθμο Ανάλυσης Βασικών Συστατικών Στοιχείων (Principal Component Analysis - PCA) [102]. Ο PCA αλγόριθμος είναι μια στατιστική διαδικασία στην οποία χρησιμοποιούνται ορθογώνιοι μετασχηματισμοί για να μετατρέψουν ένα σύνολο δειγμάτων που πιθανόν σχετίζονται μεταξύ τους, σε ένα νέο σύνολο από γραμμικά μη συσχετιζόμενα δείγματα τα οποία ονομάζονται βασικά συστατικά στοιχεία (principal components). Ο αλγόριθμος PCA μπορεί να χρησιμοποιηθεί: για την ανάλυση πίνακα σε ιδιάζουσες τιμές (Singular-Value Decomposition - SVD) [103], όπως επίσης και για την Διαγωνιοποίηση και Ιδιοαποσύνθεση του πίνακα συνδιακύμανσης ο οποίος εκφράζεται σε μορφή ιδιοτιμών και ιδιοδιανυσμάτων (Eigenvalue Decomposition) [104].



Εικόνα 2.14 PCA από μια Πολυδιάστατη Gaussian κατανομή με κέντρο το $(1, 3)$ και τυπική απόκλιση 3 στην κατεύθυνση $(0.866, 0.5)$ και 1 στην ορθογώνια σε αυτήν. Τα διανύσματα που φαίνονται είναι τα ιδιοδιανύσματα του πίνακα συνδιακύμανσης κανονικοποιημένα από την αντίστοιχη ιδιοτιμή και μετατοπισμένα έτσι ώστε τα βελάκια να είναι ίδια με το μέσο.

Βασική ιδέα

Η ιδέα μοντελοποίησης του Backgrounds με χρήση Ιδιοχώρων (Eigen Space Models) προτάθηκε από τους Oliver *et al.* [105]. Η χρήση Ιδιοχώρων στην περίπτωση των διαδοχικών frames περιγράφει το εύρος στην διακύμανση των εντάσεων των παρατηρούμενων δειγμάτων.

Εάν έχουμε ένα frame I με διαστάσεις $w \times h$, αυτό πρέπει να μετασχηματιστεί σε ένα διάνυσμα στήλης x . Αυτό μπορεί να γίνει εάν πάρουμε την πρώτη στήλη του frame και την βάλουμε σε ένα νέο διάνυσμα, στην συνέχεια πάρουμε την δεύτερη στήλη του frame και την βάλουμε στο ίδιο διάνυσμα κάτω από την πρώτη εισαγωγή και συνεχίσουμε έτσι για όλες τις στήλες του frame.

$$I_{h,w} \rightarrow x_{n,1}$$

Όπου $n = wh$. Έχοντας μια ακολουθία από N frames $x_{n,1}^1, x_{n,1}^2, \dots, x_{n,1}^N$ δημιουργείται ο παρακάτω πίνακας X . Οι στήλες του πίνακα X αντιπροσωπεύουν τα frames του βίντεο, οπότε ο υποχώρος που δημιουργείται είναι ένας wh -διάστατος χώρος.

$$X_{n,N} = [x_{n,1}^1 \ x_{n,1}^2 \ \dots \ x_{n,1}^N]$$

Στην συνέχεια υπολογίζεται ο αριθμητικός μέσος των frames, $\bar{x}_{n,1}$ και ο πίνακας συνδιασποράς των frames, $C_{n,n}$:

$$\bar{x}_{n,1} = \frac{1}{N} \sum_{i=1}^N x_{n,1}^i, \quad C_{n,n} = \frac{1}{N} \sum_{i=1}^N (x_{n,1}^i - \bar{x}_{n,1}) (x_{n,1}^i - \bar{x}_{n,1})^T$$

Επειδή ο πίνακας συνδιασποράς είναι πραγματικός και συμμετρικός, τότε ο πίνακας της συνδιασποράς διαγωνοποιείται με την μέθοδο της Ιδιοαποσύνθεσης:

$$L_{n,n} = \Phi_{n,n} C_{n,n} \Phi_{n,n}^T \quad 2.27$$

Όπου $\Phi_{n,n}$ ο πίνακας ιδιοδιανυσμάτων και $L_{n,n}$ ο διαγωνιοποιημένος πίνακας με στοιχεία στην διαγώνιο τις ιδιοτιμές. Επειδή τα frames είναι πιθανόν να μοιάζουν μεταξύ τους, υπάρχει η δυνατότητα να αναπαρασταθούν οι στήλες του πίνακα $X_{n,N}$ σε χαμηλότερης διάστασης υποχώρο. Για να μειωθούν οι διαστάσεις του υποχώρου κρατείται μόνο ένας αριθμός M από ιδιοδιανύσματα ($M < N$) τα οποία αποτελούν τα βασικά συστατικά στοιχεία (PCA). Οι μεγαλύτερες M ιδιοτιμές που περιέχονται στον πίνακα $L_{n,n}$ και τα αντίστοιχα M ιδιοδιανύσματα τους τα οποία περιέχονται στον πίνακα $\Phi_{n,n}$, κρατούνται τελικά.

Ο καινούργιος πίνακας που δημιουργείται Φ_M περιέχει M Eigenbackgrounds¹ εικόνες, από αυτόν τον πίνακα μπορεί να υπολογιστεί και πάλι η μέση τιμή:

$$\bar{x}_{n,1} = \frac{1}{M} \sum_{i=1}^M x_{n,1}^i$$

¹ Ελληνικός όρος για το 'Eigenbackgrounds' δεν υπάρχει, για αυτό τον λόγο κρατάμε την αγγλική ορολογία. Σε μια επιχειρούμενη μετάφραση θα ήταν, 'ιδιο-εικόνες του background'.

Ανίχνευση κινούμενων Αντικειμένων

Το καινούργιο frame $x_{n,1}^i$ που θα έρθει ως είσοδος, μπορεί να προσδιοριστεί στον καινούργιο υποχώρο ως εξής:

$$r_{k,1} = (x_{n,1}^i - \bar{x}_{n,1})^T \Phi_M \quad 2.28$$

Στη συνέχεια, η προβολή του $r_{k,1}$ πίσω στο χώρο της εικόνας μας δίνει την ανακατασκευασμένη εικόνα του background:

$$B_t = \Phi_M r_{k,1}^T + \bar{x}_{n,1}$$

Τέλος, η ανίχνευση κινούμενων αντικειμένων γίνεται υπολογίζοντας την απόσταση της εικόνας εισόδου και της ανακατασκευασμένης εικόνας του background:

$$|x_{n,1}^i - B_t| > T$$

Όπου T ένα όριο.

Βελτιώσεις

Αν και η μέθοδος έχει αρκετά προβλήματα τα οποία θα αναφερθούν στην συνέχεια κάποιες βελτιώσεις της μεθόδου μπορούμε να βρούμε εδώ [106].

2.2.6 Αποτελέσματα & Συμπεράσματα

Η χρήση *Μονής Gaussian Συνάρτησης* έχει αρκετά προβλήματα, παρεμφερή με αυτά της μεθόδου στην Παράγραφο 2.1: Δουλεύει καλά μόνο όταν όλα τα pixels, κατά την διαδικασία αρχικοποίησης του background, είναι pixels που ανήκουν στο background. Δε μπορεί να αντιμετωπίσει σταδιακές αλλαγές στο background.

Για παράδειγμα, όταν ένα pixel χαρακτηριστεί ως foreground για μεγάλο διάστημα, η ένταση του background στην περιοχή αυτή μπορεί να έχει αλλάξει εξαιτίας της αλλαγής στον φωτισμό και ως εκ τούτου δεν μπορεί να χαρακτηριστεί σωστά ως background πάλι. Η μέθοδος μπορεί να αντιμετωπίσει μόνο μονότροπα (unimodal) backgrounds, παρόλ' αυτά η μέθοδος είναι σχετικά απλή στην υλοποίησή της, χωρίς πολλές απαιτήσεις σε χρόνο και μνήμη.

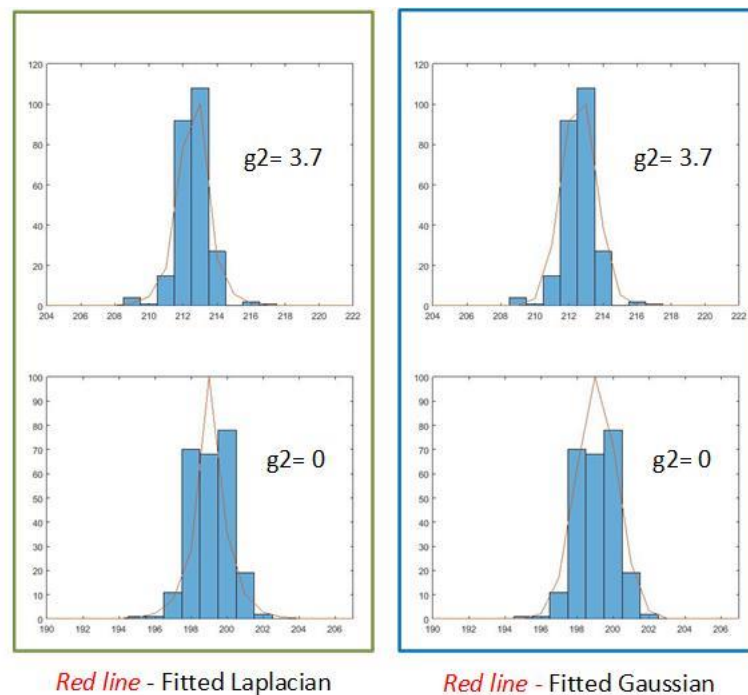
Η βασική μέθοδος *Μίξη Gaussian Συναρτήσεων (MOG)* λόγω της πολυτροπικότητας (multimodality) του μοντέλου που χρησιμοποιεί, βοηθάει στην αντιμετώπιση πολύτροπων backgrounds και στις σταδιακές αλλαγές στις συνθήκες φωτισμού.

Όπως και με τη μονή Gaussian συνάρτηση, δε χρειάζεται να γίνει αποθήκευση μεγάλου όγκου δεδομένων. Παρόλα αυτά κάποια αρνητικά του παραπάνω μοντέλου είναι ότι: ο αριθμός των K Gaussian κατανομών πρέπει να προαποφασιστεί, η σωστή αρχικοποίηση των τιμών είναι σημαντική για την αποτελεσματικότητα του αλγορίθμου, η αντιστοίχιση της δραστηριότητας του pixel σε κάποια σωστή κατανομή είναι πολλές φορές αδύνατη με την χρήση Gaussian κατανομών. Ο χρόνος εκτέλεσης της μεθόδου MOG είναι συγκριτικά μεγαλύτερος από την χρήση απλής Gaussian. [2]

Η βασική μέθοδος Μοντελοποίησης του Background με *Χρήση Συνάρτησης Πυρήνα (KDE)* είναι και αυτή μια πολύτροπη μέθοδος και αντιμετωπίζει εξίσου πολύτροπα backgrounds. Συγκριτικά με την μέθοδο MOG είναι κατάλληλη κυρίως για εξωτερικά περιβάλλοντα [62], λόγω της καλύτερης αντιμετώπισης σε προκλήσεις όπως είναι: το

κούνημα της κάμερας (camera jitter), οι κυματισμοί των κλαδιών των δέντρων (waving trees) αλλά και οι κυματισμοί του νερού (water rippling). Παρόλα αυτά η μέθοδος απαιτεί αρκετή μνήμη, καθώς είναι απαραίτητη η αποθήκευση N frames καθ' όλη την διάρκεια της διαδικασίας. Είναι επίσης και χρονοβόρα σε χρόνους εκτέλεσης, με χρονική πολυπλοκότητα $O(N * N)$.

Η βασική μέθοδος Μοντελοποίησης του Background με *Γενικευμένη Gaussian Συνάρτηση (GGF)*, αν και δεν υποστηρίζει πολύτροπα backgrounds, σύμφωνα με τους εμπνευστές της [82] [83] προσφέρει καλύτερα αποτελέσματα σε σχέση με τις τεχνικές MOG και KDE. Για τη χρησιμότητα της Γενικευμένης Gaussian Συνάρτησης (δηλαδή της επιλογής στην μοντελοποίησης του pixel είτε με Gaussian, είτε με Laplace κατανομή, σύμφωνα με την τιμή της υπερβάλλουσας κύρτωσης) δοκιμάζεται η χρήση των Laplace και Gaussian κατανομών στα ιστογράμματα των pixel του background για το βίντεο με τις ασκήσεις yoga, όπου το περιβάλλον είναι εσωτερικό.



Εικόνα 2.15 Κανονικοποιημένα αποτελέσματα μοντελοποίησης του ιστογράμματος του pixel με Gaussian κατανομή (Μπλε πλαίσιο) και με Laplace κατανομή (Πράσινο πλαίσιο), για διαφορετικές τιμές της υπερβάλλουσας κύρτωσης $g2$.

Τελικά παρατηρείται ότι η χρήση GGF κατανομών δεν είναι και τόσο κρίσιμη στη μοντελοποίηση της έντασης των pixel του background. Η αποτελεσματικότητα της μεθόδου βασίζεται κυρίως στην ασαφή λογική επιλογής των ορίων, όπως επίσης και στους μορφολογικούς αλγορίθμους που εφαρμόζονται στο τελικό αποτέλεσμα. Μια πιο αναλυτική περιγραφή σχετικά με την υλοποίηση της μεθόδου μπορούμε να βρούμε εδώ [107].

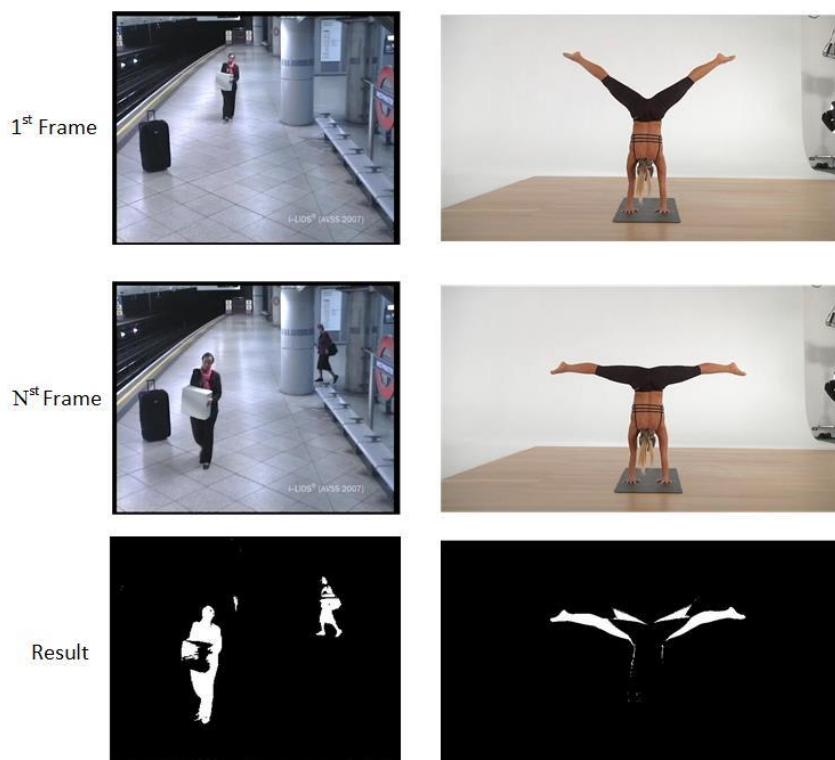
Η βασική μέθοδος Μοντελοποίησης του Background με *Χρήση Ταξινομητών (SVM)* προσφέρει ένα καλό πλαίσιο στην αντιμετώπιση των αλλαγών στις συνθήκες φωτισμού σε ένα βίντεο όπως επίσης και στην μοντελοποίηση δυναμικών backgrounds [62]. Η

μέθοδος δε βασίζεται στη μοντελοποίηση της PDF του pixel μέσω κάποιας γνωστής συνάρτησης (π.χ. Gaussian, Laplace, συναρτήσεις Πυρήνα) και επίσης δεν απαιτεί μνήμη όσο άλλες μη παραμετρικές μέθοδοι.

Μια σύγκριση των μεθόδων *MOG*, *KDE*, *SVM* γίνεται από τους Rajan *et al.* [108], όπου το οπτικό αποτέλεσμα των τριών μεθόδων για τρία διαφορετικά περιβάλλοντα είναι αρκετά πανομοιότυπο.

Τέλος, η βασική μέθοδος Μοντελοποίησης του Background με *Χρήση Εκμάθησης Υποχώρου και του Αλγορίθμου Ανάλυσης Βασικών Συστατικών Στοιχείων (SL-PCA)*, που αναλύθηκε παραπάνω, προσφέρει πάλι την δυνατότητα μοντελοποίησης της PDF του background. Η μέθοδος έχει αρκετά αρνητικά στοιχεία: συγκριτικά με τις άλλες μεθόδους είναι υπολογιστικά χρονοβόρα στην ανανέωση του background, το μοντέλο που χρησιμοποιείται είναι περιορισμένο μόνο για ασπρόμαυρες εικόνες, καθώς για πολλαπλά κανάλια χρωμάτων δεν υπάρχει ευθεία αντιστοιχία.

Και τέλος, όπως και οι παραπάνω βασικές τεχνικές, δεν αντιμετωπίζει το πρόβλημα της παρουσίας foreground αντικειμένων τα οποία δεν μετακινούνται από την αρχική τους θέση (Εικόνα 2.16, ασκήσεις योग).



Εικόνα 2.16 Στην πρώτη σειρά βλέπουμε το πρώτο frame και στην δεύτερη το τελευταίο frame που δέχεται σαν είσοδο η μέθοδος *MOG*. Στην τελευταία σειρά φαίνεται το αποτέλεσμα της μεθόδου *MOG* για το τελευταίο frame για RGB βίντεο και παραμέτρους: Εκπ. περίοδος $N=50$, $K=5$, $T=0.8$, παράγοντας εκμάθησης $a=0.005$.

2.3 Αλγόριθμοι Ομαδοποίησης για Εξαγωγή του Foreground

Η τμηματοποίηση της εικόνας (Image Segmentation) στον τομέα της μηχανικής όρασης αφορά στη διαδικασία τμηματοποίησης μιας ψηφιακής εικόνας σε ομάδες. Πιο συγκεκριμένα, η τμηματοποίηση μιας ψηφιακής εικόνας είναι η διαδικασία όπου κάθε pixel της εικόνας τιτλοφορείται (labeling) με ένα αναγνωριστικό, κάθε φορά που pixels με το ίδιο αναγνωριστικό μοιράζονται κοινά χαρακτηριστικά. Τέτοια χαρακτηριστικά μπορεί να είναι το χρώμα, η ένταση, η υφή. Σε μια εικόνα όπου το χαρακτηριστικό ως προς το οποίο γίνεται η τμηματοποίηση της εικόνας είναι το χρώμα ή ένταση, το ιστόγραμμα της εικόνας περιέχει σημαντική πληροφορία. Τέτοιες τεχνικές ονομάζονται βασισμένες στο ιστόγραμμα (Histogram-based) μέθοδοι. Άλλες τεχνικές τμηματοποίησης είναι: η εφαρμογή ορίου (thresholding), η ανίχνευση ακμών (edge detection) αλλά και η ομαδοποίηση (clustering).

Η βασική αρχή σε αυτή την κατηγορία μοντελοποίησης είναι ότι τα pixels του frame μπορούν να αναπαρασταθούν από συστοιχίες-ομάδες (clusters). Κάθε εισερχόμενο pixel του frame ταυτίζεται με μια από τις διαθέσιμες συστοιχίες και στην συνέχεια η κατηγοριοποίηση του pixel σε background ή foreground έχει να κάνει με το εάν η συγκεκριμένη συστοιχία ανήκει ή όχι σε συστοιχία που αναπαριστά background ή foreground αντίστοιχα.

2.3.1 K-mean Μοντέλο

Στη δημοσίευση των Butler *et al.* [109] προτείνεται η χρησιμοποίηση K συστοιχιών (clusters) για την αναπαράσταση των pixel σε ένα frame. Η ιδέα της προσέγγισης είναι ότι όσο περισσότερο ένα pixel παίρνει ένα συγκεκριμένο χρώμα τόσο περισσότερο είναι πιθανόν να ανήκει στο background.

Βασική ιδέα

Στο μοντέλο που χρησιμοποιείται κάθε pixel μοντελοποιείται από μια ομάδα K συστοιχιών, όπου κάθε συστοιχία αποτελείται από ένα βάρος w_k και από την μέση τιμή των pixels, η οποία ονομάζεται κέντρο βάρους (centroid) c_k της συστοιχίας. Ο λόγος που κάθε pixel αναπαριστάται από μια ομάδα και όχι από μια μονή συστοιχία, είναι εξαιτίας της πολυτροπικότητας του background. Ως εκ τούτου το μέγεθος της ομάδας εκφράζει την πολυτροπικότητα του background και η τιμή του εκτιμήθηκε από 3 έως 5. Ο αλγόριθμος που χρησιμοποιείται για την αφαίρεση του background είναι ο εξής:

Η αρχικοποίηση των τιμών w_k, c_k και κατ' επέκταση του background απαιτεί μια εκπαιδευτική περίοδο από N frames (off-line phase). Κάθε pixel του frame που έρχεται σαν είσοδο στον αλγόριθμο συγκρίνεται με τις K συστοιχίες. Ο σκοπός είναι να βρεθεί η συστοιχία με το μεγαλύτερο βάρος w_k και για αυτό τον λόγο η σύγκριση pixel και συστοιχίας γίνεται με αύξουσα σειρά σύμφωνα με τα βάρη w_k . Για να γίνει η σύγκριση αυτή υπολογίζεται η Manhattan απόσταση μεταξύ του pixel και της συστοιχίας, η οποία δεν είναι τίποτε άλλο από το άθροισμα απόλυτων τιμών και στη συνέχεια εφαρμόζεται ένα όριο καθορισμένο από τον χρήστη που καθορίζει ένα υπάρχει ταύτιση ή όχι. Ο λόγος που χρησιμοποιείται η Manhattan απόσταση και όχι η Ευκλείδεια είναι εξαιτίας

της αποφυγής υπολογισμού πράξεων. Το χρωματικό μοντέλο που χρησιμοποιείται στα πειράματα από τους Butler *et al.* [109] είναι το $Y' C_B C_R$ 4:2:2, όπου η συνιστώσα Y' εκφράζει την φωτεινότητα ενώ οι συνιστώσες $C_B C_R$ εκφράζουν το χρώμα.

Εάν δεν υπάρχει κάποια ταύτιση με κάποια από τις K συστοιχίες τότε η συστοιχία με το μικρότερο βάρος αντικαθίσταται με μια νέα, η οποία θα έχει κέντρο βάρους την τιμή του pixel και αρχικό βάρος κάποια μικρή τιμή (0,01 στην υλοποίηση των Butler *et al.* [109]). Η τιμή αυτή εκφράζει την πιθανότητα η συστοιχία να ανήκει στο background, ως εκ τούτου η τιμή της εξαρτάται από την μορφή του background. Backgrounds με γρήγορο ρυθμό αλλαγών απαιτούν και μεγαλύτερη αρχική τιμή.

Εάν από την άλλη βρεθεί ταύτιση με κάποια συστοιχία τότε τα βάρη όλων των συστοιχιών ανανεώνονται χρησιμοποιώντας την σχέση παρακάτω:

$$w'_k = \begin{cases} w_k + \frac{1}{L}(1 - w_k), & k = M_k \\ w_k + \frac{1}{L}(0 - w_k), & k \neq M_k \end{cases}$$

Όπου M_k ο δείκτης της συγκλίνουσα συστοιχίας. Η παράμετρος L είναι το αντίστροφο του παράγοντα εκμάθησης και ως εκ τούτου εκφράζει τον ρυθμό προσαρμογής του background στις αλλαγές.

Η ανανέωση του κέντρου βάρους δίνεται από τη σχέση

$$c'_k = c_k + (1/L)(x_t - c_k)$$

Ανίχνευση Κινούμενων Αντικειμένων

Η τιμή του βάρους w_k εκφράζει το σύνολο των φορών που έχει γίνει ταύτιση με τη συγκεκριμένη συστοιχία και γι' αυτό το λόγο αποθηκεύει πληροφορία για την τιμή του pixel στο χρόνο. Εάν η τιμή είναι μεγάλη ξέρουμε ότι το pixel έχει τιμή περίπου ίδια με το κέντρο βάρους c_k της συστοιχίας και σύμφωνα με την παραδοχή που έχει γίνει στην αρχή τότε η συστοιχία ανήκει στο background. Στην αντίθετη περίπτωση, εάν η τιμή είναι μικρή τότε το χρώμα του κέντρο βάρους δεν εμφανίζεται πολύ συχνά και η συστοιχία ανήκει στο foreground. Για να διασφαλιστεί ότι το άθροισμα των βαρών όλων των συστοιχιών αθροίζει στο 1 γίνεται κανονικοποίηση με το άθροισμα των βαρών.

$$w'_k = \frac{w_k}{S}, \quad \text{Όπου } S = \sum_k W_k$$

Αυτό γίνεται έτσι ώστε τα βάρη να αντιπροσωπεύουν ποσοστό επί του background για κάθε συστοιχία και για αυτό το λόγο μπορούν να αντιμετωπιστούν σαν πιθανότητες.

Τέλος, για τα εισερχόμενα pixels υπολογίζεται το άθροισμα όλων των βαρών των συστοιχιών W_k , μόνο όμως για τις συστοιχίες που το βάρος τους είναι μεγαλύτερο από την ταυτισμένη συστοιχία. Για αυτό το λόγο, όπως προαναφέρθηκε, είναι αναγκαία η ταξινόμηση με αύξουσα σειρά των βαρών w_k .

$$P = \sum_{k > M_k}^{K-1} w'_k$$

Τα pixels με τιμή P μεγαλύτερη ενός ορίου χαρακτηρίζονται ως foreground.

2.3.2 Χρήση Codebook

Στο ίδιο μήκος κύματος με τα K-mean μοντέλα, είναι η μοντελοποίηση και η αφαίρεση του background με χρήση codebook.

Βασική Ιδέα

Στη δημοσίευση των Kim *et al.* [110] προτείνεται η μοντελοποίηση του pixel με την εξής τρόπο:

Έστω ότι έχουμε μια εκπαιδευτική περίοδο και X είναι μια ακολουθία από N RGB τιμές ενός pixel για εκείνη την περίοδο: $X = \{x_1, x_2, \dots, x_N\}$. Σε κάθε pixel λοιπόν αντιστοιχεί ένα codebook της μορφής $C = \{c_1, c_2, \dots, c_L\}$, όπου L ο αριθμός των στοιχείων του codebook. Το μέγεθος του codebook L δεν είναι ίδιο για όλα τα pixel άλλα εξαρτάται από την διακύμανση των τιμών του pixel.

Κάθε στοιχείο λοιπόν του codebook c_i αποτελείται από ένα RGB διάνυσμα: $v_i = (R_i, G_i, B_i)$, καθώς και μια πλειάδα (tuple) από 6 στοιχεία: $aux_i = \langle \hat{I}_i, \hat{I}_i, f_i, \lambda_i, p_i, q_i \rangle$. Τα έξη αυτά στοιχεία περιέχουν πληροφορία για το pixel στο χρόνο:

\hat{I}, \hat{I} : η ελάχιστη και η μέγιστη τιμή της φωτεινότητας αντίστοιχα, τις οποίες το στοιχείο του codebook επιτρέπει.

f : η συχνότητα με την οποία το στοιχείο του codebook εμφανίζεται.

λ : η τιμή αυτή καταγράφει την μεγαλύτερη διάρκεια κατά την εκπαιδευτική περίοδο όπου το στοιχείο του codebook ΔΕΝ προσπελάστηκε.

p, q : η πρώτη και η τελευταία τιμή αντίστοιχα, προσπέλασης χρονικά του στοιχείου του codebook.

Κατασκευή Codebook

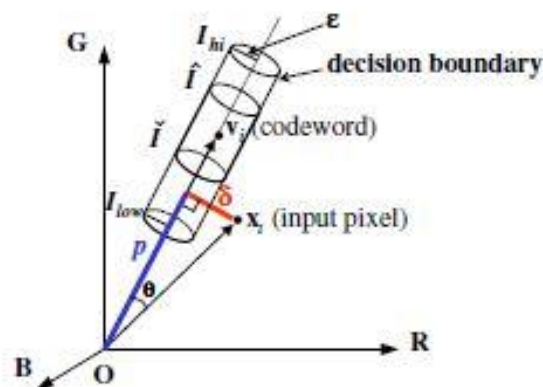
Κατά τη διαδικασία αρχικοποίησης του codebook, κάθε pixel του frame επεξεργάζεται με το εξής τρόπο:

Αρχικά υπολογίζεται η συνιστώσα της φωτεινότητας I , η οποία είναι το άθροισμα των τριών συνιστωσών του RGB μοντέλου. Στη συνέχεια, το εισερχόμενο pixel συγκρίνεται με τα στοιχεία του codebook. Για να υπάρξει ταύτιση μεταξύ των δύο θα πρέπει να ισχύουν δύο κανόνες. Πρώτον, η τιμή του I να είναι μεταξύ των τιμών \hat{I}, \hat{I} (οι οποίες τιμές έχουν κλιμακωθεί κατά δυο παράγοντες α, β) και δεύτερον η χρωματική παραμόρφωση των δύο να είναι μικρή, κάτω από ένα όριο (e_1). Για τον λόγο αυτόν χρησιμοποιείται ένα χρωματικό μοντέλο το οποίο διαχωρίζει την πληροφορία του χρώματος από την πληροφορία της φωτεινότητας. Για το εισερχόμενο pixel $x_t = (R, G, B)$ και το στοιχείο του codebook $v_i = (R_i, G_i, B_i)$ υπολογίζονται οι τιμές:

$$\begin{aligned} \|x_t\|^2 &= R^2 + G^2 + B^2 \\ \|v_i\|^2 &= R_i^2 + G_i^2 + B_i^2 \\ \langle x_t, v_i \rangle^2 &= (R_i R + G_i G + B_i B)^2 \end{aligned}$$

Και τελικά η χρωματική παραμόρφωση (color distortion) δ μεταξύ των δύο είναι:

$$p^2 = \|x_t\|^2 \cos^2 \theta = \frac{\langle x_t, v_t \rangle^2}{\|V_t\|^2}, \quad \text{colordist}(x_t, v_i) = \delta = \sqrt{\|x_t\|^2 - p^2} \quad 2.29$$


$$V = (B, C, B)$$

$$V_L = (R, G, B)$$

$$aux_I = \langle I, I, 1, t - 1, t, t \rangle$$

Σε αντίθετη περίπτωση τα στοιχεία του codebook ανανεώνονται ανάλογα:

$$V_m = \frac{f_m R_m + R}{f_m + 1}, \frac{f_m G_m + G}{f_m + 1}, \frac{f_m B_m + B}{f_m + 1}$$

$$aux_M = \langle \min\{I, \check{I}_m\}, \max\{I, \hat{I}_m\}, f_m + 1, \max\{\lambda_m, t - q_m\}, p_m, t \rangle$$

Η διαδικασία συνεχίζεται και για τα N frames της εκπαιδευτική περιόδου.

Αποφυγή Foreground Αντικειμένων

Για την αποφυγή των foreground αντικειμένων κατά την εκπαιδευτική περίοδο, χρησιμοποιείται ένα φιλτράρισμα στα στοιχεία του codebook. Για να γίνει αυτό χρησιμοποιείται η τιμή της πλειάδας λ (MNRL). Τα στοιχεία του codebook που δεν έχουν προσπελαστεί για ένα μεγάλο χρονικό διάστημα διαγράφονται (συνήθως το όριο αυτό είναι ίσο με $N/2$), έτσι το background μένει καθαρό από αυτά τα αντικείμενα.

Ανίχνευση Κινούμενων Αντικειμένων

Για την ανίχνευση των κινούμενων αντικειμένων κάθε εισερχόμενο pixel συγκρίνεται όπως και στην εκπαιδευτική περίοδο με όλα τα στοιχεία του codebook εάν ισχύουν οι δύο προϋποθέσεις παραπάνω τότε το αντικείμενο χαρακτηρίζεται ως background διαφορετικά χαρακτηρίζεται ως foreground. Η μόνη αλλαγή είναι στην επιλογή του ορίου (e_2) το οποίο είναι διαφορετικό από αυτό της εκπαιδευτικής περιόδου (e_1).

Βελτιώσεις

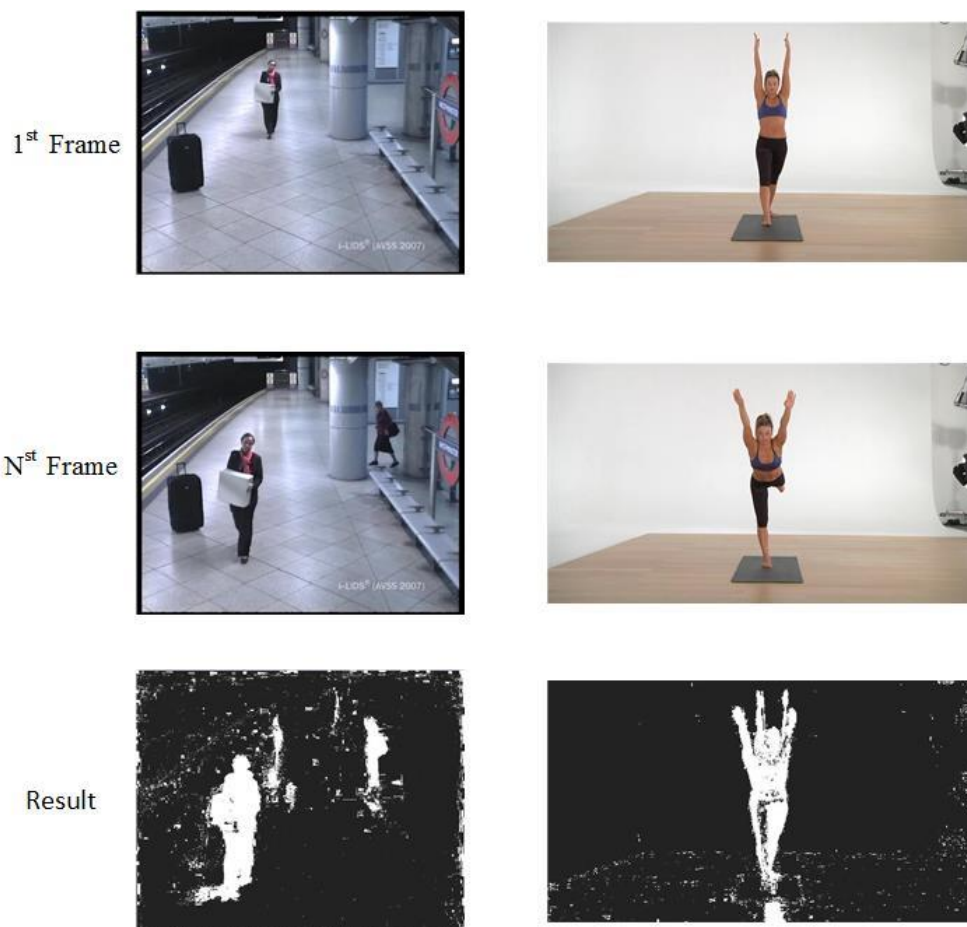
Το codebook δεν αλλάζει κατά την διάρκεια της κανονικής περιόδου (online phase), για αυτό το λόγο οι Kim *et al.* [111] προτείνουν την προσαρμοστικότητα του codebook

κατά την διάρκεια αυτής της περιόδου. Άλλες βελτιώσεις αφορούν την δομή και τα στοιχεία του codebook.

2.3.3 Αποτελέσματα & Συμπεράσματα

Οι αλγόριθμοι των παραπάνω παραγράφων αντιμετωπίζουν εξίσου πολύτροπα background, καθώς και δυναμικά backgrounds. Στη δημοσίευση των Kim *et al.* [110] η μέθοδος με το codebook συγκρίνεται με τις μεθόδους KDE και MOG, από τα πειράματα φαίνεται ότι η μέθοδος έχει εξίσου καλά αποτελέσματα με τις δυο αναφερθείσες μεθόδους, αλλά ο χρόνος επεξεργασίας της μεθόδου είναι αρκετά μεγαλύτερος από τις μεθόδους KDE και MOG.

Και τα δύο μοντέλα που περιγράφονται στις παραγράφους 2.3 υστερούν στην αντιμετώπιση των προκλήσεων που αφορούν το βίντεο με τις ασκήσεις yoga.

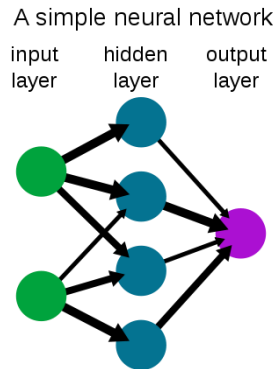


Εικόνα 2.18 Στην πρώτη σειρά βλέπουμε το πρώτο frame και στην δεύτερη το τελευταίο frame που δέχεται σαν είσοδο η μέθοδος με την χρήση Codebook. Στην τελευταία σειρά φαίνεται το αποτέλεσμα της μεθόδου για το τελευταίο frame για RGB βίντεο και παραμέτρους: Εκπαιδευτική περίοδο $N=50$, $e_1=3$, $a=0.55$, $b=1.3$.

2.4 Χρήση Νευρωνικών Δικτύων για Εξαγωγή του Foreground (Neural Networks)

Τα Νευρωνικά Δίκτυα (Neural Networks) είναι ένας τομέας της Μηχανικής Εκμάθησης (Machine Learning) και πιο συγκεκριμένα του τομέα της Βαθιάς Μάθησης (Deep Learning) που τα τελευταία χρόνια γνωρίζει μεγάλη πρόοδο. Η ονομασία και η ιδέα είναι εμπνευσμένη από τα βιολογικά νευρικά συστήματα των ζώων τα οποία εκτός των άλλων εμπεριέχουν νευρώνες. Οι νευρώνες μπορούν να επικοινωνούν μεταξύ τους με συνδέσεις, παίρνοντας αυτό σαν έμπνευση δημιουργήθηκαν τα υπολογιστικά νευρωνικά δίκτυα.

Ένα υπολογιστικό νευρωνικό δίκτυο έχει σαν κύριο δομικό στοιχείο τους κόμβους. Κάθε κόμβος δέχεται ένα σύνολο από αριθμητικές εισόδους (είτε από άλλους νευρώνες είτε από το περιβάλλον), επιτελεί έναν υπολογισμό με βάση αυτές τις εισόδους και παράγει μια έξοδο. Η εν λόγω έξοδος είτε κατευθύνεται στο περιβάλλον, είτε τροφοδοτείται ως είσοδος σε άλλους νευρώνες του δικτύου.



Εικόνα 2.19 Παράδειγμα νευρωνικού δικτύου.

Υπάρχουν τρεις τύποι νευρώνων: οι νευρώνες εισόδου, οι νευρώνες εξόδου και οι υπολογιστικοί νευρώνες ή κρυμμένοι νευρώνες. Οι νευρώνες εισόδου δεν επιτελούν κάποιον υπολογισμό, μεσολαβούν απλώς ανάμεσα στις περιβαλλοντικές εισόδους του δικτύου και στους υπολογιστικούς νευρώνες. Οι νευρώνες εξόδου διοχετεύουν στο περιβάλλον τις τελικές αριθμητικές εξόδους του δικτύου. Οι υπολογιστικοί νευρώνες πολλαπλασιάζουν κάθε είσοδο τους με το αντίστοιχο βάρος και υπολογίζουν το ολικό άθροισμα των γινομένων. Το άθροισμα αυτό τροφοδοτείται ως όρισμα στην συνάρτηση ενεργοποίησης, την οποία υλοποιεί εσωτερικά κάθε κόμβος. Η τιμή που λαμβάνει η συνάρτηση για το εν λόγω όρισμα είναι και η έξοδος του νευρώνα για τις τρέχουσες εισόδους και βάρη.

Εάν x_{ki} είναι η i -οστή είσοδος του k νευρώνα, w_{ki} το i -οστό βάρος του k νευρώνα και $\Phi(\cdot)$ η συνάρτηση ενεργοποίησης του νευρωνικού δικτύου, τότε η έξοδος y_k του k νευρώνα δίνεται από την σχέση:

$$y_k = \Phi\left(\sum_{i=0}^N x_{ki} w_{ki}\right)$$

Στον k -οστό νευρώνα υπάρχει ένα βάρος w_{k0} με ιδιαίτερη σημασία, το οποίο καλείται πόλωση ή κατώφλι (bias, threshold). Η τιμή της εισόδου του είναι πάντα η μονάδα $x_{k0} = 1$. Εάν το συνολικό άθροισμα από τις υπόλοιπες εισόδους του νευρώνα είναι μεγαλύτερο από την τιμή αυτή, τότε ο νευρώνας ενεργοποιείται. Εάν είναι μικρότερο, τότε ο νευρώνας παραμένει ανενεργός. Η ιδέα προέκυψε από τα βιολογικά νευρικά κύτταρα. Η συνάρτηση ενεργοποίησης $\Phi(\cdot)$ μπορεί να είναι βηματική (step transfer function), γραμμική (linear transfer function), μη γραμμική (non-linear transfer function), στοχαστική (stochastic transfer function). Το κύριο χαρακτηριστικό των νευρωνικών δικτύων είναι η εγγενής ικανότητα μάθησης. Η μάθηση επιτυγχάνεται μέσω της εκπαίδευσης, μίας επαναληπτικής διαδικασίας σταδιακής προσαρμογής των παραμέτρων του δικτύου (συνήθως των βαρών και της πόλωσης του) σε τιμές κατάλληλες ώστε να επιλύεται με επαρκή επιτυχία το προς εξέταση πρόβλημα [112].

Στην περίπτωση της αφαίρεσης του background σε ένα βίντεο, το background αναπαρίσταται από το μέσο όρο των βαρών ενός νευρωνικού δικτύου το οποίο έχει προκύψει μέσα από μια εκπαιδευτική περίοδο N frames. Το νευρωνικό δίκτυο μαθαίνει εν τέλει να χαρακτηρίζει τα pixel είτε ως background, είτε ως foreground. Κάθε κόμβος-νευρώνας του δικτύου κωδικοποιεί το background είτε για ένα pixel, είτε για όλο το frame. Η ικανότητα των νευρωνικών δικτύων για ενδογενή παραλληλισμό κατά την διάρκεια της ανίχνευση foreground αντικειμένων, τα καθιστά ιδανικά για αποδοτική υλοποίηση τους σε hardware.

2.4.1 Αυτό-Οργανωτικό Νευρωνικό Δίκτυο (Self-Organizing Background Subtraction–SOBS)

Αρκετές μέθοδοι έχουν αναπτυχθεί τα τελευταία χρόνια πάνω στα νευρωνικά δίκτυα. Στη συγκεκριμένη παράγραφο αναλύεται ένα Αυτό-Οργανωτικό Νευρωνικό δίκτυο όπως αυτό περιγράφεται στην δημοσίευση των Maddalena *et al.* [113] [114].

Βασική Ιδέα

Η ιδέα της παραπάνω δημοσίευσης είναι η δημιουργία ενός background μοντέλου το οποίο θα δημιουργείται με έναν αυτό-οργανωτικό τρόπο μαθαίνοντας από τις διακυμάνσεις του background. Το αυτό-οργανωτικό μοντέλο του background θα είναι ένας πίνακας (map) που θα περιέχει περιοχές με κίνηση αλλά και στάσιμες περιοχές, οπότε ο αλγόριθμος θα πρέπει να αναγνωρίζει τις περιοχές με κίνηση και ανάλογα να ανανεώνει το background.

Το χρησιμοποιούμενο νευρωνικό δίκτυο αποτελείται από έναν πίνακα 2 διαστάσεων, όπου τα στοιχεία του αποτελούν τους κόμβους του δικτύου. Ο συγκεκριμένος πίνακας επιτρέπει την αναπαράσταση των δειγμάτων, που ελήφθησαν κατά την εκπαιδευτική περίοδο, σε χαμηλότερης διάστασης απεικόνιση από αυτή που κανονικά χρειάζονται για να αναπαρασταθούν, ενώ ταυτόχρονα διατηρεί και την τοπολογία των περιοχών (δεδομένων). Η εκμάθηση του νευρωνικού δικτύου επιτυγχάνεται με την σταθμισμένη ανανέωση των γειτονικών τιμών του κόμβου-νευρώνα, για τον οποίο υπάρχει ομοιότητα (σύμφωνα με μια μετρική) με το εισερχόμενο δεδομένο.

Πρακτικά, για κάθε pixel της εικόνας δημιουργείται ένας $n \times n$ πίνακας βαρών. Το χρωματικό μοντέλο που χρησιμοποιείται για την αναπαράσταση είναι το HSV, το οποίο διαχωρίζει την πληροφορία της φωτεινότητας από την χρωματική. Αυτό σημαίνει ότι ο πίνακας με τα βάρη $c_i, i = 1, \dots, n^2$ είναι στην ουσία ένας τρισδιάστατος πίνακας.

Η αρχικοποίηση του πίνακα βαρών c προκύπτει από την τιμή του pixel για το πρώτο frames I_0 . Εάν για παράδειγμα (h, s, v) είναι οι συνιστώσες του HSV χρωματικού μοντέλου για το pixel στην θέση (x, y) και $c = (c_1, c_2, \dots, c_{n^2})$ ο πίνακας βαρών του, τότε κάθε στοιχείο του $c_i, i = 1, \dots, n^2$, θα είναι ένα τρισδιάστατο διάνυσμα όπου $c_i = (h, s, v)$.

Ο πίνακας που θα δημιουργηθεί για μια εικόνα με N σειρές και M στήλες, θα είναι ένα πίνακας B με $n \times N$ σειρές και $n \times M$ στήλες. Ο πίνακας B αποτελεί μια πρώτη εκτίμηση για το background και μπορεί να χρησιμοποιηθεί για την ανίχνευση foreground αντικειμένων.

Ανίχνευση Foreground Αντικειμένων

Ένα εισερχόμενο pixel p_t για το t frame του βίντεο συγκρίνεται με τον πίνακα βαρών του συγκεκριμένου pixel $C(p_t) = (c_1(p_t), \dots, c_{n^2}(p_t))$ για να προσδιοριστεί εάν υπάρχει κάποιος φορέας βάρους $c_m(p_t)$ που να ικανοποιεί την παρακάτω σχέση:

$$d(c_m(p_t), p_t) = \min_{i=1, \dots, n^2} d(c_i(p_t), p_t) < \varepsilon \quad 2.30$$

Όπου $d(\cdot)$ η συνάρτηση για την μέτρηση της ομοιότητας των χρωμάτων του $c_m(p_t)$ και του p_t . Το ε είναι ένα όριο, το οποίο είναι διαφορετικό για εκπαιδευτική και κανονική περίοδο. Η μετρική $d(\cdot)$ για το HSV hexcone χρωματικό μοντέλο είναι η ευκλείδεια απόσταση και δίνεται από την σχέση παρακάτω:

$$d(p_i, p_j) = \|(v_i s_i \cos(h_i), v_i s_i \sin(h_i), v_i) - (v_j s_j \cos(h_j), v_j s_j \sin(h_j), v_j)\|_2^2$$

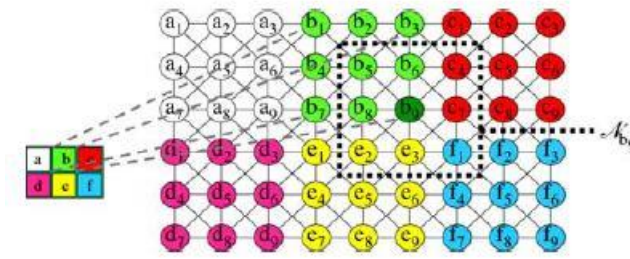
Εάν υπάρχει τελικά τέτοιος φορέας βάρους $c_m(p_t)$ που να ικανοποιεί την σχέση παραπάνω (2.31) το pixel (x, y) χαρακτηρίζεται ως background αλλιώς ως foreground. Πριν γίνει ο τελικός διαχωρισμός υπάρχει ένα ακόμα τελικό στάδιο όπου το pixel ελέγχεται έτσι ώστε να διαπιστωθεί εάν πρόκειται για σκιά αντικειμένου η κανονικό foreground αντικείμενο. Για τον παραπάνω έλεγχο είναι απαραίτητο να προκαθοριστούν κάποιοι παράμετροι.

Ανανέωση Background

Εάν ο φορέας βάρους $c_m(p_t)$ βρεθεί για το pixel p_t και βρίσκεται στην θέση (\bar{x}, \bar{y}) στην background εικόνα B , τότε τόσο ο φορέας βάρους $c_m(p_t)$ καθώς και τα $n \times n$ γειτονικά του ανανεώνονται σύμφωνα με ένα σχήμα επιλεκτικού μέσου όρου (selective weighted running average):

$$B_t(i, j) = (1 - a_{i,j}(t)) B_{t-1}(i, j) + a_{i,j}(t) p_t$$

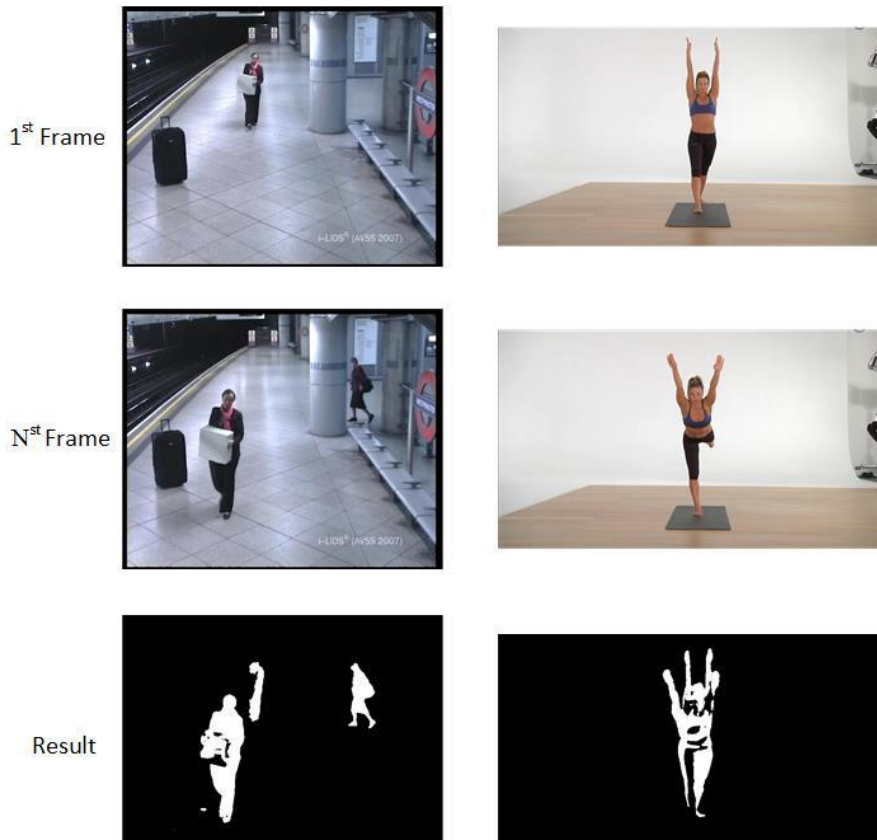
Για $i = \bar{x} - \lfloor \frac{n}{2} \rfloor, \dots, \bar{x} + \lfloor \frac{n}{2} \rfloor, j = \bar{y} - \lfloor \frac{n}{2} \rfloor, \dots, \bar{y} + \lfloor \frac{n}{2} \rfloor$. Όπου $a_{i,j}(t)$ είναι ο παράγοντας εκμάθησης και ο υπολογισμός του λαμβάνει υπόψιν την περίοδο εκτέλεσης (εκπαιδευτική ή κανονική), καθώς και προκαθορισμένες τιμές που έχουν σταθμιστεί από τα βάρη των γειτονικών βαρών και μιας Gaussian ρήτρας.



Εικόνα 2.20 Παράδειγμα εικόνας (Αριστερά), ο νευρωνικός χάρτης (Δεξιά). N_{bg} είναι η περιοχή όπου γίνεται η ανανέωση του background για $c_m(p_t) = b_9$.

Η τιμή του n είναι συνήθως ίση με 3, καθώς για μεγαλύτερο αριθμό ο χρόνος εκτέλεσης του αλγορίθμου μεγαλώνει εκθετικά. Για την αρχικοποίηση του background και κατ' επέκταση του νευρωνικού χάρτη απαιτείται μια εκπαιδευτική περίοδο όπου ο αλγόριθμος θα επαναλαμβάνει την παραπάνω διαδικασία για K αριθμό frames.

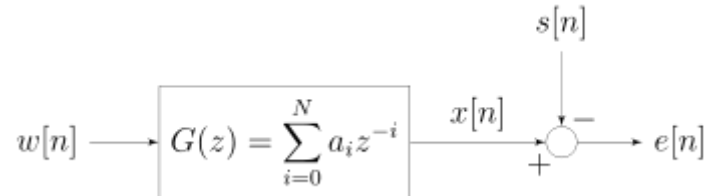
Τέλος, τα αποτελέσματα της μεθόδου είναι και εδώ μη ικανοποιητικά (Εικόνα 2.21), καθώς προκλήσεις που αναφέρθηκαν παραπάνω πάλι δεν αντιμετωπίστηκαν από την μέθοδο.



Εικόνα 2.21 Στην πρώτη σειρά βλέπουμε το πρώτο frame και στην δεύτερη το τελευταίο frame που δέχεται σαν είσοδο η μέθοδος με την χρήση Νευρωνικού Δικτύου. Στην τελευταία σειρά φαίνεται το αποτέλεσμα της μεθόδου για το τελευταίο frame για RGB βίντεο και παραμέτρους: Εκπαιδευτική περίοδο $K=50$, $n=3$, $e1=0.5$, $e2=0.005$, $c2=0.1$.

2.5 Μοντελοποίηση του Background με Θεωρία Εκτίμησης

Άλλη μια κατηγορία αφαίρεσης του background σε ένα βίντεο, την οποία αναφέρουμε για λόγους πληρότητας, είναι η εκτίμηση του background με χρήση φίλτρων. Στην κατεύθυνση της αφαίρεσης του background, κάθε pixel της εικόνας που διαφέρει σημαντικά από την προβλεπόμενη τιμή του, χαρακτηρίζεται ως foreground. Ένα από τα φίλτρα που χρησιμοποιείται είναι το Wiener φίλτρο που υπολογίζει το μέσο τετραγωνικό σφάλμα (mean square error-MSE) μεταξύ της εκτίμησης του σήματος και του επιθυμητού σήματος.



Εικόνα 2.22 Wiener Φίλτρο με είσοδο διακριτό σήμα $w[n]$.

Εκτός από το Wiener φίλτρο, έχουν προταθεί και άλλα φίλτρα. Στον πίνακα παρακάτω φαίνονται οι τρεις βασικές κατηγορίες φίλτρων, καθώς και οι δημοσιεύσεις πάνω σε αυτές.

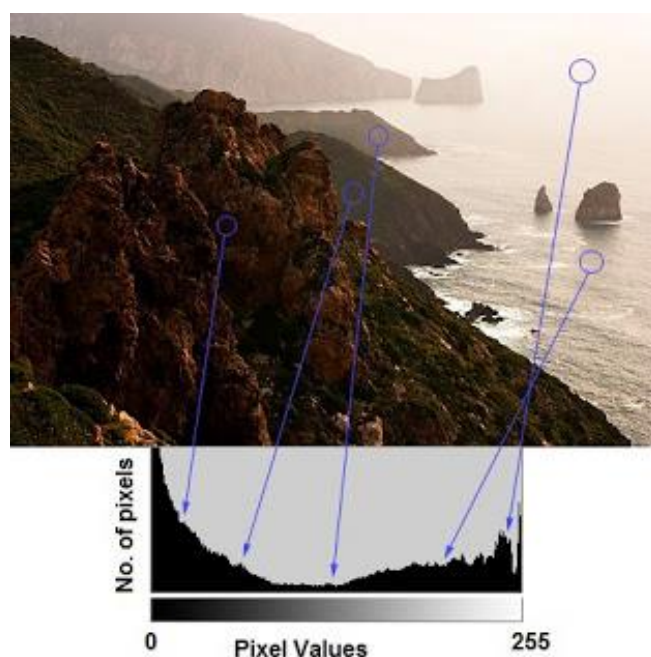
Μοντέλα Εκτίμησης	Δημοσιεύσεις
Wiener φίλτρο	Toyama <i>et al.</i> [115]
Kalman φίλτρο	Messelodi <i>et al.</i> [116]
Chebyshev φίλτρο	Chang <i>et al.</i> [117]

Πίνακας 2-3 Δημοσιεύσεις για διάφορα μοντέλα εκτίμησης.

3. Κεφάλαιο – Προτεινόμενη Μέθοδος Ανίχνευσης Foreground Αντικειμένων

3.1 Εισαγωγή

Στο Κεφάλαιο 2 παρουσιάστηκαν παραδοσιακές τεχνικές αφαίρεσης του background. Δυστυχώς, το σύνολο των τεχνικών αυτών δε μπορεί να αντιμετωπίσει το γεγονός ότι το foreground αντικείμενο κατά τη διάρκεια εκτέλεσης της μεθόδου μετακινείται ελάχιστα, στο σύνολο του ή σε κάποια σημεία του, από την αρχική του θέση. Αυτός ο περιοριστικός παράγοντας σε ένα βίντεο, μας οδηγεί στο συμπέρασμα ότι πρέπει να γίνει αξιοποίηση της χωρικής πληροφορίας σε επίπεδο frame. Με τον όρο χωρική πληροφορία σε ένα frame νοείται η ομαδοποίηση των pixels σε ομάδες ανάλογα με κάποιο χαρακτηριστικό τους, όπως το χρώμα για παράδειγμα. Το ιστόγραμμα μιας εικόνας μπορεί να μας δώσει πληροφορίες σχετικά με ομοιογενείς περιοχές όπως φαίνεται και στην εικόνα παρακάτω (Εικόνα 3.1). Γι' αυτό το λόγο η έννοια των Gaussian Mixture Models (GMM), εκτός από την αφαίρεση του background σε ένα βίντεο (Παράγραφος 2.2.2), μπορεί να χρησιμοποιηθεί και για την τμηματοποίηση της εικόνας (image segmentation). Τελικά, ο αλγόριθμος τμηματοποίησης που επιλέγεται είναι η μέθοδος K-means που μοιάζει ως προς στην φύση του με τον αλγόριθμο K-mean που χρησιμοποιείται για την αφαίρεση του background (Παράγραφο 2.3.1).



Εικόνα 3.1 Παράδειγμα εικόνας (πάνω) και το ιστόγραμμα της (κάτω).

Εκτός από τη χωρική πληροφορία για κάθε frame, πρέπει να υπάρχει και μηχανισμός ανίχνευσης των foreground αντικειμένων παρόμοιο με εκείνους που παρουσιάστηκαν στο προηγούμενο κεφάλαιο. Ένας απλός αλγόριθμος

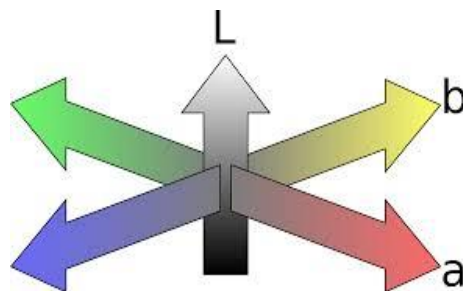
μοντελοποίησης του background, είναι η μέθοδος της μέσης τιμής που αναφέρεται παραπάνω στην δημοσίευση των Wren *et al.* [14] (Παράγραφος 2.2.1). Στην συγκεκριμένη υλοποίηση, η background εικόνα είναι η μέση τιμή των προηγούμενων frames. Ο συγκεκριμένος αλγόριθμος, εάν και μη βέλτιστος στην αφαίρεση του background, επιλέχθηκε καθώς σε συνδυασμό με κάποιους μορφολογικούς αλγορίθμους μπορεί να προσδιορίσει μια περιοχή ενδιαφέροντος όπου το foreground αντικείμενο εμπεριέχεται στο σύνολο του ή ένα τμήμα αυτού.

Στη συνέχεια προτείνεται μια απλή τεχνική ανακατασκευής του background (background reconstruction) αφαιρώντας από το frame την χρονική στιγμή $t = 1$ τα εν δυνάμει foreground αντικείμενα. Στη συγκεκριμένη τεχνική, το χρώμα των εν δυνάμει foreground περιοχών αντικαθίσταται με το επικρατέστερο χρώμα των γειτονικών τους περιοχών που όμως ανήκουν στο background. Ο προσδιορισμός των εν δυνάμει foreground περιοχών προκύπτει συνδυάζοντας τις δύο πληροφορίες παραπάνω: την μια που προκύπτει από την τμηματοποίηση της εικόνας και την άλλη που προκύπτει από την ανίχνευση των περιοχών κίνησης. Η ανακατασκευασμένη εικόνα που θα δημιουργηθεί θα αποτελεί μια πρώτη εκτίμηση για το δεύτερο background μοντέλο που η προτεινόμενη μέθοδος χρησιμοποιεί για την τελική εξαγωγή των foreground αντικειμένων. Τέλος, η προσαρμοστική ανανέωση του δεύτερου background μοντέλου γίνεται συνδυάζοντας την πληροφορία και από τις τρεις εικόνες παραπάνω (δυαδική εικόνα με τις περιοχές κίνησης, τμηματοποιημένη εικόνα, δυαδική εικόνα με τις τελικές περιοχές foreground).

3.2 Θεωρητικό Υπόβαθρο

3.2.1 Χρωματικό Μοντέλο $L^*a^*b^*$

Το χρωματικό μοντέλο που χρησιμοποιείται στην προτεινόμενη μέθοδο και για τα δύο μοντέλα αναπαράστασης της εικόνας είναι το $L^*a^*b^*$ καθώς σαν χρωματικό μοντέλο διαχωρίζει την χρωματική πληροφορία από την πληροφορία της έντασης της φωτεινότητας με αποτέλεσμα να είναι πιο αποτελεσματικό σε περιβάλλοντα με θόρυβο.



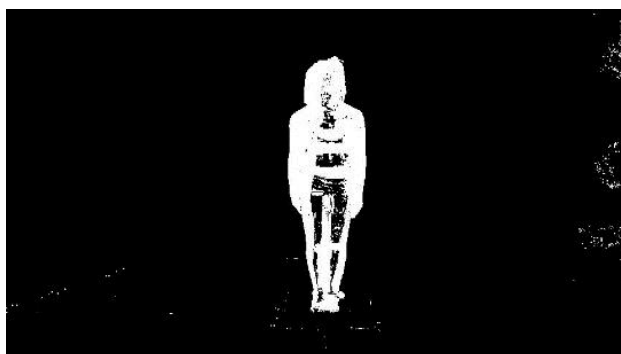
Εικόνα 3.2 Χρωματικό Μοντέλο $L^*a^*b^*$

Το χρωματικό μοντέλο CIE Lab ή $L^*a^*b^*$ παρουσιάστηκε από την CIE το 1976. Το μοντέλο αυτό αναλύει το χρώμα του κάθε pixel σε τρεις συνιστώσες: L, a, b. Η συνιστώσα L εκφράζει την ένταση της φωτεινότητας (Luminosity). Η χρωματική συνιστώσα a αναπαριστά το χρώμα ανάμεσα στο κόκκινο και το συμπληρωματικό του

πράσινο, ενώ η χρωματική συνιστώσα b αναπαριστά το χρώμα ανάμεσα στο μπλε και το συμπληρωματικό του κίτρινο. Ένα από τα πλεονεκτήματα του χρωματικού μοντέλου $L^*a^*b^*$ είναι ότι η διαβάθμιση των χρωμάτων είναι από τις πλησιέστερες στην ανθρώπινη αντίληψη.

3.2.2 Ανίχνευση Περιοχών Κίνησης

Η ανίχνευση των περιοχών κίνησης πραγματοποιείται μοντελοποιώντας το background με χρήση μέσης τιμής [14]. Η επιλογή του συγκεκριμένου μοντέλου έναντι κάποιου πιο εξελιγμένου όπως των Stauffer και Grimson [21] έγινε αφενός λόγω της απλότητας στην υλοποίηση και αφετέρου εξαιτίας του περιοριστικού παράγοντα που υπάρχει στο βίντεο. Ο περιοριστικός παράγοντας θέλει το foreground αντικείμενο να μετακινείται ελάχιστα από την αρχική του θέση με αποτέλεσμα ομοιογενείς περιοχές του foreground αντικειμένου να μην αναγνωρίζονται σωστά. Στην Εικόνα 2.16 φαίνεται το αποτέλεσμα χρήσης της μεθόδου των Stauffer και Grimson [21] για το βίντεο με τις ασκήσεις yoga. Παρατηρείται ότι το συγκεκριμένο μοντέλο δεν ανταποκρίνεται αποτελεσματικά στον συγκεκριμένο περιορισμό. Σε αντίθεση με την χρήση μέσης τιμής για την μοντελοποίηση του background, όπου η μέθοδος παράγει περισσότερα λανθασμένα ως foreground (false positive) pixel, παρόλα αυτά η χρήση της ενδείκνυται για το συγκεκριμένο σκοπό καθώς ανιχνεύει περιοχές με κίνηση με μεγαλύτερη ευαισθησία (Εικόνα 3.3).



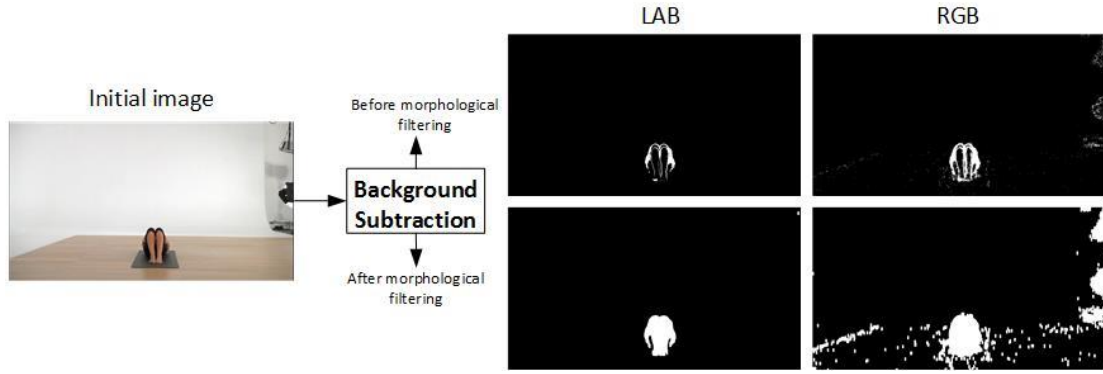
*Εικόνα 3.3 Το αποτέλεσμα της μοντελοποίησης του background με χρήση μέσης τιμής και εφαρμογή ενός ορίου στη χρωματική απόσταση μεταξύ της τρέχουσας εικόνας και της background εικόνας για το χρωματικό μοντέλο $L^*a^*b^*$.*

Στη συνέχεια, με την εφαρμογή κάποιων μορφολογικών φίλτρων και αλγορίθμων διασφαλίζεται στο μέτρο του δυνατού, ο εγκλεισμός του foreground αντικειμένου μέσα στην περιοχή κίνησης (Εικόνα 3.4). Ο κύριος λόγος που χρησιμοποιήθηκε το χρωματικό μοντέλο $L^*a^*b^*$ έναντι του RGB είναι εξαιτίας της καλύτερης απαλοιφής του θορύβου στην τελική δυαδική εικόνα. Το πρόβλημα γίνεται ακόμα πιο έντονο έπειτα από την εφαρμογή των μορφολογικών τεχνικών (Εικόνα 3.5).

Ο σκοπός της παραπάνω μεθοδολογίας είναι η ευρεία ανίχνευση περιοχών κίνησης χωρίς όμως την ανάγκη για ακριβή εξαγωγή των foreground αντικειμένων. Το αποτέλεσμα της συγκεκριμένης μεθόδου αποτελεί την πρώτη background εικόνα που θα χρησιμοποιηθεί στην συνέχεια για τις επόμενες τεχνικές.



Εικόνα 3.4 Αποτέλεσμα αφαίρεσης του background META την χρήση μορφολογικών αλγορίθμων.



Εικόνα 3.5 Παράδειγμα αφαίρεσης του background για τα χρωματικά μοντέλα RGB και $L^*a^*b^*$.

Μαθηματική Αναπαράσταση

Έστω ότι έχουμε μια εκπαιδευτική περίοδο χρονικής διάρκειας T frames. Κάθε frame ενός βίντεο έχει διαστάσεις $M \times N$, κάθε pixel του frame αποτελείται από ένα διάνυσμά 3 τιμών στο $L^*a^*b^*$ χρωματικό μοντέλο. Για παράδειγμα, το pixel στην θέση (x, y) του frame t της εκπαιδευτικής περιόδου, θα αναπαριστάται ως εξής: $p_t = (L_t^*, a_t^*, b_t^*)$. Για όλα τα T frames της εκπαιδευτικής περιόδου πρέπει να υπολογιστεί η μέση τιμή του κάθε pixel. Αυτό που θα προκύψει θα αποτελεί την εκτίμηση του background για το συγκεκριμένο pixel και μπορεί να χρησιμοποιηθεί στην συνέχεια για την εξαγωγή των foreground αντικειμένων. Επειδή κάθε pixel αποτελείται από ένα διάνυσμα τριών τιμών, τότε το αποτέλεσμα που θα προκύψει θα είναι ένας πίνακας τριών διαστάσεων:

$$\begin{aligned} \text{background_model}_{(x,y)_{L,a,b}} &= \frac{1}{T} \sum_{t=1}^T p_t(x, y) \\ &= \frac{1}{T} \sum_{t=1}^T (L_t^*(x, y), a_t^*(x, y), b_t^*(x, y)) \end{aligned} \quad 3.1$$

Ανίχνευση Περιοχών Κίνησης

Κάθε εισερχόμενο pixel συγκρίνεται με το background μοντέλο που δημιουργήθηκε μετά την εκπαιδευτική περίοδο. Η σύγκριση θα γίνει μετρώντας την Ευκλείδεια απόσταση μεταξύ αυτών των δύο τρισδιάστατων διανυσμάτων. Ο οργανισμός International Commission on Illumination (CIE) έχει ονομάσει την απόσταση αυτή ΔE_{ab}^* . Έστω ότι έχουμε δύο pixel $p_i = (L_i^*, a_i^*, b_i^*)$ και $p_j = (L_j^*, a_j^*, b_j^*)$ η απόσταση μεταξύ των δύο θα είναι:

$$\Delta E_{ab}^* = \sqrt{(L_i^* - L_j^*)^2 + (a_i^* - a_j^*)^2 + (b_i^* - b_j^*)^2} \quad 3.2$$

Για να έχουν παρατηρήσιμη διαφορά τα δυο υπό σύγκριση διανύσματα και κατ' επέκταση χρώματα, η τιμή ΔE_{ab}^* πρέπει να είναι μεγαλύτερη του 2.3 [118]. Για λόγους μείωσης των υπολογισμών η τετραγωνική ρίζα στην παραπάνω εξίσωση μπορεί να απαλειφθεί χωρίς να επηρεάσει κατά πολύ το αποτέλεσμα, στην προτεινόμενη μέθοδο η ρίζα έχει απαλειφθεί.

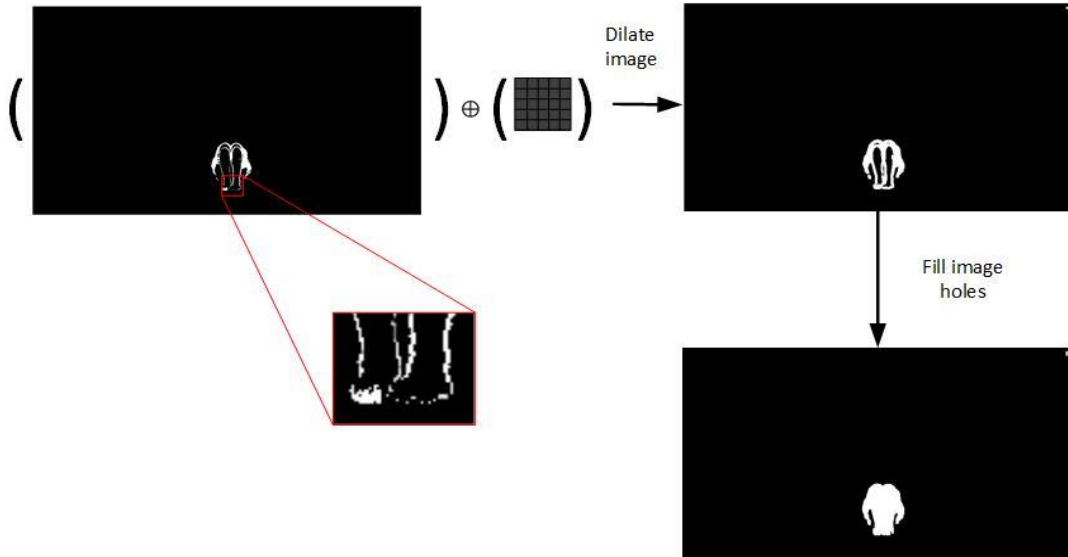
Η εξαγωγή των foreground αντικειμένων θα γίνει με την εφαρμογή ενός ορίου στην τιμή της εξίσωσης ΔE_{ab}^* .

$$foreground_mask_1 = \begin{cases} 0 & \Delta E_{ab}^* < T_{background_1} \\ 1 & else \end{cases} \quad 3.3$$

Το όριο $T_{background_1}$ μπορεί να είναι συναρτήσει της διακύμανσης του pixel κατά την διάρκεια των frames όπως και στην υλοποίηση των Wren *et al.* [14] ή μπορεί να είναι σταθερό και να έχει προαποφασιστεί από τον χρήστη. Στην προτεινόμενη μέθοδο το όριο είναι σταθερό. Από την πειραματική διαδικασία προκύπτει ότι το όριο πρέπει να είναι μεγαλύτερο, όταν το βίντεο είναι εξωτερικού χώρου και έχουμε αρκετό θόρυβο. Ενδεικτικές τιμές του ορίου $T_{background_1}$ είναι από 6~20.

Επειδή το παραπάνω μοντέλο μπορεί να αφήσει κενά στο περίγραμμα της σιλουέτας με αποτέλεσμα στα επόμενα στάδια η προτεινόμενη μέθοδος να γίνει ασταθής, για καλύτερη ανίχνευση των περιοχών κίνησης εφαρμόζεται ένα μορφολογικό φίλτρο καθώς και ένας αλγόριθμος για το 'γέμισμα' του εσωτερικού χώρου του περιγράμματος σε μια δυαδική εικόνα. Οι χώροι αυτοί αντιπροσωπεύουν τις ομοιόμορφες περιοχές του foreground αντικείμενου που εξαιτίας της μικρής κίνησης του δεν παράγει μεγάλη διακύμανση και ως εκ τούτου δεν αναγνωρίζεται σωστά ως foreground.

Στην προτεινόμενη μέθοδο χρησιμοποιείται η μορφολογική πράξη της διαστολής (dilation) μεταξύ ενός μορφολογικού δομικού στοιχείου και της δυαδικής εικόνας του προηγούμενου βήματος. Σαν μορφολογικό δομικό στοιχείο χρησιμοποιείται ένα τετράγωνο 5×5. Τέλος εφαρμόζεται ένας αλγόριθμος για το 'γέμισμα' των εσωτερικών χώρων, βασισμένος στην μορφολογική ανακατασκευή (συνάρτηση *imagefill*) [119].



Εικόνα 3.6 Μορφολογική πράξη διαστολής μεταξύ δυαδικής εικόνας και ενός τετραγώνου 5×5 και στην συνέχεια ‘γέμισμα’ των εσωτερικών χώρων.

Αρά, η τελική δυαδική εικόνα για την πρώτη τεχνική μοντελοποίησης του background θα προκύπτει ως εξής:

$$se = \begin{bmatrix} 1 & \dots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \dots & 1 \end{bmatrix}$$

$$foreground_mask_1 = foreground_mask_1 \oplus se$$

$$foreground_mask_1 = imagefill(foreground_mask_1)$$

Εάν η εικόνα έχει αρκετό θόρυβο μπορεί να χρησιμοποιηθεί μια βαθμίδα αφαίρεσης περιοχών, οι οποίες συγκριτικά με τις διαστάσεις της εικόνας είναι πολύ μικρές. Πριν εφαρμοστεί η παραπάνω τεχνική η εικόνα $foreground_mask_1$ υφίσταται την μορφολογική πράξη της συστολής, άρα:

$$foreground_mask_1 = foreground_mask_1 \ominus se$$

Έστω L_i μια περιοχή της εικόνας $foreground_mask_1$ για την οποία ισχύει ότι $L_i = 1$ σε όλη την έκταση της. Εάν η έκταση της L_i είναι κάτω από ένα όριο τότε μάλλον η περιοχή αυτή προέρχεται από θόρυβο και δεν πρέπει να συμπεριληφθεί στο τελικό αποτέλεσμα. Το όριο αυτό είναι ποσοστιαίο σε σχέση με την συνολική έκταση της εικόνας και συμβολίζεται με T_{perc} που χρησιμοποιείται παρακάτω στην προτεινόμενη μέθοδο.

Ανανέωση του Background

Το σχήμα ανανέωσης του background είναι ένα IIR φίλτρο ή όπως συνηθίζεται να λέγεται στην αγγλική ορολογία running average [14]:

$$\begin{aligned} background_model_{1(x,y)_{L,a,b}} \\ = (1 - \alpha)background_model_{1(x,y)_{L,a,b}} + \alpha p_i(x,y) \end{aligned} \quad 3.4$$

Όπου α παράγοντας εκμάθησης και p_i το τρέχον pixel στην θέση (x, y) του frame i της κανονικής περιόδου.

3.2.3 Τμηματοποίηση Εικόνας

Μια τεχνική τμηματοποίησης της εικόνας είναι: η μέθοδος Gaussian Mixture Models (GMM), η οποία σε άλλη της μορφή χρησιμοποιείται και για την αφαίρεση του background σε ένα βίντεο. Στη συγκεκριμένη μέθοδο γίνεται μοντελοποίηση του ιστογράμματος των pixel του frame έτσι ώστε κάθε πιθανή κορυφή του ιστογράμματος να μοντελοποιηθεί από μια κανονική κατανομή, έτσι κάθε κατανομή αντιπροσωπεύει μια συστάδα (cluster). Κάθε pixel της εικόνας ανήκει σε κάποια i κατανομή, εάν έχει μεγαλύτερη πιθανότητα να ανήκει σε αυτήν την κατανομή σε σχέση με τις άλλες (posterior probability).

Μια άλλη τεχνική τμηματοποίησης είναι η μέθοδος K-means που επίσης χρησιμοποιείται για την αφαίρεση του background. Ομοίως και εδώ τα δεδομένα εισόδου αφορούν τα pixels σε ένα frame. Περιληπτικά η ιδέα του αλγορίθμου είναι ότι έχοντας ένα σετ από δεδομένα (x_1, x_2, \dots, x_n) , όπου κάθε δεδομένο είναι d -διάστασης, ο αλγόριθμός K-means στοχεύει στον διαχωρισμό των n παρατηρήσεων σε k ($k \leq n$) συστάδες (clusters) $S = \{S_1, S_2, S_k\}$, έτσι ώστε να ελαχιστοποιηθεί μέσα στον ίδιο cluster το άθροισμα των τετραγώνων.

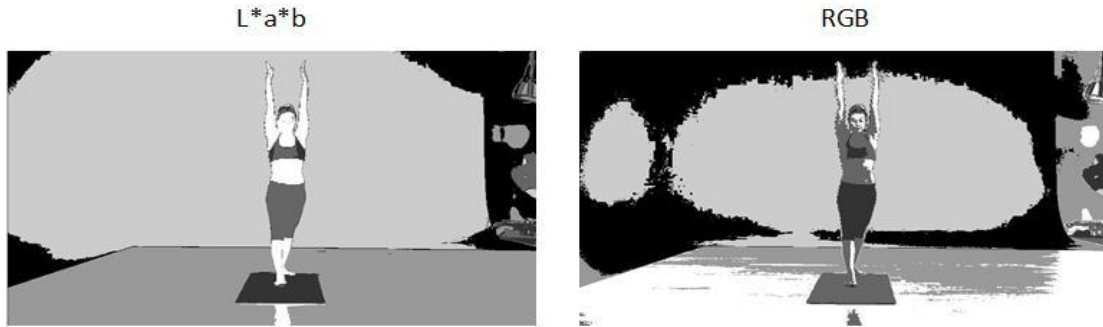
$$\arg \min_S \sum_{i=1}^k \sum_{x \in S_i} \|x - \mu_i\|^2$$

Όπου μ_i είναι η μέση τιμή για τον cluster S_i .

Το αποτέλεσμα είναι ο διαχωρισμός των δεδομένων εισόδου σε διαγράμματα Voronoi. Η βασική ιδέα του αλγορίθμου έχει υλοποιηθεί από τον Lloyd [120]. Ενώ μια βελτίωση του βασικού αλγορίθμου έχει προταθεί από τους David και Vassilvitskii [121]. Η βασική διαφορά των δύο μεθόδων έχει να κάνει στο ότι η πρώτη μέθοδος (Gaussian Mixture Models) δίνει την πιθανότητα ένα pixel να ανήκει σε κάποια i κατανομή, μετά από την εφαρμογή των GMM στα δεδομένα (soft clustering), ενώ η δεύτερη (K-means) κατηγοριοποιεί το pixel σε έναν μόνο cluster (hard clustering). Και στις δύο περιπτώσεις ο αριθμός των clusters πρέπει να προαποφασιστεί πριν την εφαρμογή της εκάστοτε μεθόδου.

Και οι δυο μέθοδοι παρουσιάζουν αρκετά παρόμοιο οπτικό αποτέλεσμα με λίγο καλύτερο αυτό της μεθόδου K-means, για αυτό τον λόγο αλλά και για λόγους ευκολίας στην υλοποίηση προτιμήθηκε η χρήση του αλγορίθμου K-means.

Το χρωματικό μοντέλο RGB χρησιμοποιείται σπάνια σε τεχνικές τμηματοποίησης της εικόνας (Εικόνα 3.7). Τα συνηθέστερα χρωματικά μοντέλα που χρησιμοποιούνται είναι το HSV και το $L^*a^*b^*$, τα οποία ανταποκρίνονται αρκετά ικανοποιητικά σε περιβάλλοντα με θόρυβο [122]. Στην προτεινόμενη μέθοδο για την τμηματοποίηση της εικόνας σε συστάδες χρησιμοποιήθηκαν και οι τρεις συνιστώσες από το χρωματικό μοντέλο $L^*a^*b^*$, σε άλλες υλοποιήσεις προτείνεται η χρήση μόνο των συνιστωσών a^*b^* [123].



Εικόνα 3.7 Παράδειγμα τμηματοποίησης της εικόνας με τον αλγόριθμο K-means και απεικόνιση με χρήση διαγραμμάτων Voronoi, για τα χρωματικά μοντέλα L^*a^*b και RGB και αριθμό συστάδων $K=6$.

Επειδή ο αλγόριθμος K-means είναι αρκετά χρονοβόρος στην εκτέλεση του, προτείνεται η δημιουργία ενός codebook το οποίο θα περιέχει τα βασικά χρώματα της εικόνας. Το codebook θα δημιουργείται αρχικά με την εκτέλεση του αλγορίθμου K-means για το πρώτο frame της εκπαιδευτικής περιόδου και στην συνέχεια κατά την διάρκεια της κανονικής περιόδου θα ανανεώνεται σύμφωνα με μια διαδικασία που περιγράφεται παρακάτω. Κάθε pixel του frame της κανονικής περιόδου θα ταξινομείται σε έναν cluster, ο οποίος θα έχει τον μεγαλύτερο βαθμό ομοιότητας από τους υπόλοιπους. Η ομοιότητα θα υπολογίζεται πάλι μετρώντας την Ευκλείδεια απόσταση μεταξύ των δύο χρωμάτων (του cluster και του pixel) και έτσι θα προκύψει η ταξινόμηση όλων των pixels σε συστάδες. Το μέγεθος του codebook και κατ' επέκταση ο αριθμός των συστάδων K του αλγορίθμου K-means, που θα εκτελεστεί για το πρώτο frame, θα προαποφασίζεται από τον χρήστη.

Η ιδέα της χρήσης codebook το οποίο χρησιμοποιείται στην προτεινόμενη μέθοδο έναντι αυτού που περιγράφεται στην δημοσίευση των Kim *et al.* [110], είναι ότι το codebook της προτεινόμενης μεθόδου αφορά όλα τα pixels του frame σε αντίθεση με τους Kim *et al* [110] όπου κάθε pixel είχε το δικό του codebook.

Μαθηματική Αναπαράσταση

Έστω ένα frame F της εκπαιδευτικής περιόδου με διαστάσεις $M \times N$ και μοντέλο χρωματικής αναπαράστασης το $L^*a^*b^*$. Το μέγεθος του codebook συμβολίζεται ως $\text{Codebook}_{\text{size}}$. Το codebook θα δημιουργηθεί με την εφαρμογή του αλγορίθμου K-means στο αρχικό frame F :

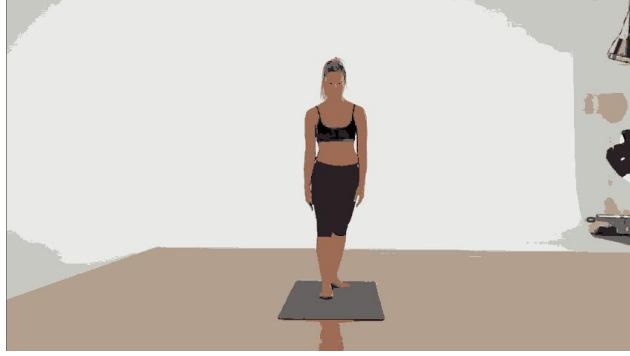
$$\text{codebook} = kmeans(F, \text{Codebook}_{\text{size}})$$

Το codebook θα είναι ένας πίνακας με διαστάσεις $\text{Codebook}_{\text{size}} \times 3$, όπου κάθε σειρά του πίνακα αντιπροσωπεύει την μέση τιμή στο χρωματικό μοντέλο $L^*a^*b^*$ για μια συγκεκριμένη συστάδα k .

Κατά τη διάρκεια της κανονικής περιόδου ένα εισερχόμενο pixel p_i συγκρίνεται με κάθε σειρά του codebook. Η σύγκριση γίνεται πάλι με τον υπολογισμό της ποσότητας ΔE_{ab}^* και η ταξινόμηση του συγκεκριμένου pixel σε κάποια συστάδα, γίνεται για την συστάδα στην οποία καταγράφηκε η μικρότερη τιμή της ποσότητας ΔE_{ab}^* :

$$color_image_{(x,y)} = k \mid \arg \min_k \Delta E_{ab}^* \quad 3.5$$

Η εικόνα `color_image` που προκύπτει έχει ταξινομήσει κάθε pixel του frame σε μια μοναδική συστάδα. Σε μια ενδεχόμενη ανακατασκευή της τιτλοφορημένης εικόνας `color image` με τις μέσες τιμές των συστάδων θα προέκυπτε η παρακάτω εικόνα:



Εικόνα 3.8 Τμηματοποίηση της εικόνας και αναπαράσταση με τιμή για κάθε pixel την μέση τιμή της συστάδα, $K=6$.

Ανανέωση του Codebook

Επειδή το background σε ένα βίντεο είναι δυναμικό με αποτέλεσμα αντικείμενα να εισέρχονται ή να εξέρχονται από το σκηνικό καταγραφής, θα πρέπει να επιληφθεί ένας μηχανισμός ανανέωσης του Codebook.

Η ανανέωση θα γίνει κοιτώντας την τιμή ΔE_{ab}^* που αντιπροσωπεύει τον βαθμό ομοιότητας μεταξύ του εισερχόμενου pixel και κάθε στοιχείου του codebook. Εάν η τιμή αυτή είναι μεγάλη πρέπει να ελεγχθεί εάν το χρώμα αυτό πρέπει να συμπεριληφθεί ή όχι στο codebook. Μια απλή λύση είναι να οριστεί ένα νέο codebook και ένας μετρητής (counter) για κάθε στοιχείο του νέου codebook, η διάρκεια ζωής του codebook και του μετρητή θα είναι για ένα frame. Εάν ο counter έχει τιμή πάνω από ένα όριο τότε το χρώμα του νέου codebook πρέπει να συμπεριληφθεί στο τελικό codebook. Το όριο αυτό εκφράζεται σε ποσοστό επί της συνολικής εικόνας. Τέλος, εάν ένα χρώμα από το codebook δεν έχει προσπελαστεί για αρκετά frames αυτό σημαίνει ότι πρέπει να διαγραφεί από το codebook. Αρά ο μηχανισμός ανανέωσης ενεργοποιείται όταν :

$$\arg \min_k \Delta E_{ab}^* > T_{instert}$$

Όπου $T_{instert}$ το όριο για τον περαιτέρω έλεγχο του χρώματος του pixel p_i . Επειδή ο μηχανισμός εισαγωγής νέου χρώματος στο codebook, για λόγους ταχύτερης εκτέλεσης, δεν πρέπει να ενεργοποιείται πολύ συχνά το όριο $T_{instert}$ πρέπει να είναι αρκετά μεγάλο. Από τα πειράματα το όριο αυτό προσδιορίστηκε γύρω στο 100~300. Για παράδειγμα ένα frame του βίντεο με τις ασκήσεις yoga στο οποίο δεν εισέρχονται νέα αντικείμενα, με όριο $T_{instert} = 200$ θα έχει ως αποτέλεσμα από τα 409920 pixels που έχει το frame ο μηχανισμός ανανέωσης να ενεργοποιηθεί μόνο για περίπου 300 pixels.

Το νέο codebook θα συμβολίζεται ως $codebook_{new}$ και έχει μέγεθος j . Το pixel p_i συγκρίνεται με το $codebook_{new}$, αυτή την φορά μετρίεται η ευκλείδεια απόσταση μεταξύ των συνιστωσών a^*b^* :

$$\Delta E_{new}^* = \sqrt{(a_{p_i}^* - a_{codebook_{new}}^*)^2 + (b_{p_i}^* - b_{codebook_{new}}^*)^2} \quad 3.6$$

$$\arg \min_j \Delta E_{new}^*$$

Ο λόγος που γίνεται αυτό είναι εξαιτίας της συνιστώσας L^* , η οποία περιέχει την πληροφορία για την φωτεινότητα. Στην προτεινόμενη μέθοδο θέλουμε η εισαγωγή να αφορά μόνο τα πραγματικά χρώματα αντικειμένων που εισέρχονται στο background και όχι τυχόν θόρυβο λόγω αντανάκλασεων στην επιφάνεια του αντικειμένου.

Εάν τώρα η παραπάνω ποσότητα είναι μεταξύ ενός ορίου, το οποίο για ευκολία θα είναι συναρτήσει του $T_{instert}$ και συγκεκριμένα $T_{instert}/\text{"ακέραιος αριθμός"}$, ο counter για το συγκεκριμένο χρώμα του πίνακα $codebook_{new}$ θα αυξάνεται κατά ένα αλλιώς θα δημιουργείται ένα νέο χρώμα στον πίνακα $codebook_{new}$:

$$\begin{cases} counter_{codebook_{new_j}} + 1 & \text{if } \arg \min_j \Delta E_{new}^* \leq T_{instert}/\text{"ακέραιος αριθμός"}, \\ codebook_{new} = instert(p_i) & \text{else} \end{cases} \quad 3.7$$

Και τέλος η εισαγωγή στο τελικό codebook γίνεται εάν:

$$counter_{codebook_j} > T_{perc} \quad 3.8$$

Όπου T_{perc} εκφράζει το ποσοστό επί του frame που ένα χρώμα πρέπει να καλύπτει για να εισαχθεί στο τελικό codebook. Συνήθως είναι $0.005 \sim 0.01\% \times \text{size}(\text{frame})$.

Ο λόγος που χρησιμοποιείται το όριο $T_{instert}/\text{"ακέραιος αριθμός"}$ είναι διότι για να μετρηθεί το ποσοστό κάποιου χρώματος επί του frame, θα πρέπει να γίνει μια διακριτοποίηση στα χρώματα του frame. Έστω ο αριθμός των pixels που ενεργοποιείται ο μηχανισμός είναι 300, πράγμα που σημαίνει ότι θα δημιουργηθούν 300 νέα χρώματα που όμως δεν θα εισέρθουν στο codebook εξαιτίας του ποσοστιαίου ορίου T_{perc} . Εάν όμως τα χρώματα χωριστούν σε έναν αριθμό από ομάδες ανάλογα με το πόσο μοιάζουν μεταξύ τους, τότε προκύπτουν τα βασικά χρώματα του frame που ανάλογα στην συνέχεια αποφασίζεται η εισαγωγή τους ή όχι. Ο "ακέραιος αριθμός" είναι μεταξύ 5 έως 10.

3.3 Προτεινόμενη Μεθοδολογία

3.3.1 Απλή Συσχέτιση Τμηματοποιημένης Εικόνας και Περιοχών Κίνησης

Για το συνδυασμό των δύο παραπάνω μεθόδων μια απλή τεχνική εξαγωγής των foreground αντικειμένων μπορεί να γίνει χρησιμοποιώντας την τμηματοποίηση της εικόνας σε ομοιογενείς περιοχές (Εικόνα 3.8) και της δυαδική εικόνας όπου καταγράφονται οι περιοχές κίνησης (Εικόνα 3.4). Εάν μια περιοχή της τμηματοποιημένης εικόνας ανήκει κατά ένα ποσοστό και άνω στην δυαδική εικόνα, τότε η περιοχή αυτή χαρακτηρίζεται ως foreground αλλιώς χαρακτηρίζεται ως background. Μια παρόμοια υλοποίηση αναφέρεται και στη δημοσίευση των X H Fang *et al.* [124] στην οποία χρησιμοποιείται ένας βελτιωμένος αλγόριθμος MOG για την αφαίρεση του background και τελικά την δημιουργία της δυαδικής εικόνας. Τέλος, η τμηματοποίηση κάθε frame του βίντεο προκύπτει από τον αλγόριθμο Mean-Shift.

Μαθηματική Αναπαράσταση

Έστω R μια ομοιογενής περιοχή η οποία προκύπτει μετά από την τμηματοποίηση της εικόνας και F_1 αναπαριστά τη δυαδική εικόνα που προκύπτει μετά την αφαίρεση του background. Εάν το pixel στη θέση (x, y) είναι περιοχή κίνησης τότε η τιμή της εικόνας F_1 στο σημείο θα είναι $F_1(x, y) = 1$, αλλιώς $F_1(x, y) = 0$.

Η έκταση της περιοχή R που καλύπτει την εικόνα F_1 συμβολίζεται με M_R και ισχύει ότι:

$$M_R(x, y) = \begin{cases} 1 & F_1(x, y) \text{ is foreground} \\ 0 & \text{else} \end{cases}, \quad \forall (x, y) \in R \quad 3.9$$

Εάν η έκταση της περιοχής R συμβολίζεται με S_R τότε η τελική εξαγωγή των foreground αντικειμένων θα προκύπτει ως εξής:

$$F_2(x, y) = \begin{cases} 1 & \frac{\text{sum}(M_R(x, y) = '1')}{S_R} > T_{prop} \\ 0 & \text{else} \end{cases}, \quad \forall (x, y) \in R \quad 3.10$$

Η ποσότητα $\text{sum}(M_R(x, y) = '1')/S_R$ δηλώνει το ποσοστό κάλυψης της τμηματοποιημένης εικόνας R από την περιοχής κίνησης που ανήκει ταυτόχρονα και στην περιοχή R ($M_R(x, y) = '1'$). Το T_{prop} εκφράζει το ποσοστό για το οποίο πάνω από αυτό η περιοχή R θεωρείται foreground.

Συμπεράσματα

Η παραπάνω τεχνική εκμεταλλεύεται εν μέρει τη χωρική πληροφορία της εικόνας αλλά η μέθοδος είναι ασταθής σε περιπτώσεις όπου το αντικείμενο κατά την διάρκεια του βίντεο μετακινείται ελάχιστα, με αποτέλεσμα περιοχές του foreground να εξαφανίζονται ξαφνικά και να χαρακτηρίζονται λανθασμένα ως background. Ακόμη, περιοχές του background μπορεί να χαρακτηριστούν λανθασμένα ως foreground εάν είναι εγκλεισμένες από foreground περιοχές. Το ποσοστιαίο όριο T_{prop} εισάγει έναν συμβιβασμό (trade-off) στην μέθοδο, για μεγάλες τιμές προκύπτουν περισσότερα pixel

χαρακτηρισμένα λανθασμένα ως background (false negative) ενώ για μικρές τιμές προκύπτουν περισσότερα pixel χαρακτηρισμένα λανθασμένα ως foreground (false positive). Στην Εικόνα 3.9 φαίνονται οι δυο περιορισμοί της μεθόδου.



Εικόνα 3.9 Παράδειγμα εξαγωγής foreground αντικειμένου στο βίντεο με τις ασκήσεις yoga. Το ποσοστιαίο όριο επικάλυψης για την ανίχνευση είναι $T_{prop} = 0.7$ και αριθμός συστάδων $K=8$. Σε κόκκινο πλαίσιο δεξιά της εικόνας φαίνονται οι δυο περιορισμοί.

3.3.2 Προτεινόμενη Μέθοδος Ανακατασκευής του Background (Background Reconstruction)

Εισαγωγή

Στην παρακάτω Παράγραφο περιγράφεται μια απλή τεχνική ανακατασκευής του background (Background Reconstruction) η οποία θα προσπαθήσει να λύσει το πρόβλημα που δημιουργεί ο περιοριστικός παράγοντας παραπάνω. Ο συγκεκριμένος περιορισμός θέλει περιοχές του background να χαρακτηρίζονται λανθασμένα ως foreground εάν είναι εγκλεισμένες από foreground περιοχές.

Οι παραδοσιακές μέθοδοι αφαίρεσης του background σε ένα βίντεο έχουν σαν στόχο την μοντελοποίηση και την δημιουργία μιας background εικόνας και στη συνέχεια την αφαίρεση του τρέχον frame από την background εικόνα έτσι ώστε να προκύψει η εξαγωγή των foreground αντικειμένων. Το ζητούμενο σε αυτές τις τεχνικές δεν είναι η πιστή αναπαράσταση της background εικόνας αλλά η εξαγωγή των foreground αντικειμένων. Η ανακατασκευή του background έχει σαν στόχο την δημιουργία μιας background εικόνας με ακριβή αναπαράσταση του background σκηνικού χωρίς τα κινούμενα αντικείμενα. Οι διάφορες προκλήσεις αφορούν τα foreground αντικείμενα τα οποία βρίσκονται τον περισσότερο ή όλο τον χρόνο μέσα σε μια περιοχή του background.

Στην αναφορά του Dirk S. F. [125] προτείνεται μια μέθοδος ανακατασκευής του background (SimMat) εκμεταλλευόμενη την πληροφορία από τις γειτονικές περιοχές σε συνδυασμό με την πληροφορία που υπάρχει εξαιτίας της κίνησης των foreground αντικειμένων. Η μέθοδος είναι αποτελεσματική ακόμα και όταν μια περιοχή του background είναι ορατή κάτω από το 50% του συνολικού χρόνου. Η μέθοδος βασίζεται στο γεγονός ότι μια περιοχή που καλύπτεται από κάποιο foreground αντικείμενο για χρόνο t , είναι πιθανό ότι μια άλλη γειτονική περιοχή θα καλύπτεται περίπου για τον ίδιο χρόνο t . Η μέθοδος δεν μπορεί να αντιμετωπίσει περιπτώσεις όπου περιοχές του

background καλύπτονται στο σύνολο του χρόνου από foreground αντικείμενα. Στην αναφορά του Peng Chen [126] προτείνεται η ανακατασκευή του background με χρήση Gaussian Mixture Models, στην οποία λαμβάνεται υπόψιν η χωρική και χρονική πληροφορία προκειμένου να δημιουργηθούν πολλαπλά μοντέλα από τα οποία χρησιμοποιώντας ένα αλγόριθμο ενισχυμένης εκμάθησης (Sarsa(λ)), να προκύψει βάσει μέτρησης της αξιοπιστίας η κατάλληλη τιμή για το pixel. Για την ενίσχυση της αποτελεσματικότητας του παραπάνω αλγορίθμου χρησιμοποιούνται επίσης pre-processing αλγόριθμοι που αφορούν την μείωση του θορύβου καθώς και την ανίχνευση και αφαίρεση των σκιών. Η παραπάνω μεθοδολογία εφαρμόστηκε μόνο σε βίντεο από κυκλώματα παρακολούθησης κυκλοφορίας και ως εκ τούτου δεν υπάρχει σύγκριση για πιο απαιτητικά περιβάλλοντα.

Άλλες, πιο εξελιγμένες τεχνικές ανακατασκευής του background [127] [128] χρησιμοποιούν γειτονικά κελιά από pixels (patch-based) και χρησιμοποιώντας χαρακτηριστικά υφών (texture features) για την εικόνα [127], την πληροφορία της οπτικής ροής (optical flow) και τα χρώματα της εικόνας από την άλλη [128] γίνεται ανακατασκευή περιοχών του βίντεο όπου δεν υπάρχει καμία πληροφορία για το background καθώς foreground αντικείμενα το καλύπτουν πλήρως (video inpainting). Και οι δύο παραπάνω τεχνικές απαιτούν τις περιοχές που είναι προς ανακατασκευή να είναι οριοθετημένες από τον χρήστη.

Ο υπολογισμός των χαρακτηριστικών της εικόνας (π.χ. οπτική ροή) όπως και η συσχέτιση των χαρακτηριστικών σε χώρο-χρονικό επίπεδο αποτελεί μια χρονοβόρα και πολύπλοκη διαδικασία. Επιπροσθέτως, ο περιοριστικός παράγοντας στο βίντεο με τις ασκήσεις yoga όπου το αντικείμενο δε μετακινείται σε κάποια σημεία του αποτελεί έναν ανασταλτικό παράγοντα για την μέτρηση κάποιων χαρακτηριστικών της εικόνας, όπως της οπτικής ροής και των χαρακτηριστικών της υφής.

Στην προτεινόμενη μέθοδο το αρχικό στάδιο ανακατασκευής της background εικόνας πραγματοποιείται αντικαθιστώντας τις εν δυνάμει foreground περιοχές με την πληροφορία των γειτονικών περιοχών τους, που ανήκουν όμως στο background. Εκμεταλλευόμενοι την πληροφορία του codebook και κατ' επέκταση των βασικών χρωμάτων ενός frame, οι εν δυνάμει foreground περιοχές αντικαθίσταται με το επικρατέστερο χρώμα από το codebook. Για την αναγνώριση των εν δυνάμει foreground περιοχών χρησιμοποιείται μια μέθοδος που μοιάζει με εκείνη της Παραγράφου 3.3.1 και περιγράφεται παρακάτω, σε αυτή την περίπτωση είναι απαραίτητος ο ορισμός μιας υπόθεσης.

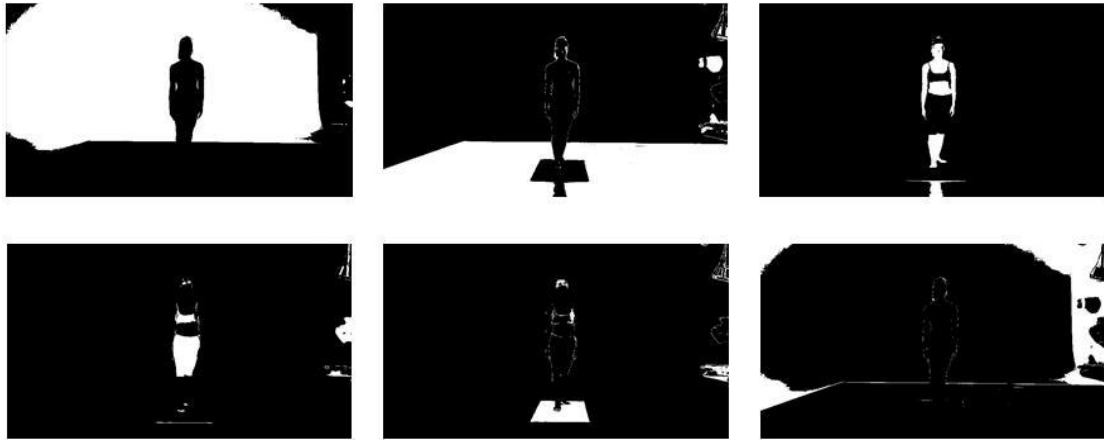
Στη συνέχεια, χρησιμοποιώντας ένα σχήμα επιλεκτικής ανανέωσης του background βασιζόμενο σε περιοχές κίνησης, η αρχική ανακατασκευασμένη εικόνα του background προσαρμόζεται κατά το μέτρο του δυνατού στα πραγματικά δεδομένα της τμηματοποιημένης εικόνας (Παράγραφος 3.2.3). Το κύριο ζητούμενο στην εργασία είναι η εξαγωγή των foreground αντικειμένων, οπότε η χρήση codebook για την αναπαράσταση της τρέχουσας εικόνας απλοποιεί αρκετά τις μαθηματικές πράξεις που χρειάζεται ο αλγόριθμος και προσφέρει έναν καλό σκελετό (framework).

Ένα μειονέκτημα της προτεινόμενης μεθόδου παραπάνω είναι ότι σε απαιτητικά περιβάλλοντα όπου το background περιέχει πολλά χρώματα, θα υπάρχουν πολλές περιοχές ανακατασκευασμένες με λανθασμένο χρώμα. Σε ενδεχόμενη μετακίνηση

λοιπόν του αντικειμένου από την αρχική του θέση, θα έχει ως αποτέλεσμα το σκηνικό του background που τώρα είναι ακάλυπτο και έχει λάθος χρώμα μετά την αρχική ανακατασκευή να αναγνωριστεί λανθασμένα ως foreground (Δείτε αποτελέσματα Κεφαλαίου 4). Το πρόβλημα λύνεται εν μέρει με έναν μεγαλύτερο παράγοντα εκμάθησης για το πρώτο σχήμα μοντελοποίησης του background (Παράγραφος 3.2.2).

Μαθηματική Αναπαράσταση

Έστω $R_{1,...,K}$ οι περιοχές που προκύπτουν μετά από την τμηματοποίηση της εικόνας σύμφωνα με την μέθοδο της Παραγράφου 3.2.3, το σύνολο των περιοχών είναι όσες και οι συστάδες K .



Εικόνα 3.10 Παράδειγμα τμηματοποίησης της εικόνας σε $K=6$ αριθμό συστάδων. Με λευκό χρώμα φαίνονται οι περιοχές για κάθε συστάδα κ .

Με F_1 αναπαρίσταται η δυαδική εικόνα που προκύπτει μετά την αφαίρεση του background σύμφωνα με την μέθοδο της Παραγράφου 3.2.2. Εάν το pixel στη θέση (x, y) είναι περιοχή κίνησης τότε η τιμή της εικόνας F_1 στο σημείο θα είναι $F_1(x, y) = 1$, αλλιώς $F_1(x, y) = 0$.

Η έκταση της κάθε περιοχής $R_{1,...,K}$ που καλύπτει την εικόνα F_1 συμβολίζεται με $M_{R_{1,...,K}}$ και ισχύει ότι:

$$M_{R_i}(x, y) = \begin{cases} 1 & F_1(x, y) \text{ is foreground} \\ 0 & \text{else} \end{cases}, \forall (x, y) \in R_i \quad 3.11$$

Όπου $i = 1, \dots, K$

Η συνολική έκταση της συστάδας κ συμβολίζεται με S_{R_i} , για $i = 1, \dots, K$

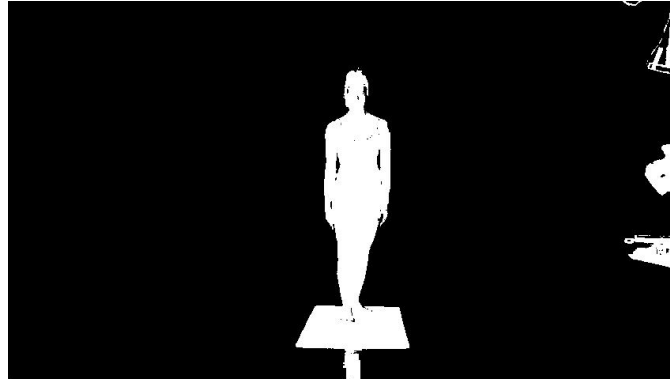
Για να εξασφαλιστεί ότι οι περιοχές των foreground αντικειμένων θα αναγνωριστούν και θα αφαιρεθούν στην συνέχεια, η προτεινόμενη μέθοδος βασίζεται στο γεγονός ότι η έκταση των foreground αντικειμένων θα είναι συγκριτικά μικρότερη των background περιοχών. Αυτή η υπόθεση εάν και εισάγει έναν περιοριστικό παράγοντα στην προτεινόμενη μέθοδο παρόλα αυτά η υπόθεση αυτή ικανοποιείται από πολλά βίντεο πραγματικών συνθηκών όπως θα δούμε στην συνέχεια.

Οπότε η έκταση της συστάδας i (M_{R_i}) ανήκει ή όχι στο foreground εάν:

$$F_2(x, y) = \begin{cases} i & \text{sum}(M_{R_i}(x, y) = '1')/S_{R_i} > T_{prop} \\ 0 & \text{else} \end{cases}, \quad \forall (x, y) \in R_i \quad 3.12$$

Όπου $i = 1, \dots, K$

Οι συστάδες που συμπεριλαμβάνονται στο τελικό αποτέλεσμα θα συμβολίζονται με M_{R_j} , όπου $j = i - n$ και n ο αριθμός των συστάδων που δεν συμπεριλήφθηκαν τελικά στο αποτέλεσμα. Για να έχει αποτέλεσμα η παραπάνω μέθοδος πρέπει $n \neq 0$. Το όριο T_{prop} πρέπει να είναι μικρό (0.04~0.3) έτσι ώστε να μην αποκλειστεί κάποια περιοχή του foreground από το αποτέλεσμα.



Εικόνα 3.11 Αποτέλεσμα προσδιορισμού εν δυνάμει foreground περιοχών (λευκές περιοχές) σύμφωνα με τον αλγόριθμο παραπάνω και $T_{prop} = 0.1$, $K = 6$.

Έχοντας σα δεδομένο την παραπάνω εικόνα η ανακατασκευή του background εφόσον το αντικείμενο δεν κινείται από την αρχική του θέση πρέπει γίνει χρησιμοποιώντας την πληροφορία από τις γειτονικές περιοχές των foreground αντικειμένων:

Έστω L μια ομοιογενής περιοχή. Όλα τα pixel της περιοχής L σύμφωνα με την τμηματοποίηση της εικόνας αποτελούν μέρος της έκτασης της συστάδας j (M_{R_j}).

$$L \in M_{R_j}$$

$$M(L) = \begin{cases} 1 & \forall (x, y) \in L \\ 0 & \text{else} \end{cases}$$

Έχοντας την εικόνα $M(L)$ και εφαρμόζοντας ένα μορφολογικό φίλτρο διαστολής με μορφολογικό δομικό στοιχείο κάποιο τετράγωνο μεγάλης διάστασης (π.χ. 30×30), θα προσδιοριστούν οι γειτονικές περιοχές αυτής της foreground περιοχής (περιοχή L_1) συν όμως και της αρχικής foreground περιοχής L . Για να αποκλειστεί η ίδια η περιοχή L χρησιμοποιείται η λογική πράξη του XOR (περιοχή L_2). Στην συνέχεια επειδή δεν θέλουμε να εισαχθούν στην τελική περιοχή (περιοχή L_3) περιοχές του foreground σύμφωνα με την εικόνα F_2 , εξαιρούνται από αυτήν. Ίδια διαδικασία ακολουθείται και για τις υπόλοιπες ομοιογενείς περιοχές των συστάδων j .

Εξισώσεις :

3.13

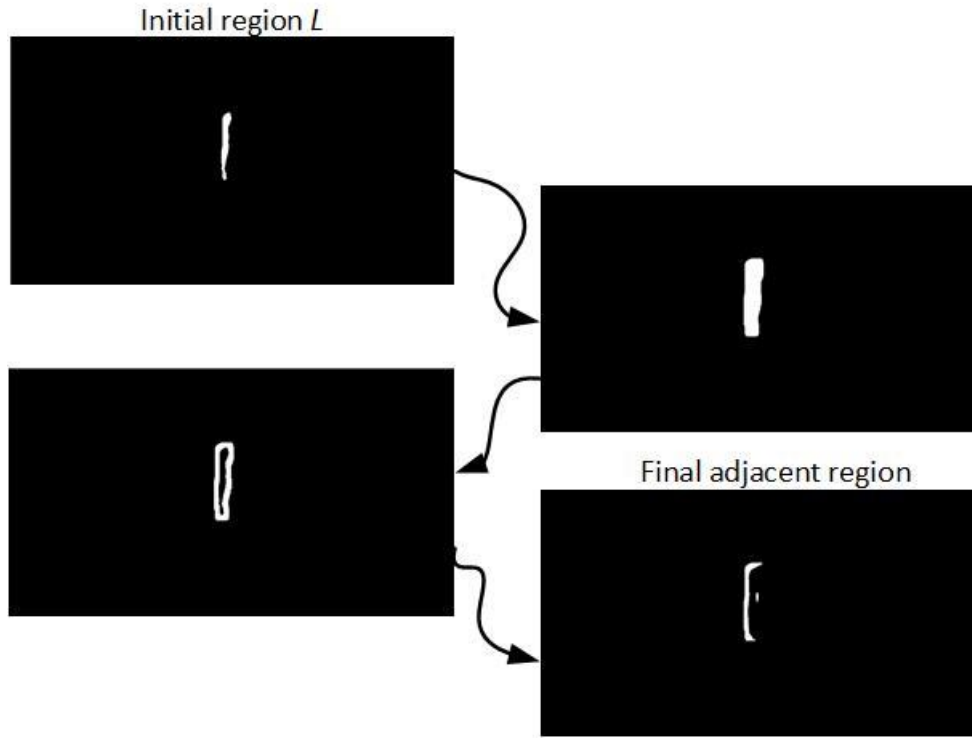
$$se = \begin{bmatrix} 1 & \dots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \dots & 1 \end{bmatrix}$$

$$L_1 = M(L) \oplus se$$

$$L_2 = M(L) XOR L_1$$

$$L_3 = L_2 \notin M_{R_j}$$

Παράδειγμα:



Εικόνα 3.12 Παράδειγμα υπολογισμού γειτονικών περιοχών μιας foreground περιοχής (λευκή περιοχή).

Στην συνέχεια έστω S_{L3_i} η συνολική έκταση για κάθε συστάδα i που εμπεριέχεται στην περιοχή L_3 , το επικρατέστερο χρώμα για την περιοχή L προκύπτει:

$$L = i \mid \arg \max_{i=1, \dots, n} S_{L3_i}$$

Η background εικόνα που θα προκύψει συμβολίζεται ως: background_model₂ (Εικόνα 3.13).

Η ανακατασκευή του background γίνεται στο πρώτο μόνο frame της κανονικής περιόδου. Στα υπόλοιπα frames θα πρέπει η background εικόνα (Εικόνα 3.13) να ανανεωθεί σύμφωνα με τις τιμές της εικόνας *color_image* που έχει προκύψει στην Παράγραφο 3.2.3 και σύμφωνα με ένα σχήμα επιλεκτικής ανανέωσης περιοχών του background. Ο λόγος που η ανανέωση του background γίνεται από την

τμηματοποιημένη εικόνα (*color_image*) και όχι από ένα πραγματικό frame του βίντεο είναι εξαιτίας ότι η ανακατασκευή του background έγινε σε επίπεδο χρωμάτων από το codebook. Το ίδιο θα ισχύει και για την εξαγωγή των foreground αντικειμένων στην συνέχεια όπου θα χρησιμοποιηθεί η *color_image* για κάθε frame και όχι τα πραγματικά χρώματα του frame.



Εικόνα 3.13 Τελικό αποτέλεσμα ανακατασκευής του background από τα χρώματα του codebook χρησιμοποιώντας τον παραπάνω αλγόριθμο. Για καλύτερα αποτελέσματα έχει χρησιμοποιηθεί και ένα τετραγωνικό median φίλτρο μεγάλης διάστασης στο τελικό αποτέλεσμα.

Εξαγωγή Foreground Αντικειμένων

Η ανίχνευση των foreground περιοχών για τα frames μετά την ανακατασκευή του background θα γίνει υπολογίζοντας την Ευκλείδεια απόσταση μεταξύ της *color_image* ($p_{color} = L^*a^*b^*$) και του background ($p_{back} = L^*a^*b^*$) που προκύπτει από την ανακατασκευή του background:

$$\Delta E_{ab}^* = \sqrt{(L_{color}^* - L_{back}^*)^2 + (a_{color}^* - a_{back}^*)^2 + (b_{color}^* - b_{back}^*)^2} \quad 3.14$$

$$foreground_final_mask = \begin{cases} 0 & \Delta E_{ab}^* < T_{background_2}, \\ 1 & else \end{cases}$$

Το όριο $T_{background_2}$ είναι μεγαλύτερο του $T_{background_1}$, καθώς η σύγκριση γίνεται μεταξύ διακριτών τιμών του codebook με αποτέλεσμα να υπάρχει δυνατότητα επιλογής μεγαλύτερου εύρους τιμών χωρίς να επηρεάζει το τελικό αποτέλεσμα. Συνήθως 80~150.

Ανανέωση Background Εικόνας

Η επιλεκτική ανανέωση του background θα γίνει σε επίπεδο περιοχών σύμφωνα με ένα παρόμοιο σχήμα της Παραγράφου 3.3.1. Μια περιοχή $R \in foreground_final_mask$ και ισχύει:

$$M_R = i, \quad \forall (x, y) \in R \ \& \ i \in Codebook, \quad \text{Όπου } i = 1, \dots, K$$

Η περιοχή M_R θα ανανεώνεται εάν επικαλύπτεται σε ένα ποσοστό και πάνω από την αντίστοιχη περιοχή της δυαδική εικόνα F_1 που προκύπτει από την μέθοδο της Παραγράφου 3.2.2, έστω S_{F_1} η έκταση αυτής και S_R αντίστοιχα.

$$Upd(x, y) = \begin{cases} 1 & S_R/S_{F_1} > T_{prop} \\ 0 & else \end{cases}, \quad \forall (x, y) \in R \quad 3.15$$

Το ποσοστιαίο όριο T_{prop} αυτή την φορά θα είναι μεγαλύτερο από ότι κατά την διαδικασία της ανακατασκευής του background. Συνήθως 0.3~0.8

$background_model_2(x, y) =$

$$\begin{cases} (1 - \alpha)background_model_2(x, y) + \alpha color_image_i(x, y) & if Upd(x, y) = '1' \\ do nothing & else \end{cases}, \quad 3.16$$

Όπου α παράγοντας εκμάθησης. Σε σύγκριση με τον παράγοντα εκμάθησης της Παραγράφου 3.2.2, η τιμή του α πρέπει να είναι ίση ή λίγο μεγαλύτερη. Επειδή κατά τη διάρκεια της κανονικής περιόδου η 2^η background εικόνα χρειάζεται κάποια frames μέχρι να προσαρμοστεί στα πραγματικά χρώματα από το codebook, θεωρούμε ότι η κανονική περίοδος ξεκινάει μετά από 10~20 frames της κανονικής περιόδου.

Η εικόνα Upd πολλές φορές περιέχει κάποιες λανθασμένες εκτιμήσεις περιοχών (εξαιτίας του θορύβου), που έχουν χαρακτηριστεί ως foreground ενώ στην πραγματικότητα είναι background περιοχές. Αυτές οι περιοχές δεν θα ανανεωθούν σύμφωνα με το παραπάνω σχήμα ανανέωσης του background, καθώς αποτελούν λανθασμένες foreground περιοχές. Αυτό δεν αποτελεί πρόβλημα για την μέθοδο καθώς μετά την αρχική ανακατασκευή του background οι περιοχές αυτές σε βάθος χρόνου θα ανανεώνονται με τις πραγματικές τιμές του background ακόμα και εάν λόγω θορύβου περάσουν κάποια frames χωρίς ανανέωση.

Συμπεράσματα

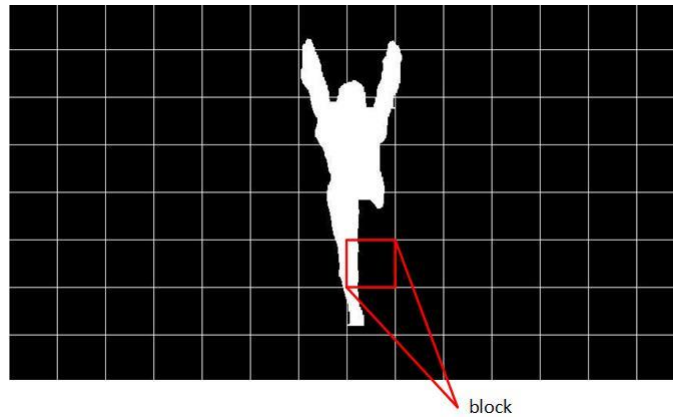
Τι θα γίνει όμως όταν μια περιοχή του foreground λόγω αργής κίνησης εξαφανιστεί ξαφνικά από την εικόνα F_1 που αποτελεί την εκτίμηση για την περιοχή κίνησης? Το αποτέλεσμα θα είναι το χρώμα από το codebook για το foreground αντικειμένου να περάσει στην background εικόνα με αποτέλεσμα την λανθασμένη εξαγωγή των foreground αντικειμένων στην συνέχεια.



Εικόνα 3.14 Παράδειγμα της εικόνας Upd α) την χρονική στιγμή t και την χρονική στιγμή $t+1$ b) μια περιοχή της εξαφανίζεται ξαφνικά λόγω αργής κίνησης του αντικειμένου (κόκκινο πλαίσιο).

3.3.3 Ασαφή Λογική Ανανέωσης της Περιοχής Κίνησης

Για την αποφυγή του δεύτερου περιοριστικού παράγοντα που περιγράφεται παραπάνω απαιτείται η δημιουργία ενός μηχανισμού αντιμετώπισης. Με τον όρο ακαριαία νοείται η απουσία μιας περιοχής σε επόμενο frame δεδομένου ότι η περιοχή αυτή αναγνωρίζεται από την μέθοδο της Παραγράφου 3.2.2 στο προηγούμενο frame (Εικόνα 3.14). Για την αντιμετώπιση του περιορισμού, δημιουργήθηκε ένα πλαίσιο (grid) πάνω στην εικόνα $foreground_mask_1$ και η εικόνα διασπάστηκε σε blocks. Στην συνέχεια συνδυάζοντας δύο πληροφορίες αποφασίζεται η ανανέωση του block με το τρέχον χρονικά frame ή η αντικατάσταση του από το αντίστοιχο block του προηγούμενου χρονικά frame.



Εικόνα 3.15 Παράδειγμα τμηματοποίησης της εικόνας σε blocks μεγέθους 61×60 .

Μαθηματική Αναπαράσταση

Έστω F_1 το πρώτο frame προς επεξεργασία με διαστάσεις $M \times N$. Οι διαστάσεις του κάθε block συμβολίζονται με $b_{size1} \times b_{size2}$. Η εικόνα F_1 χωρίζεται σε $M \times N / b_{size1} \times b_{size2}$ αριθμό blocks που αποθηκεύονται στον πίνακα *Previous_Block*. Ο πίνακας *Previous_Block* έχει διαστάσεις $[b_{size1} \times b_{size2} \times Num_of_blocks]$, οπότε η αναφορά στο block i από τα συνολικά blocks (Num_of_blocks) της εικόνας θα γίνεται ως εξής:

$$Previous_Block_i[1:b_{size1} \times 1:b_{size2} \times i]$$

Η διάσπαση της εικόνας σε blocks και στην συνέχεια η αποθήκευσή τους θα γίνει ξεκινώντας από αριστερά προς τα δεξιά και κατά μήκος του οριζόντιου άξονα.

Επειδή χρειάζεται άλλη μια πληροφορία που είναι απαραίτητη για την διατήρηση και την ανανέωση του πίνακα με τα blocks, απαιτείται στο παραπάνω frame η εύρεση κάποιων χαρακτηριστικών για την περιοχή που αντιπροσωπεύει τον άνθρωπο. Στη συγκεκριμένη υλοποίηση της προτεινόμενης μεθόδου η περιοχή που αντιπροσωπεύει τον άνθρωπο θα είναι αυτή με την μεγαλύτερη έκταση (χρονική στιγμή $t=1$). Εάν στο βίντεο υπήρχε και άλλο ή άλλα άτομα θα έπρεπε με παρόμοιο τρόπο να ληφθούν

υπόψιν. Τα χαρακτηριστικά που χρειάζονται είναι το κέντρο βάρους (Centroid), το οριοθετημένο τετράγωνο (Bounding Box) γύρω από το αντικείμενο του ενδιαφέροντος μας και η συνολική έκταση της περιοχής (Area).

$$[Centroid_I, BoundingBox_I, area_I] = regionprops(F_1)$$

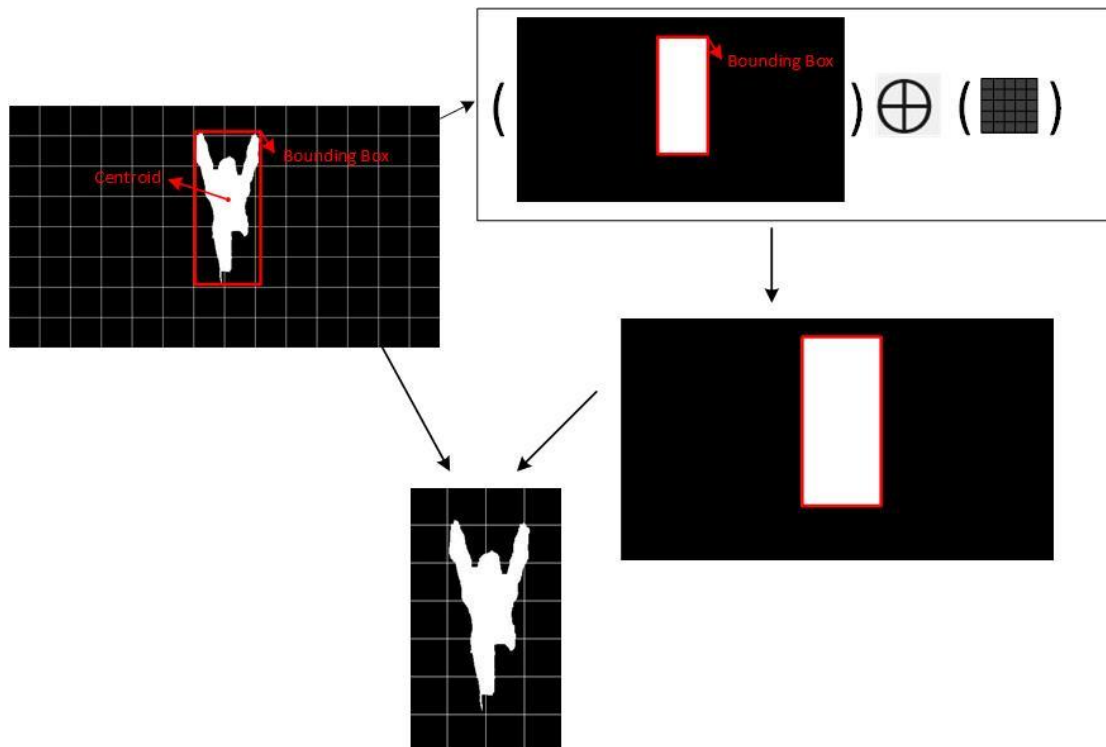
$$[Centroid_{\kappa}, BoundingBox_{\kappa}, area_{\kappa}] = \kappa \mid \arg \min_{i=1, \dots, I} (area_i)$$

Όπου I ο συνολικός αριθμός των αντικειμένων στην δυαδική εικόνα F_1 , και κ το αντικείμενο που αφορά τον άνθρωπο. Πληροφορίες για την συνάρτηση *regionprops* μπορούν να βρεθούν εδώ [129].

Για τα επόμενα frames (χρονική στιγμή $t > 1$) της εικόνας ($F_{2, \dots, n}$) θα δημιουργείται με τον ίδιο τρόπο ο πίνακας με τα blocks με την ονομασία όμως *Current_Block*. Στην συνέχεια θα υπολογίζονται πάλι οι τιμές $[Centroid_I, BoundingBox_I, area_I]$ και το αντικείμενο που θα αφορά τον άνθρωπο θα εντοπίζεται βρίσκοντας την ελάχιστη ευκλείδεια απόσταση μεταξύ του κέντρου βάρους του αντικειμένου κ (πού υπολογίστηκε από το προηγούμενο βήμα) και των αντικειμένων I του τρέχοντος frame.

$$[Centroid_j, BoundingBox_j, area_j] = j \mid \arg \min_{i=1, \dots, I} ((Centroid_{\kappa} - Centroid_i)^2)$$

Ο λόγος που χρησιμοποιείται το κέντρου βάρους και όχι η συνολική έκταση του αντικειμένου είναι διότι εάν στο σκηνικό καταγραφής εισέλθει κάποιο αντικείμενο μεγαλύτερο από το αρχικό τότε η μέθοδος θα αποτύχει.



Εικόνα 3.16 Παράδειγμα χρησιμότητας του Bounding Box και τελικώς τα blocks τα οποία είναι ικανά να ανανεωθούν ή όχι. Τα υπόλοιπα blocks παραμένουν ως έχει.

Η πληροφορία *BoundingBox* (Εικόνα 3.16) του αντικειμένου, θα χρησιμοποιηθεί με το εξής τρόπο: Η ανανέωση της εικόνας (F_2, \dots, n) με τις περιοχές κίνησης πρέπει να γίνει μόνο στις περιοχές όπου πραγματικά υπάρχει άνθρωπος καθώς εάν περιλάμβανε όλη την εικόνα, εξαιτίας του θορύβου μεγάλες περιοχές θα μπορούσαν ακαριαία να αλλάζουν ως προς την ταξινόμηση τους (background, foreground) με αποτέλεσμα η μέθοδος να αποτυγχάνει. Το παραλληλόγραμμο (*BoundingBox*) που σχηματίζεται συν μιας μικρής αύξησης στις διαστάσεις του μέσω μορφολογικού φίλτρου, έτσι ώστε να συμπεριλαμβάνει και περιοχές του foreground που ίσως χαθούν, θα υποδεικνύει στην μέθοδο ποιες περιοχές είναι κατάλληλες προς ανανέωση. Η διάσταση του τετράγωνου δομικού στοιχείου που χρησιμοποιείται για την μορφολογική πράξη της διαστολής υποδηλώνει και το περιθώριο γειννίας με τα άλλα blocks. Συνήθως 30×30~50×50.

Έστω *Current_Block_{upd}* τα blocks τα οποία είναι για περαιτέρω διερεύνηση. Εάν ένα block έχει αλλάξει κατά πολύ σε σχέση με το αντίστοιχο του προηγούμενου frame και επίσης η έκταση του μειώθηκε, τότε κατά πάσα πιθανότητα το block αυτό πρέπει να αλλαχθεί με το προηγούμενο χρονικά. Εάν δεν ισχύει κάτι από αυτά παραπάνω τότε το block παραμένει ως έχει.

$$\text{sum}(\text{and}(\text{Current_Block}_i, \text{Previous_Block}_i) = '1')$$

$$\text{sum}(\text{Current_Block}_i) = '1'$$

$$\text{sum}(\text{Previous_Block}_i) = '1'$$

Η ομοιότητα μετριέται με τη λογική πράξη του AND μεταξύ του *Current_Block_i* και του *Previous_Block_i* και στην συνέχεια την μέτρηση των pixels που έχουν τιμή ίση με true. Επίσης υπολογίζεται ο αριθμός των pixels του block που είναι χαρακτηρισμένα ως foreground και για τα δύο χρονικά διαφορετικά blocks.

$$\frac{\text{sum}(\text{and}(\text{Current_Block}_i, \text{Previous_Block}_i) = '1')}{\text{sum}(\text{Current_Block}_i) = '1'} > T_{cov}$$

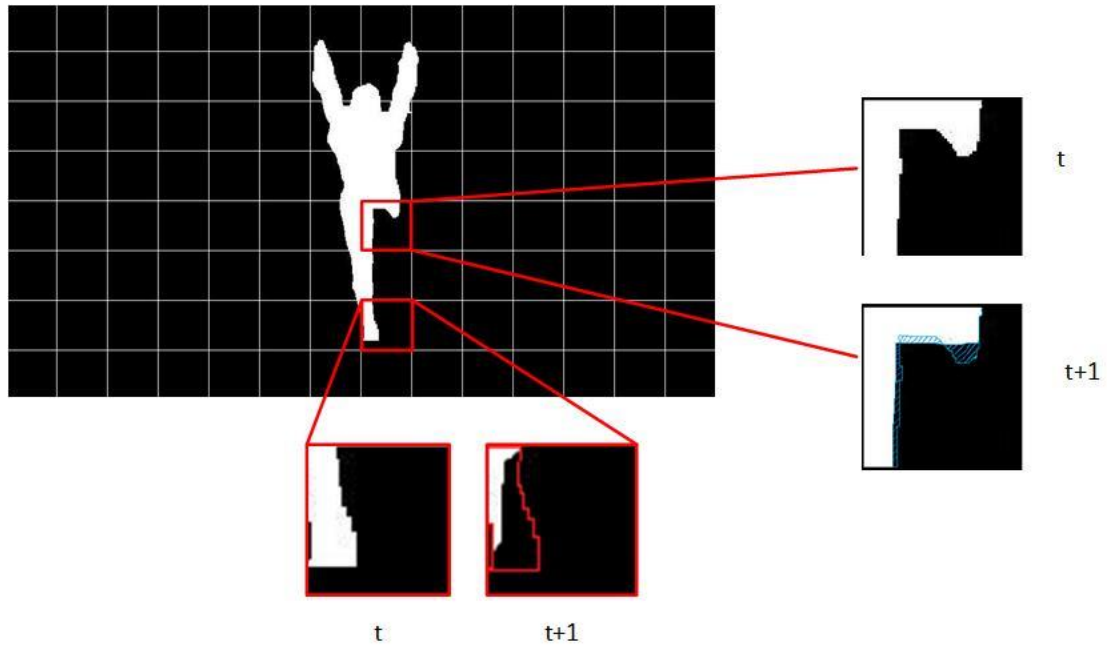
$$\text{sum}(\text{Previous_Block}_i) = '1' > \text{sum}(\text{Current_Block}_i) = '1'$$

Το block το οποίο η έκταση τους μειώθηκε μεταξύ δύο διαδοχικών frames πρέπει:

Εάν ο λόγος παραπάνω είναι μικρότερος ενός ορίου τότε σημαίνει ότι μια περιοχή του foreground εξαφανίστηκε ακαριαία και πρέπει να αντικατασταθεί από το προηγούμενο χρονικά block. Εάν όχι τότε απλά σημαίνει ότι το αντικείμενο απλά μετατοπίστηκε και το block δεν χρειάζεται να αντικατασταθεί με το προηγούμενο χρονικά. Το όριο είναι συνήθως τα 2/3 της έκτασης του block.

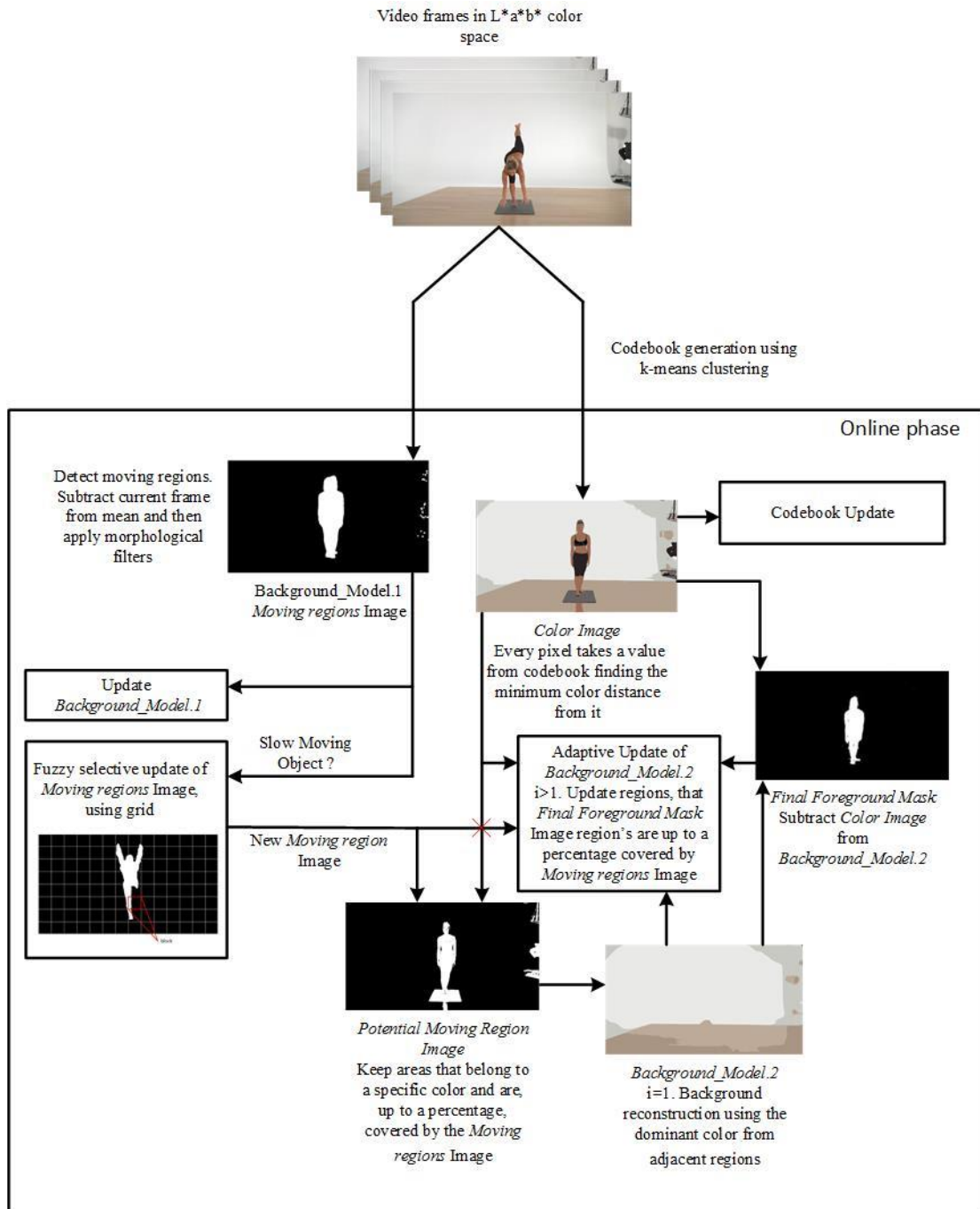
Συμπεράσματα

Όπως προαναφέρεται, η συγκεκριμένη μέθοδος εφαρμόζεται σε περιβάλλοντα όπου ένα αντικείμενο μετακινείται ελάχιστα από την αρχική του θέση. Η προτεινόμενη μέθοδος μπορεί να αντιμετωπίσει ένα αντικείμενο. Για περισσότερα από ένα αντικείμενα πρέπει να ληφθεί υπόψιν διαφορετική υλοποίηση που όμως δεν είναι διαφορετική σε λογική από την υπάρχουσα. Τέλος, ο παραπάνω μηχανισμός για λόγους χρονικής καθυστέρησης αλλά και μη αποτελεσματικότητας μπορεί να παραληφθεί σε περιβάλλοντα παρακολούθησης κυκλοφορίας.



Εικόνα 3.17 Δεξιά φαίνονται τα pixels του block που αλλάζουν εξαιτίας της κίνησης του αντικειμένου (με μπλε περίγραμμα), ενώ σε κόκκινο πλαίσιο τα pixels του block του οποίο εξαφανίστηκαν λόγω αστοχίας της μεθόδου της Παραγράφου 1.

3.3.4 Γραφικό Σχήμα Προτεινόμενης Μεθόδου (Flow Chart)



4. Κεφάλαιο – Αποτελέσματα Υλοποίησης

4.1 Μέτρηση Αποτελεσματικότητας – Απόδοσης Αλγορίθμου

Η υλοποίηση της προτεινόμενης μεθόδου έγινε στο λογισμικό της Matlab© αποκλειστικά. Η μέτρηση της αποτελεσματικότητας μιας μεθόδου μπορεί να γίνει σε αρκετούς τομείς. Παρακάτω αναφερόμαστε στον τομέα μέτρησης της ικανότητας του αλγορίθμου να κατηγοριοποιεί σωστά τα pixels στις δύο κατηγορίες (background-foreground), με όσο το δυνατόν λιγότερες λανθασμένες εκτιμήσεις.

Για την εκτίμηση της ανιχνευτικής ικανότητας έχουν προταθεί διάφορες προσεγγίσεις. Στην περίπτωση της εξαγωγής των foreground αντικειμένων σε ένα βίντεο, ο αλγόριθμός έχει να επιλέξει ανάμεσα σε background και foreground για κάθε pixel ενός frame. Η μέτρηση της αποτελεσματικότητας θα αφορά το κατά πόσο ο αλγόριθμος ταξινομεί τα pixel στις σωστές κατηγορίες.

Στην προτεινόμενη μέθοδο υπολογίζονται οι τιμές true positive rate (*TPR*) και false positive rate (*FPR*). Το *FPR* εκφράζει το ποσοστό επί των συνολικών background pixels, των pixels που έχουν χαρακτηριστεί λανθασμένα ως foreground αντικείμενα από τον αλγόριθμο, ενώ το *TPR* εκφράζει το ποσοστό επί των συνολικών foreground pixels, των pixels που έχουν ορθά χαρακτηριστεί ως foreground. Οι τιμές *FPR* και *TPR* είναι γνωστές και ως sensitivity, specificity αντίστοιχα και για τον υπολογισμό τους είναι απαραίτητη η μέτρηση των τιμών true positive (TP), true negative (TN), false positive (FP), false negative (FN). Οι ποσότητες *TPR*, *FPR* θα υπολογίζονται για κάθε frame του βίντεο και θα αναπαριστώνται γραφικά σε δυο γραφικές παραστάσεις. Τέλος, μια σύντομη αναφορά θα γίνει στον τομέα της χρονικής πολυπλοκότητας του αλγορίθμου.

Ο όρος Ground Truth που συχνά αναφέρεται στη μέτρηση της αποτελεσματικότητας νοείται θεωρητικά ως η πλήρως σωστή ταξινόμηση των δεδομένων στις περιοχές ταξινόμησης χωρίς καμία λάθος ταξινόμηση.

		GROUND TRUTH	
		Presence	Absence
PREDICTED	Presence	True Positive (TP)	False Positive (FP)
	Absence	False Negative (FN)	True Negative (TN)

$$FPR = \frac{FP}{FP + TN} ; TPR = \frac{TP}{TP + FN}$$

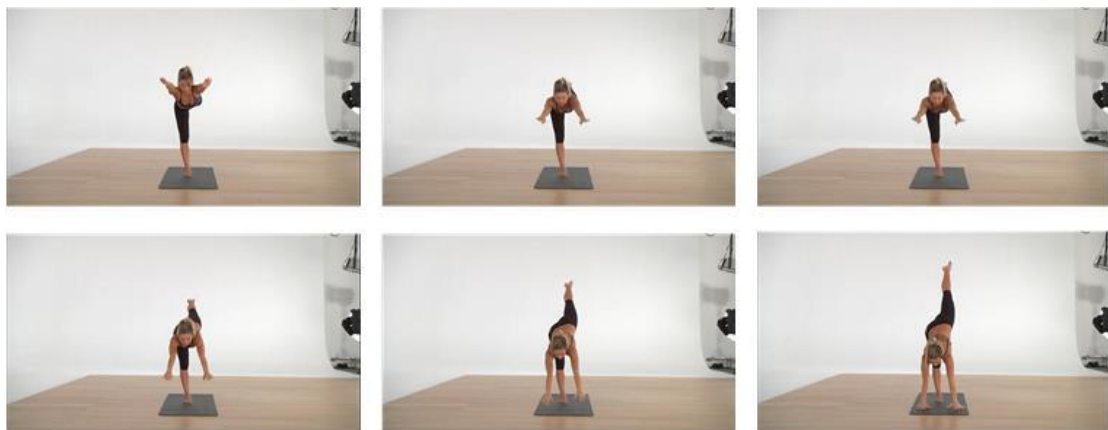
Εικόνα 4.1 Πίνακας Confusion για τις τιμές TP, TN, FP, FN και FPR TPR.

4.2 Δεδομένα Προς Επεξεργασία

Για τον έλεγχο της απόδοσης της προτεινόμενης μεθόδου απαιτείται ένα σύνολο δεδομένων από διαφορετικά περιβάλλοντα. Στόχος της προτεινόμενης εργασίας ήταν η εξαγωγή foreground αντικειμένων τα οποία κινούνται σχετικά αργά, με αποτέλεσμα ομοιογενείς περιοχές του foreground αντικειμένου να μην είναι ανιχνεύσιμες από την μέθοδο. Αυτόν τον περιοριστικό παράγοντα έχουν και τα βίντεο εκγύμνασης και συγκεκριμένα τα βίντεο με τις ασκήσεις yoga όπου η κίνηση του ανθρώπου είναι σχετικά αργή. Εντούτοις, η προτεινόμενη μέθοδος εφαρμόστηκε και σε άλλα βίντεο με διαφορετικές συνθήκες.

4.2.1 Βίντεο με Ασκήσεις Yoga (60 frames, resolution 854×480, 25 fps)

Βίντεο με αργή κίνηση ενός μόνο ατόμου εκτελώντας ασκήσεις yoga: Για το συγκεκριμένο βίντεο είναι ενεργοποιημένη η τεχνική της Ασαφούς Λογικής Ανανέωσης της Περιοχής Κίνησης (Παράγραφος 3.3.3) όπως περιγράφεται παραπάνω, στα επόμενα βίντεο δεν χρησιμοποιείται ο μηχανισμός.



Εικόνα 4.2 Ενδεικτικά frames από το βίντεο.

Παράμετροι Μεθόδου

Αριθμός Χρωμάτων K: 6

Εκπαιδευτική Περίοδος: 20 frames

Παράγοντας εκμάθησης 1^{ου} Μοντέλου (α_1): 0.06

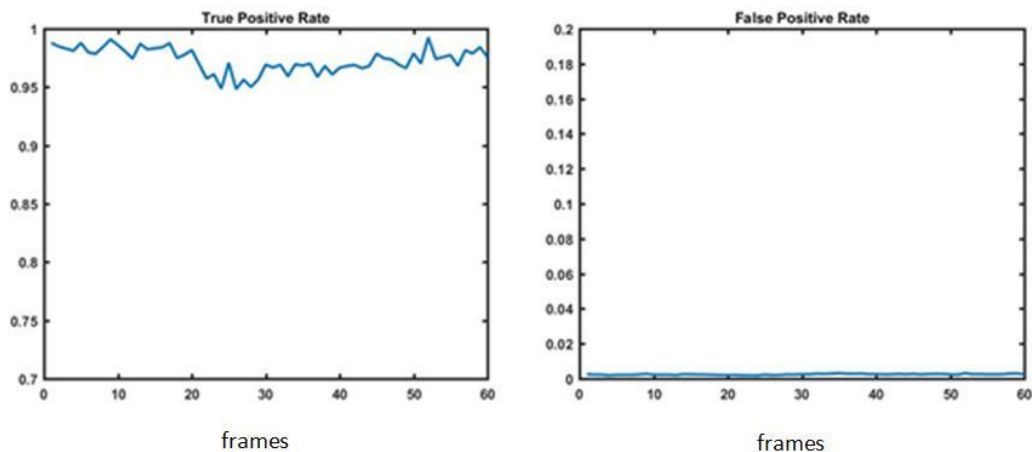
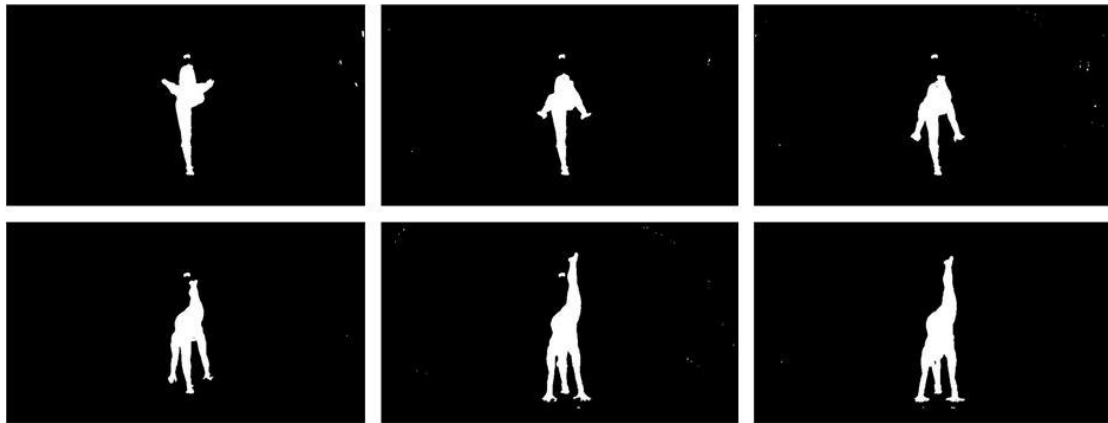
Παράγοντας εκμάθησης 2^{ου} Μοντέλου (α_2): 0.13

Όριο Εξαγωγής Foreground αντικειμένων 1^{ου} Μοντέλου ($T_{\text{background}_1}$): 10

Όριο Εξαγωγής Foreground αντικειμένων 2^{ου} Μοντέλου ($T_{\text{background}_2}$): 120

Όριο Επικάλυψης Περιοχών (Κατά την ανακατασκευή του Background): 0.1

Όριο Επικάλυψης Περιοχών (Κατά την ανανέωση του background): 0.5



Εικόνα 4.3 Αποτελέσματα χρήσης μεθόδου και μέτρηση αποτελεσματικότητας.

4.2.2 Βίντεο από Κάμερα Παρακολούθησης σε Σταθμό του Μετρό (i-LIDS dataset [130]) (190 frames, resolution 720×576, 30 fps)

Εξωτερικό περιβάλλον με δύσκολες συνθήκες φωτισμού και προσθήκη κατά την διάρκεια του βίντεο δύο ακόμη ατόμων: Κατά τη διάρκεια του βίντεο το άτομο επικαλύπτεται εν μέρει με ιδίου χρώματος αντικείμενο του background (camouflage).

Παράμετροι Μεθόδου

Αριθμός Χρωμάτων K: 10

Εκπαιδευτική Περίοδος: 20 frames

Παράγοντας εκμάθησης 1^{ου} Μοντέλου (α_1): 0.1

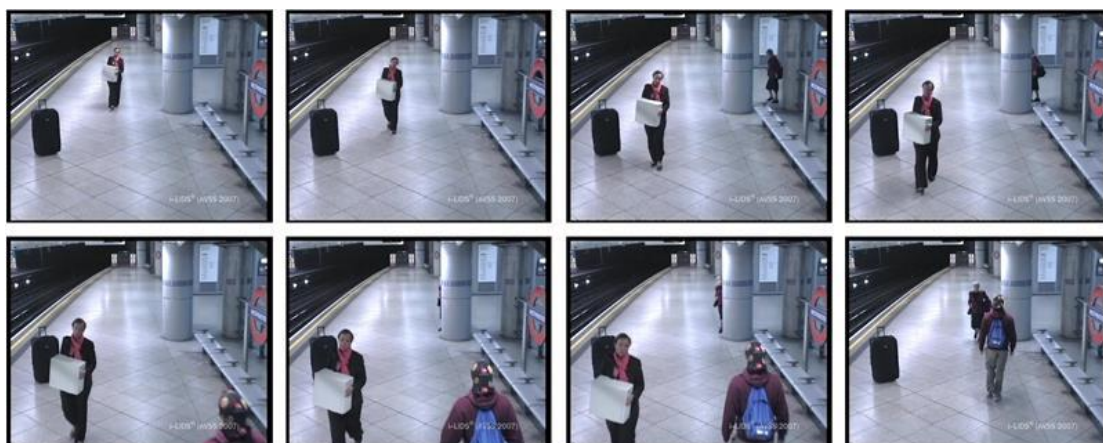
Παράγοντας εκμάθησης 2^{ου} Μοντέλου (α_2): 0.2

Όριο Εξαγωγής Foreground αντικειμένων 1^{ου} Μοντέλου ($T_{background1}$): 20

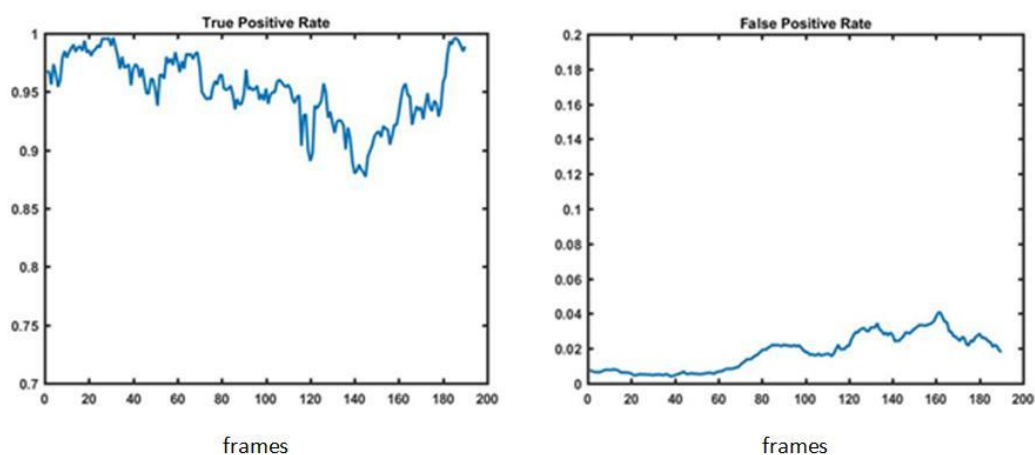
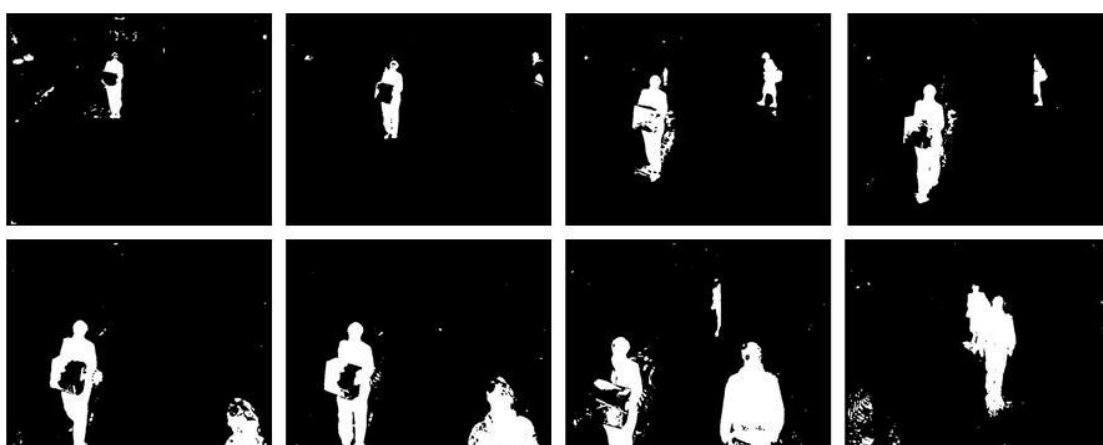
Όριο Εξαγωγής Foreground αντικειμένων 2^{ου} Μοντέλου ($T_{background2}$): 150

Όριο Επικάλυψης Περιοχών (Κατά την ανακατασκευή του Background): 0.04

Όριο Επικάλυψης Περιοχών (Κατά την ανανέωση του background): 0.8



Εικόνα 4.5 Ενδεικτικά frames από το βίντεο.

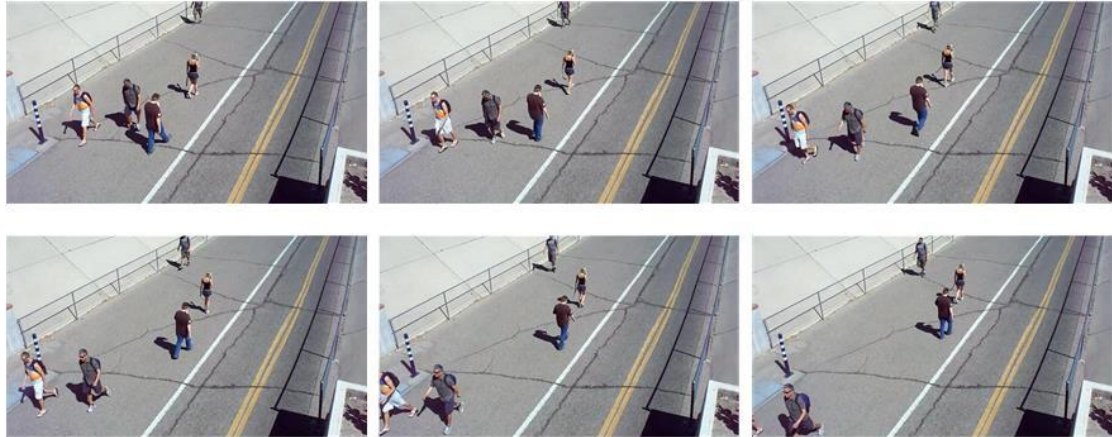


Εικόνα 4.4 Αποτελέσματα χρήσης μεθόδου και μέτρηση αποτελεσματικότητας.

Η μικρή πτώση (TPR) στην απόδοση του αλγορίθμου από το 120^ο έως 160^ο frame ευθύνεται στην επικάλυψη του ατόμου εν μέρει με ιδίου χρώματος αντικείμενο του background (camouflage). Πιθανόν λανθασμένα ως foreground pixels είναι αποτέλεσμα της σκιάς γύρω από το αντικείμενο με αποτέλεσμα η background εικόνα να μολύνεται.

4.2.3 Βίντεο από Κάμερα Παρακολούθησης σε Δρόμο 1 ([131]) (50 frames, resolution 854×480, 24 fps)

Εξωτερικό περιβάλλον με σκιά και αρκετά άτομα κατά την διαδικασία αρχικοποίησης του background: Η σκιά επειδή έχει μεγάλη χρωματική διαφορά στις ground truth εικόνες θεωρείται ως foreground



Εικόνα 4.6 Ενδεικτικά frames από το βίντεο.

Παράμετροι Μεθόδου

Αριθμός Χρωμάτων K: 6

Training Period: 10 frames

Παράγοντας εκμάθησης 1^{ου} Μοντέλου (α_1): 0.1

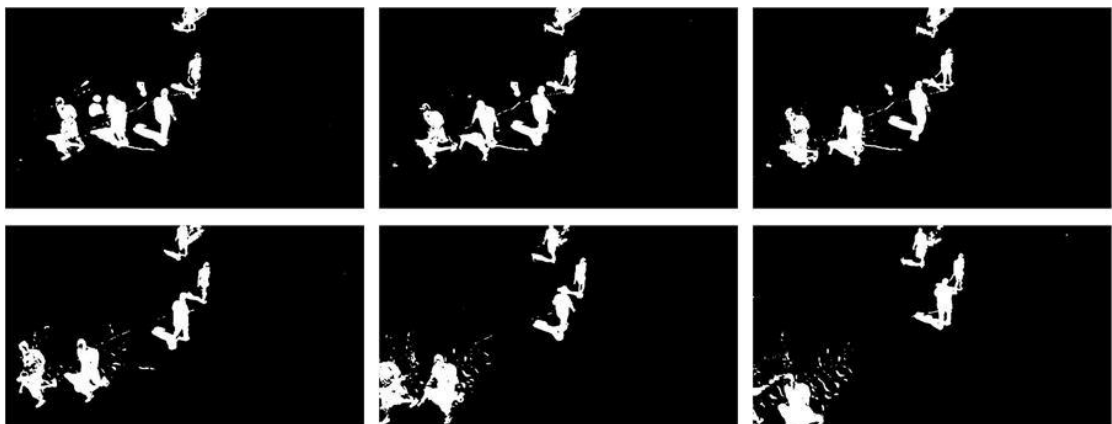
Παράγοντας εκμάθησης 2^{ου} Μοντέλου (α_2): 0.2

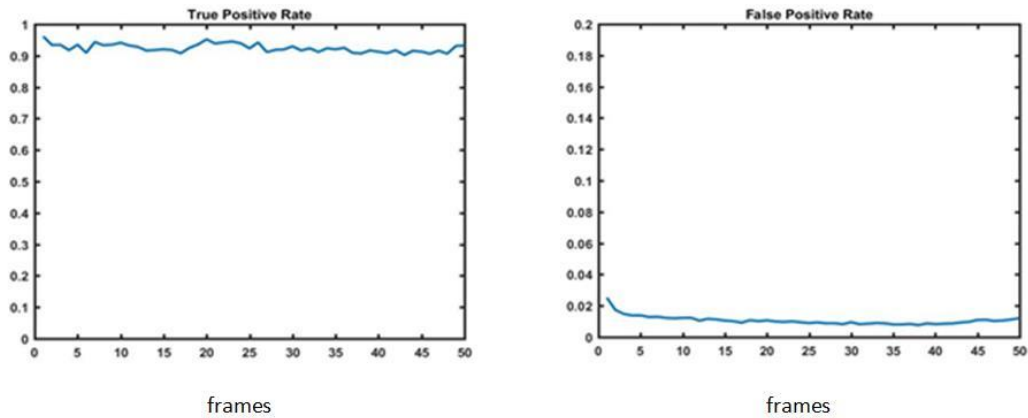
Όριο Εξαγωγής Foreground αντικειμένων 1^{ου} Μοντέλου ($t_{background1}$): 20

Όριο Εξαγωγής Foreground αντικειμένων 2^{ου} Μοντέλου ($t_{background2}$): 100

Όριο Επικάλυψης Περιοχών (Κατά την ανακατασκευή του Background): 0.2

Όριο Επικάλυψης Περιοχών (Κατά την ανανέωση του background): 0.7





Εικόνα 4.7 Αποτελέσματα χρήσης μεθόδου και μέτρηση αποτελεσματικότητας.

4.2.4 Βίντεο από Κάμερα Παρακολούθησης σε Δρόμο 2 ([132]) (60 frames, resolution 1020×580, 30 fps)

Εξωτερικό περιβάλλον με αρκετά άτομα και χρώμα περιοχής foreground περίπου ίδιο με εκείνου του background: Το χρώμα παντελονιού του ενός ατόμου είναι ίδιο με το χρώμα του background



Εικόνα 4.8 Ενδεικτικά frames από το βίντεο.

Παράμετροι Μεθόδου

Αριθμός Χρωμάτων K: 10

Training Period: 20 frames

Παράγοντας εκμάθησης 1^{ου} Μοντέλου (α_1): 0.1

Παράγοντας εκμάθησης 2^{ου} Μοντέλου (α_2): 0.2

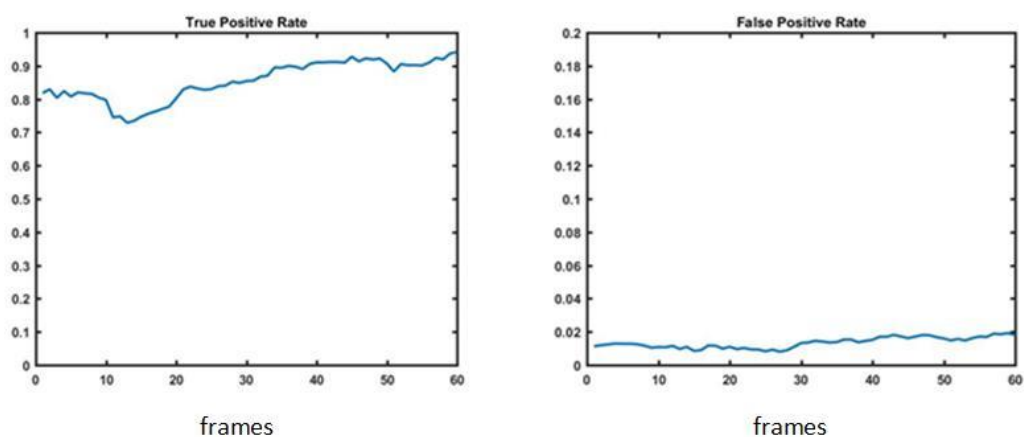
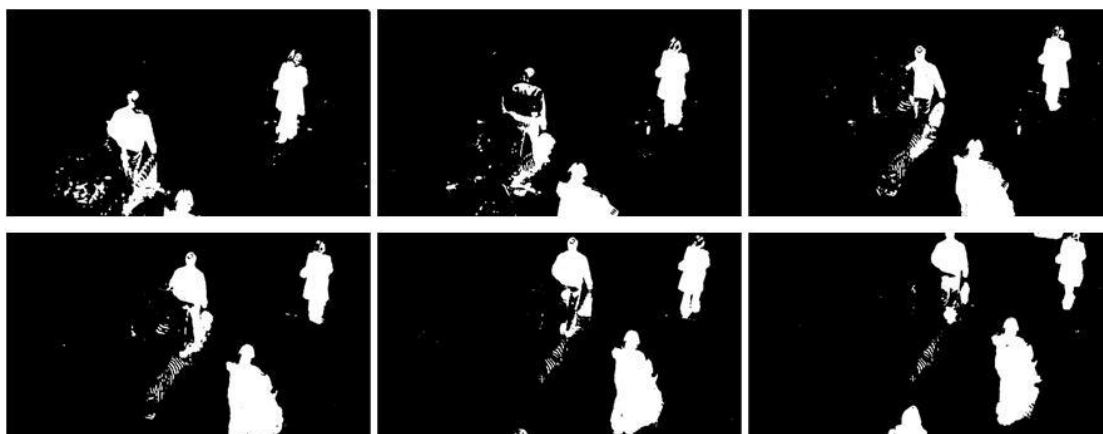
Όριο Εξαγωγής Foreground αντικειμένων 1^{ου} Μοντέλου ($T_{\text{background}_1}$): 10

Όριο Εξαγωγής Foreground αντικειμένων 2^{ου} Μοντέλου ($T_{\text{background}_2}$): 100

Όριο Επικάλυψης Περιοχών (Κατά την ανακατασκευή του Background): 0.2

Όριο Επικάλυψης Περιοχών (Κατά την ανανέωση του background): 0.7

Αποτέλεσμα

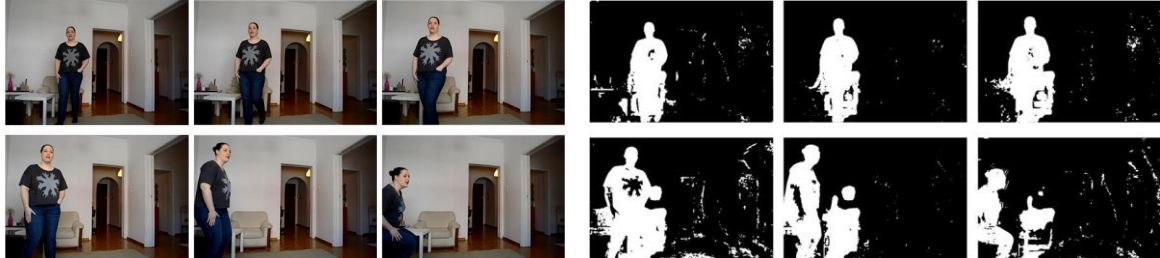


Εικόνα 4.9 Αποτελέσματα χρήσης μεθόδου και μέτρηση αποτελεσματικότητας.

Η μέθοδος είναι λιγότερη αποτελεσματική όταν ο άνθρωπος πλησιάζει το σκούρο αντικείμενο, λόγω ίδιου χρώματος background ένα μέρος του ανθρώπου εξαφανίζεται (camouflage). Επίσης το παντελόνι του ατόμου έχει περίπου ίδιο χρώμα με το background με αποτέλεσμα την λανθασμένη ανίχνευση του.

4.2.5 Βίντεο από Εσωτερικό Περιβάλλον με Ανομοιόμορφο Background

Λιγότερο Αποτελεσματική η μέθοδος σε βίντεο με πολλά background αντικείμενα και δύσκολο φωτισμό.



Εικόνα 4.10 Ενδεικτικά frames βίντεο και αποτελέσματα χρήσης μεθόδου .

4.3 Χρόνοι Εκτέλεσης Αλγορίθμων

Χρησιμοποιώντας τρία παραπάνω βίντεο, τριών διαφορετικών διαστάσεων μετρήσαμε χρόνους εκτέλεσης του αλγορίθμου. Τα στάδια του αλγορίθμου φαίνονται παρακάτω καθώς και οι χρόνοι εκτέλεσης δίπλα, στάδια του αλγορίθμου που δεν είναι σημαντικά παραλήφθηκαν.

Calibration phase (‘Pay’ the cost only once) sec/frame				Pc Specifications	
	Indoor (640×424)	Yoga (854×480)	Outdoor (1020×580)		
<i>K-means Clustering</i>	12	8	22		
<i>Background Reconstruction(1st frame)</i>	7	4	50		
Online phase msec/frame					
<i>Calculate Background model1</i>	80	110	180	<i>Operating System</i>	Windows 7 Ultimate
<i>Calculate Background model2</i>	1400	1200	2000	<i>CPU</i>	Intel i5-4460 3.2GHz(4 cores)
<i>Update Codebook</i>	110	100	300	<i>RAM</i>	8GB
<i>Background Selectivity Update</i>	110	100	2000		
<i>Fuzzy Selectivity Update (optional)</i>	-	90	-		

Για το βίντεο με τις ασκήσεις yoga η αρχική εικόνα τμηματοποιήθηκε σε $K=5$ βασικές περιοχές, ενώ για τα άλλα δύο βίντεο το K ήταν ίσο με 10. Είναι λογικό ότι για μεγαλύτερο αριθμό K ο αλγόριθμος γίνεται πιο αργός κατά την διαδικασία του K -means clustering. Για την μέθοδο της ανακατασκευής του background τόσο οι διαστάσεις της εικόνας όσο και ο αριθμός των ομάδων K παίζει σημαντικό ρόλο στην χρονική καθυστέρηση του αλγορίθμου, ειδικά στο βίντεο (outdoor) η ανακατασκευή του background είναι χρονικά ασύμφορη.

Κατά την διάρκεια της online λειτουργίας τα δύο βίντεο (indoor, yoga) παρουσιάζουν περίπου την ίδια χρονική καθυστέρηση, με μια μικρή διαφορά να οφείλεται στον αριθμό K . Στο τελευταίο κατά σειρά βίντεο η χρονική καθυστέρηση αυξάνεται ραγδαία κατά την επιλεκτική ανανέωση του background πράγμα που οφείλεται στην τμηματοποίηση της εικόνας σε πολύ μικρές περιοχές και την επεξεργασία αυτών.

5. Κεφάλαιο - Συμπεράσματα

Η προτεινόμενη μέθοδος υλοποιήθηκε λαμβάνοντας κυρίως υπόψιν τους δυο βασικούς περιοριστικούς παράγοντες που υπάρχουν σε ένα βίντεο με ασκήσεις γυμναστικής: αργή κίνηση foreground αντικειμένου και παρουσία του foreground αντικειμένου κατά την διάρκεια της εκπαιδευτικής περιόδου. Η προτεινόμενη μέθοδος σε αντίθεση με τις παραδοσιακές τεχνικές εξαγωγής της σιλουέτας ενός ατόμου αντιμετωπίζει επαρκώς τους περιορισμούς αυτούς. Από τα αποτελέσματα του Κεφαλαίου 4 παρατηρούνται τα εξής:

- Εάν το background ενός βίντεο είναι σχετικά ομοιόμορφο χωρίς πολλά αντικείμενα και κατ' επέκταση χρώματα, τότε η μέθοδος εκτός από βέλτιστη είναι και υπολογιστικά υποφερτή καθώς τα στάδια της ανακατασκευής και ανανέωσης του background εκτελούνται πιο γρήγορα εξαιτίας των λιγότερων περιοχών που δέχεται ως είσοδο ο αλγόριθμος παραπάνω (Εικόνα 4.3).
- Εάν το background ενός βίντεο δεν είναι ομοιόμορφο τότε η πιθανή μετακίνηση του αντικειμένου από την αρχική του θέση, θα έχει ως αποτέλεσμα την λανθασμένη ανίχνευση της αρχικής του θέσης ως foreground (Εικόνα 4.10). Αυτό συμβαίνει εξαιτίας της λανθασμένης αρχικής ανακατασκευής του background. Λόγω της προσαρμοστικότητας όμως του background οι περιοχές εκείνες εν τέλει θα συγκλίνουν στα πραγματικά δεδομένα με την πάροδο κάποιων λίγων frames.
- Η δημιουργία και η συντήρηση ενός codebook για τις ομοιόμορφες περιοχές της εικόνα προσφέρει αρκετά σημαντικά πλεονεκτήματα. Με την κατηγοριοποίηση των pixels σε βασικές ομάδες (clusters) δημιουργείται ένα πλαίσιο αντιμετώπισης του προβλήματος όπου το πρόβλημα οριοθετείτε σε διακριτά στοιχεία. Συν της άλλης, η τελική εξαγωγή της σιλουέτας του ατόμου γίνεται χρησιμοποιώντας την background εικόνα που έχει δημιουργηθεί από τα χρώματα του

codebook. Αυτό έχει ως αποτέλεσμα την επιλογή από τον χρήστη, ορίων μεγαλύτερου εύρους χωρίς να επηρεαστεί το τελικό αποτέλεσμα.

- Εξαιτίας της χρησιμοποίησης του codebook σε βίντεο από εσωτερικά περιβάλλοντα, παρατηρείται η εξάλειψη των ήπιων σκιών. Ήπιες σκιές από τα foreground αντικείμενα είναι εκείνες που δεν αλλάζουν δραματικά το χρώμα των background περιοχών.
- Η ανανέωση του codebook αλλά και των δύο background εικόνων γίνεται με τέτοιο τρόπο έτσι ώστε οποιεσδήποτε αλλαγή από την κίνηση αντικειμένων από και προς το σκηνικό καταγραφής, να προσαρμόζεται στα πραγματικά δεδομένα. Για το λόγο αυτό ο ρυθμός ανανέωσης των background εικόνων πρέπει να επιλεγεί κατάλληλα από τον χρήστη.
- Μικρές περιοχές λανθασμένα αναγνωρισμένες ως foreground γύρω από τα κινούμενα αντικείμενα, οφείλονται στην δυαδική εικόνα που σχηματίζεται και αφορά την επιλεκτική ανανέωση της 2^η Background εικόνας. Αν και το πρόβλημα δεν είναι έντονο, μια ενδεχόμενη λύση (χρησιμοποιώντας την δυαδική εικόνα της Παραγράφου 3.2.2), θα προκαλούσε περισσότερα false positive αποτελέσματα.
- Πρόβλημα καμουφλάζ. Όταν υπάρχει επικάλυψη περιοχών του foreground και του background και ο αλγόριθμος που δημιουργεί την τμηματοποιημένη εικόνα, αποδώσει το ίδιο χρώμα από το codebook στις περιοχές αυτές τότε η ανίχνευση του foreground αποτυγχάνει.

Τέλος, στα πλαίσια πιθανής μελλοντικής βελτιστοποίησης της προτεινόμενης μεθόδου, θα μπορούσαν να διερευνηθούν και άλλοι αλγόριθμοι που αφορούν την αρχική ανακατασκευή της background εικόνας, έτσι ώστε στην συνέχεια να προκύπτει η εξαγωγή της σιλουέτας αποτελεσματικότερα.

6. Βιβλιογραφία

- [1] T. Bouwmans, F. Porikli, B. Höferlin and A. Vacavant, Background Modeling and Foreground Detection for Video Surveillance, vol. Chapter 1, New York: Chapman and Hall/CRC, 2014.
- [2] N. Joubert, "Background modelling and subtraction for object detection in video," 2009. [Online]. Available: <http://appliedmaths.sun.ac.za/~wbrink/students/NJoubert2009.pdf>. [Accessed Jan 2018].
- [3] Y. Dedeoğlu, B. U. Töreyn, U. Güdükbay and A. Enis Çetin, "Silhouette-Based method for object classification and human action recognition in video," *ECCV'06 Proceedings of the 2006 international conference on Computer Vision in Human-Computer*, pp. 64-77, May 2006.
- [4] B. Lee and M. Hedley, "Image & Vision Computing New Zealand (IVCNZ '02)," pp. 315-320, 2002.
- [5] McFarlane N. and Schofield C., "Segmentation and tracking of piglets in images," *BMVA 1995*, pp. 187-320, 1995.
- [6] S. Sulaiman, A. Hussain, S. Samad and M. Mustafa, "Human Silhouette Extraction Using Background Modeling and Subtraction Techniques," *Information Technology Journal*, pp. 155-159, 2008.
- [7] Q. Zang and R. Klette, "Evaluation of an adaptive composite Gaussian model in video surveillance," Aug 2002.
- [8] B. White and M. Shah, "Automatically tuning background subtraction parameters using particle swarm optimization.," *IEEE International Conference on Multimedia and Expo, ICME 2007*, pp. 1826-1829, 2007.
- [9] M. Harville, G. Gordon and J. Woodfill, "Foreground segmentation using adaptive mixture models in color and depth," in *International Workshop on Detection and Recognition of Events in Video*, Jul 2001.
- [10] E. Baf, T. Bouwmans and B. Vachon, "A fuzzy approach for background subtraction," *International Conference on Image Processing ICIP 2008*, pp. 2648-2651, Oct 2008.
- [11] M. Hazewinkel, "Mode, Encyclopedia of Mathematics," Springer Science+Business Media B.V. / Kluwer Academic Publishers, 1994. [Online]. Available: <https://www.encyclopediaofmath.org/index.php/Mode>. [Accessed Jun 2018].
- [12] J. Zheng, Y. Wang, N. Nihan and E. Hallenbeck, "Extracting Roadway Background Image: A mode based approach," *Journal of Transportation Research Report*, no. No 1944, pp. 82-88, Mar 2006.

- [13] M. Hazewinkel, "Moment, Encyclopedia of Mathematics," Springer Science+Business Media B.V. / Kluwer Academic Publishers, 1994. [Online]. Available: <http://www.encyclopediaofmath.org/index.php?title=Moment&oldid=25957>. [Accessed Jun 2018].
- [14] C. Wren , A. Azarbayejani, T. Darrell and A. Pentland, "Pfinder: Real-Time Tracking of the Human Body," *IEEE Transactions on Pattern Analysis*, vol. Volume 19, no. No. 7, pp. 780-785, Jul 1997.
- [15] T. Zhen, M. Zhenjiang and W. Yanli, "Background Subtraction Using Running Gaussian Average and Frame Difference," *Entertainment Computing – ICEC 2007*, pp. 411-414, Sep 2007.
- [16] C. Martin, «BACKGROUND SUBTRACTION USING RUNNING GAUSSIAN AVERAGE: A COLOR,» Seminar aus Bildverarbeitung und Mustererkennung, 2014.
- [17] O. B. Devi, N. S. Paul and Y. J. Singh, "ROBUST STATISTICAL APPROACH FOR EXTRACTION OF MOVING HUMAN SILHOUETTES FROM VIDEOS," *International Journal on Information Theory (IJIT)*, pp. Vol. 3, No. 3, Jul 2014.
- [18] M. Zhao, N. Li and C. Chen, "Robust automatic video object segmentation technique," in *International Conference on Image Processing*, Sep 2007.
- [19] R. C. Gonzalez and R. E. Woods, *Digital Image Processing (3rd Edition)*, Pearson Education, 2008.
- [20] N. Friedman and S. Russell, "Image Segmentation in Video Sequences: A Probabilistic Approach," *UAI 1997*, pp. 175-181, 1997.
- [21] C. Stauffer και Grimson W., «Adaptive background mixture models for real-time tracking,» *CVPR 1999*, pp. 246-252, 1999.
- [22] S. Atev, O. Masoud and N. Papanikolopoulos, "Practical Mixtures of Gaussians with Brightness Monitoring," *ITS 2004*, pp. 423-428, 2004.
- [23] Q. Zang and R. Klette, "Parameter Analysis for Mixture of Gaussians," CITR Technical Report 188, Auckland University, 2006.
- [24] T. Bouwmans , F. Baf and B. Vachon, "Background Modeling using Mixture of Gaussians for Foreground Detection – A survey," *Recent Patents on Computer Science*, vol. Vol 1, no. No 3, pp. 219-237, Nov 2008.
- [25] "Improved adaptive Gaussian mixture model for background subtraction,," *International Conference Pattern Recognition*, vol. Vol 2, pp. 28-31, 2004.
- [26] J. Cheng, J. Yang, Y. Zhou and Y. Cui, "Flexible background mixture models for foreground segmentation," *Image and Vision Computing*, vol. Volume 24, pp. 473-482, 2006.

- [27] A. Shimada, D. Arita and R. Taniguchi, "Dynamic Control of Adaptive Mixture-of-Gaussians Background Model," *AVSS 2006*, p. 5, Nov 2006.
- [28] R. Tan, H. Huo, J. Qian and T. Fang, "Traffic Video Segmentation using Adaptive-K Gaussian Mixture Model," *IWICPAS 2006*, pp. 125-134, Aug 2006.
- [29] L. Carminati and J. Benois-Pinau, "Gaussian Mixture Classification for Moving Object Detection in Video Surveillance Environment," *ICIP 2005*, 2005.
- [30] V. Morellas, L. Pavlidis and P. Tsiamyrtzis, "DETER: detection of events for threat evaluation and recognition," *Machine Vision and Applications*, vol. Vol 15, pp. 29-45, Jun 2003.
- [31] D. Lee, "Online Adaptive Gaussian Mixture Learning for Video Applications," *ECCV Workshop on Statistical Methods for Video Processing*, May 2004.
- [32] Y. Zhang, Z. Liang, Z. Hou, H. Wang and M. Tan, "An Adaptive Mixture Gaussian Background Model with Online Background Reconstruction and Adjustable Foreground Mergence Time for Motion Segmentation," *ICIT 2005*, pp. 23-27, Dec 2005.
- [33] M. Amintoosi, F. Farbiz, M. Fathy, M. Analoui and N. Mozayani, "QR decomposition-based algorithm for background subtraction," *ICASSP 2007*, 2007.
- [34] A. Lepisk, "The use of Optic Flow within Background Subtraction, Master Thesis," Royal Institute of Technology, Nada, Sweden, Jun 2005.
- [35] B. Han and X. Lin, "Update the GMMs via adaptive Kalman filtering," *Proceedings of SPIE*, vol. Vol 5960, no. Issue 3, pp. 1506-1515, 2005.
- [36] H. Wang and D. Suter, "A Re-Evaluation of Mixture-of-Gaussian Background Modeling," *ICASSP 2005*, pp. 1017-1020, Mar 2005.
- [37] J. Lindstrom, F. Lindgren, K. Lindstrom, J. Holst and Holst U., "Background and Foreground Modeling Using an Online EM Algorithm," *VS 2006 in conjunction with ECCV 2006*, 2006 2006.
- [38] C. Stauffer and E. Grimson, "Learning Patterns of Activity Using Real-Time Tracking," *PAMI 2000*, vol. Vol 22, pp. 747-757, 2000.
- [39] J. Landbaso and M. Pardas, "Cooperative Background Modelling using Multiple Cameras Towards Human Detection in Smart-Rooms," *EUSIPCO 2006*, Sep 2006.
- [40] D. Park, J. Kim, Cho S. and Chung S., "Motion Detection in Complex and Dynamic Backgrounds," *PSIVT 2006*, pp. 545-552, Dec 2006.
- [41] A. Mittal and D. Huttenlocher, "Scene Modeling for Wide Area Surveillance and Image Synthesis," *CVPR 2000*, vol. Vol 2, pp. 160-167, Jun 2000.

- [42] P. KaewTraKulPong and R. Bowden, "An Improved Adaptive Background Mixture Model for Real-time Tracking with Shadow Detection," *AVBS 2001*, Sep 2001.
- [43] P. KaewTraKulPong and R. Bowden, "Adaptive Visual System for Tracking Low Resolution Color Targets," *BMVC 2001*, vol. Vol 1, pp. 243-252, Sep 2001.
- [44] P. KaewTraKulPong and R. Bowden, "A Real-Time Adaptive Visual Surveillance System for Tracking Low Resolution Color Targets In Dynamically Changing Scenes," *Journal of Image and Vision Computing*, vol. Vol 21, no. Issue 10, pp. 913-929, Sep 2003.
- [45] D. Lee, "Improved Adaptive Mixture Learning for Robust Video Background Modeling," *IAPR Workshop on Machine Vision for Applications*, pp. 443-446, Dec 2002.
- [46] M. Harville, G. Gordon and J. Woodfill, "Foreground segmentation using adaptive mixture models in color and depth," *Workshop on Detection and Recognition of Events in Video*, Jul 2001.
- [47] F. Porikli, "Human Body Tracking by Adaptive Background Models and Mean-Shift Analysis," *PETS 2003*, Mar 2003.
- [48] Y. Liu, H. Ai and G. Xu, "Moving object detection and tracking based on background subtraction," *Proceedings of SPIE*, vol. Vol 4554, pp. 62-66, 2001.
- [49] A. Pnevmatikakis and L. Polymenakos, "2D Person Tracking Using Kalman Filtering and Adaptive Background Learning in a Feedback Loop," *CLEAR Workshop 2006, LNCS 4122*, pp. 151-160, 2006.
- [50] A. Stergiou, A. Pnevmatikakis and L. Polymenakos, "The AIT Outdoor tracker for Vehicle and Pedestrians in CLEAR 2007," *CLEAR Workshop 2007*, pp. 148-159, 2007.
- [51] P. Power and J. Schoonees, "Understanding Background Mixture Models for Foreground Segmentation," *Imaging and Vision Computing New Zealand*, Nov 2002.
- [52] M. Leotta και J. Mundy, «Learning Background and Shadow Appearance with 3-D Vehicle Models,» *BMVC 2006*, Sep 2006.
- [53] W. Zhang, X. Fang, X. Yang and Q. Wu, "Spatiotemporal Gaussian mixture model to detect moving objects in dynamic scenes," *Journal of Electronic Imaging*, vol. Vol 16, no. Issue 2, Apr 2007.
- [54] Y. Ren, C. Chua and Y. Ho, "Motion Detection with Non-stationary Background," *ICIAP 2002*, pp. 78-83, 2002.
- [55] D. Lee, "Better Foreground Segmentation for Static Cameras via New Energy Form and Dynamic Graph-cut," *ICPR 2006*, 2006.

- [56] D. Lee, "Effective Gaussian Mixture Learning for Video Background," *PAMI 2005*, vol. Vol 27, pp. 827-832, 2005.
- [57] J. Landbaso, M. Pardas and L. Xu, "Hierarchical Representation of Scenes using Activity Information," *ICASSP 2005*, pp. 677-680, Mar 2005.
- [58] P. Withagen, F. Groen and K. Schutte, "EMswitch: a multi-hypothesis approach to EM background modeling," *ACIVS 2003*, Sep 2003.
- [59] M. Haque, M. Murshed and M. Paul, "A Hybrid Object Detection Technique from Dynamic Background Using Gaussian Mixture Models," *IEEE International Workshop on Multimedia Signal, MMSP 2008*, pp. 915-920, Oct 2008.
- [60] M. Haque, M. Murshed and M. Paul, "Improved Gaussian Mixtures for Robust Object Detection by Adaptive Multi-Background Generation," *ICPR 2008*, Dec 2008.
- [61] M. Haque, M. Murshed and M. Paul, "On Stable Dynamic Background Generation Technique using Gaussian Mixture Models for Robust Object Detection," *AVSS 2008*, pp. 41-48, Sep 2008.
- [62] T. Bouwmans, "Recent Advanced Statistical Background Modeling for Foreground Detection - A Systematic Survey," vol. 4, no. (3).<hal-00644746>, pp. 147-176, 2011.
- [63] "Kernel_density_estimation," [Online]. Available: https://en.wikipedia.org/wiki/Kernel_density_estimation. [Accessed Jan 2018].
- [64] A. Elgammal, D. Harwood and L. Davis, "Non-parametric Model for Background Subtraction," *ECCV 2000*, pp. 751-767, Jun 2000.
- [65] A. Tavakkoli, "Automatic Video Object Plane Extraction using non-Parametric Kernel Density Estimation," *Mathematical Methods in Computer Vision*, May 2005.
- [66] A. Tavakkoli, G. Nicolescu M and G. Bebis, "Automatic Statistical Object Detection for Visual Surveillance," *SSIAI 2006*, Mar 2006.
- [67] C. Ianasi, V. Gui, C. Toma and D. Pescaru, "A Fast Algorithm for Background Tracking in Video Surveillance, Using Nonparametric Kernel Density Estimation," *Facta Universitatis, Series: Electronics and Energetics*, vol. Vol 18, no. No. 1, pp. 127-144, Apr 2005.
- [68] T. Tanaka, A. Shimada, D. Arita and R. Taniguchi, "Non-parametric Background and Shadow Modeling for Object Detection," *ACCV 2007*, pp. 159-168, Nov 2007.
- [69] T. Tanaka, A. Shimada, D. Arita and R. Taniguchi, "A Fast Algorithm for Adaptive Background Model Construction Using Parzen Density Estimation," *AVSS 2007*, pp. 528-533, Sep 2007.

- [70] Z. Zivkovic , "Efficient adaptive density estimation per image pixel for the task of background subtraction," *Pattern Recognition Letters*, vol. Vol 27, no. No. 7, pp. 773-780, Jun 2006.
- [71] S. Cvetkovic, P. Bakker and J. Schirris, "Background Estimation and Adaptation Model With Light-Change Removal for Heavily Down-Sampled Video Surveillance Signals," *ICIP 2006*, pp. 1829-1832, 2006.
- [72] S. Witherspoon και M. Zhang, «Coefficient Polynomial Kernel Density Estimation for Visualization,» *International Conference on Modeling and Simulation*,, p. 567, Jun 2007.
- [73] R. Ramezani, P. Angelov and X. Zhou, "A Fast Approach to Novelty Detection in Video Streams using Recursive Density Estimation," *IEEE Symposium on Intelligent Systems*, pp. 142-147, Sep 2008.
- [74] Y. Mao and P. Shi, "Multimodal background model with noise and shadow suppression for moving object detection," *Journal of Southeast University*, vol. Vol 20, no. No. 4, pp. 423-426, Dec 2004.
- [75] Y. Mao and P. Shi, "Diversity sampling based kernel density estimation for background modeling," *Journal of Shanghai University*, vol. Vol 9, no. Issue 6, pp. 506-509, Dec 2005.
- [76] P. Tang, L. Gao and Z. Liu, "Salient Moving Object Detection Using Stochastic Approach Filtering," *ICIG 2007*, pp. 530-535, 2007.
- [77] A. Tavakkoli, M. Nicolescu and G. Bebis, "Robust Recursive Learning for Foreground Region Detection in Videos with Quasi-Stationary Backgrounds," *ICPR 2006*, Aug 2006.
- [78] A. Tavakkoli, M. Nicolescu and G. Bebis, "An Adaptive Recursive Learning Technique for Robust Foreground Object Detection," *ECCV 2006*, May 2006.
- [79] "Maximum_likelihood_estimation," [Online]. Available: https://en.wikipedia.org/wiki/Maximum_likelihood_estimation. [Accessed Jun 2018].
- [80] M. Varanasi and B. Aazhang, "Parametric generalized Gaussian density estimation," *Journal of the Acoustical Society of America*, pp. 1404-1415, 1989.
- [81] M. Do and M. Vetterli, "Wavelet-based Texture Retrieval Using Generalised Gaussian Density and Kullback-Leibler Distance," *Transaction on Image Processing*, pp. 146-158, Feb 2002.
- [82] H. Kim, R. Sakamoto, I. Kitahara, T. Toriyama and K. Kogure, "Robust Silhouette Extraction Technique Using Background Subtraction," *Meeting on Image Recognition and Understand, MIRU2007*, Jul 2007.
- [83] H. Kim, R. Sakamoto, I. Kitahara, Toriyama and K. Kogure, "Robust Foreground Extraction Technique Using Gaussian Family Model and Multiple Thresholds," *ACCV2007, LNCS 4843*, pp. 758-768, Nov 2007.

- [84] M. s. Allili, N. Bouguila and D. Ziou, "Robust Video Foreground Segmentation by Using Generalized Gaussian Mixture Modeling," *Canadian Conference on Computer and Robot Vision, CRV 2007*, pp. 503-509, 2007.
- [85] M. s. Allili, N. Bouguila and D. Ziou, "Finite Generalized Gaussian Mixture Modelling and Application to Image and Video Foreground Segmentation," *Journal of Electronic Imaging*, 2008.
- [86] V. Vapnik and C. Cortes, "Support-vector networks," *Machine Learning (journal)*, pp. 273-297, 1995.
- [87] H. Lin, T. Liu and J. Chuang, "A probabilistic SVM approach for background scene initialization," *ICIP 2002*, vol. Vol 3, pp. 893-896, Sep 2002.
- [88] "Sigmoid_function," [Online]. Available: https://en.wikipedia.org/wiki/Sigmoid_function. [Accessed Jan 2018].
- [89] "Posterior_probability," [Online]. Available: https://en.wikipedia.org/wiki/Posterior_probability. [Accessed Jan 2018].
- [90] "Optical_flow," [Online]. Available: https://en.wikipedia.org/wiki/Optical_flow. [Accessed Jan 2018].
- [91] A. Smola and B. Scholkopf, "A tutorial on support vector regression," *NeuroCOLTS Technical Report Series NC2-TR-1998-030*, Oct 1998.
- [92] J. Ma and J. Theiler, "Accurate on-line support vector regression," *Journal on Neural Computation*, vol. Vol 15, pp. 2683-2703, 2003.
- [93] J. Wang, G. Bebis, M. Nicolescu, M. Nicolescu and R. Miller, "Improving target detection by coupling it with tracking," *Machine Vision and Application*, pp. 1-19, 2008.
- [94] J. Wang, G. Bebis and R. Miller, "Robust Video-Based Surveillance by Integrating Target Detection with Tracking," *CVPR 2006*, Jun 2006.
- [95] A. Tavakkoli, A. Ambardekar, M. Nicolescu and S. Louis, "A Novelty Detection Approach for Foreground Region Detection in Videos with Quasi-stationary Backgrounds," *ISVC 2006*, pp. 40-49, Nov 2006.
- [96] A. Tavakkoli, A. Ambardekar, M. Nicolescu and S. Louis, "A Genetic Approach to Training Support Vector Data Descriptors for Background Modeling in Video Data," *ISVC 2007*, Nov 2007.
- [97] A. Tavakkoli, M. Nicolescu, M. Nicolescu and G. Bebis, "Incremental SVDD Training: Improving Efficiency of Background Modeling in Videos," *International Conference on Signal and Image Processing*, p. 623, Aug 2008.
- [98] A. Tavakkoli, M. Nicolescu, M. Nicolescu and G. Bebis, "Efficient Background Modeling through Incremental Support Vector Data Description," *ICPR 2008*, Dec 2008.

- [99] C. Coniglio, C. Meurie, O. L'ezoray and M. Berbineau, "A Graph Based People Silhouette Segmentation Using Combined Probabilities Extracted from Appearance, Shape Template Prior, and Color Distributions," *ACIVS 2015*, pp. 299-310, 2015.
- [100] M. Hassaballah, A. A. Abdelmgei and H. A. Alshazly, *Image Features Detection, Description*, Springer International Publishing Switzerland 2016, 2016.
- [101] «Multilinear_subspace_learning,» [Ηλεκτρονικό]. Available: https://en.wikipedia.org/wiki/Multilinear_subspace_learning. [Πρόσβαση Jan 2018].
- [102] "Principal_component_analysis," [Online]. Available: https://en.wikipedia.org/wiki/Principal_component_analysis. [Accessed Jan 2018].
- [103] «Singular-value_decomposition,» [Ηλεκτρονικό]. Available: https://en.wikipedia.org/wiki/Singular-value_decomposition. [Πρόσβαση Jan 2018].
- [104] "Eigendecomposition_of_a_matrix," [Online]. Available: https://en.wikipedia.org/wiki/Eigendecomposition_of_a_matrix. [Accessed Jan 2018].
- [105] N. Oliver, B. Rosario and A. Pentland, "A Bayesian Computer Vision System for Modeling Human Interactions," *ICVS 1999*, Jan 1999.
- [106] T. Bouwmans, "Subspace Learning for Background Modeling: A Survey," *Recent Patents on Computer Science*, vol. Vol 2, no. No 3, pp. 223-234, Nov 2009.
- [107] K. H. Karstensen, "Silhouette Extraction using Graphics Processing Units," Master's Thesis, University of Oslo, Oslo, May 2012.
- [108] P. Rajan and S. Prakash, "Moving Foreground Object Detection and Background Subtraction Using Adaptive-K GMM: A Survey," *International Journal of Advance Research in Computer Science and Management Studies*, vol. Issue 1, pp. 301-308, Jan 2014.
- [109] D. E. Butler, J. V. Bove and S. Sridharan, "Real-Time Adaptive Foreground/Background Segmentation," *EURASIP Journal on Applied Signal Processing 2005*, vol. 14, pp. 2292-2304, 25 Aug 2005.
- [110] K. Kim, T. Chalidabhongse, D. Harwood and L. Davis, "Background modeling and subtraction by codebook construction," *IEEE International Conference on Image Processing, ICIP 2004*, 2004.
- [111] K. Kim, T. Chalidabhongse, D. Harwood and L. Davis, "Real time foreground background segmentation using codebook model," *Real time Imaging*, vol. 11, no. (3), pp. 167-256, 2005.

- [112] "https://el.wikipedia.org," Artificial neural network, [Online]. Available: https://el.wikipedia.org/wiki/%CE%9D%CE%B5%CF%85%CF%81%CF%89%CE%BD%CE%B9%CE%BA%CF%8C_%CE%B4%CE%AF%CE%BA%CF%84%CF%85%CE%BF. [Accessed Jan 2018].
- [113] L. Maddalena and A. Petrosino, "A self organizing approach to background subtraction for visual surveillance applications," *IEEE Transactions on Image Processing*, vol. Vol. 17, no. No. 7, pp. 1729-1736, Jul 2008.
- [114] L. Maddalena and A. Petrosino, "A fuzzy spatial coherence-based approach to background/foreground separation for moving object detection," *Neural Computing & Application (2010)*, pp. 179-186, Jun 2009.
- [115] K. Toyama, J. Krumm, B. Brumitt and B. Meyers, "Wallflower: Principles and Practice of Background Maintenance," *International Conference on Computer Vision*, pp. 255-261, Sep 1999.
- [116] S. Messelodi, C. M. Modena, N. Segata and M. Zanin, "A Kalman filter based background updating algorithm robust to sharp illumination changes," *ICIAP 2005*, vol. Vol 3617, pp. 163-170, Sep 2005.
- [117] R. Chang, T. Ghandi and M. Trivedi, "Vision modules for a multi sensory bridge monitoring approach," *ITSC 2004*, pp. 971-976, Oct 2004.
- [118] G. Sharma, "Digital Color Imaging Handbook," Webster, New York, CRC Press, 2003, pp. 31-32.
- [119] P. Soille, "Morphological Image Analysis: Principles and Applications," Springer-Verlag, 1999, pp. 173-174.
- [120] S. P. Lloyd, "Least Squares Quantization in PCM," *IEEE Transactions on Information Theory*, vol. Vol. 28, pp. 129-137, 1982.
- [121] D. Arthur and S. Vassilvitskii, "K-means++: The Advantages of Careful Seeding," *SODA '07: Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 1027-1035, 2007.
- [122] D. J. Bora, A. K. Gupta και F. A. Khan, «Comparing the Performance of L*A*B* and HSV Color Spaces with Respect to Color Image Segmentation,» *International Journal of Emerging Technology and Advanced Engineering*, τόμ. Vol. 5, αρ. Issue 2, Feb 2015.
- [123] "Mathworks," Mathworks, [Online]. Available: <https://www.mathworks.com/help/images/examples/color-based-segmentation-using-k-means-clustering.html>. [Accessed Jan 2018].
- [124] X. H. Fang, W. Xiong, B. J. Hu and L. T. Wang, "A Moving Object Detection Algorithm Based on Color Information," *Journal of Physics: Conference Series*, pp. 384-387, 2006.
- [125] D. S. Farin, "Automatic Video Segmentation Employing Object/Camera Modeling," Technische Universiteit Eindhoven, Eindhoven, 2005.

- [126] P. Chen, "Image background reconstruction by Gaussian mixture based model reinforced with temporal-spatial confidence," *Journal of Algorithms & Computational Technology*, pp. 23-30, Jan-Mar 2016.
- [127] A. Newson, A. Almansa, M. Fradet, Y. Gousseau and P. Perez, "Video Inpainting of Complex Scenes," *Society for Industrial and Applied Mathematics: J. Imaging Science*, vol. Vol. 7, no. No. 4, pp. 1993-2019, Oct 2014.
- [128] J.-B. Huang, S. B. Kang, N. Ahuja and J. Kopf, "Temporally Coherent Completion of Dynamic Video," *ACM Trans. Graph.*, pp. 35, 6, Article 196, Nov 2016.
- [129] "https://www.mathworks.com," [Online]. Available: <https://www.mathworks.com/help/images/ref/regionprops.html>. [Accessed Feb 2018].
- [130] "i-LIDS dataset," in *IEEE International Conference on Advanced Video and Signal based Surveillance*, London (United Kingdom), Sep 2007.
- [131] Pedestrian overpass video, "<http://briefcam.com>," Nov 2017. [Online]. Available: <https://www.youtube.com/watch?v=aUdKzb4LGJI>. [Accessed Jan 2018].
- [132] B. Ben and I. Reid, "Stable Multi-Target Tracking in Real-Time Surveillance Video," *CVPR*, pp. 3457-3464, Jun 2011.
- [133] Yoga training video, "<https://app.hoolio.com/>," Sep 2016. [Online]. Available: <https://www.dropbox.com/s/uo5kj5ygdbbcsai/Temp.mp4?dl=0>. [Accessed Jan 2018].
- [134] D. Baltieri, R. Vezzani and R. Cucchiara, "3DPes: 3D People Dataset for Surveillance and Forensics," *Proceedings of the 1st International ACM Workshop on Multimedia access to 3D Human Objects*, pp. 59-64, nov 2011.
- [135] "Tensor," [Online]. Available: <https://en.wikipedia.org/wiki/Tensor>. [Accessed Jan 2018].
- [136] H. Hu, L. Xu and H. Zhao, "A spherical codebook in yuv color space for moving object detection," *Sensor Letters*, vol. 10, no. (1), pp. 177-189, Jan 2012.
- [137] X. Deng, J. Bu, Z. Yang, C. Chen and Y. Liu, "A block-based background model for video surveillance," *IEEE International Conference on Acoustics Speech and Signal Processing, ICASSP 2008*, pp. 1013-1016, Mar 2008.

Πηγές Εικόνων

[Εικόνα 2.4]

https://en.wikipedia.org/wiki/Unimodality#/media/File:Normal_distribution_pdf.svg

[Εικόνα 2.7]

https://en.wikipedia.org/wiki/Mean#/media/File:Visualisation_mode_median_mean.svg

[Εικόνα 2.8]

<http://tinyheero.github.io/assets/mixture-model-mixtools-1.png>

[Εικόνα 2.9]

https://en.wikipedia.org/wiki/Kernel_density_estimation#/media/File:Comparison_of_1D_bandwidth_selectors.png

[Εικόνα 2.10]

<https://www.mathworks.com/help/stats/kernel-distribution.html>

[Εικόνα 2.11]

https://en.wikipedia.org/wiki/Generalized_normal_distribution#/media/File:Generalized_normal_densities.svg

[Εικόνα 2.12]

<http://littledatablog.com/blog/visualizing-leptokurtic-vs-platykurtic>

[Εικόνα 2.13]

[https://en.wikipedia.org/wiki/Support_vector_machine#/media/File:Svm_separating_hyperplanes_\(SVG\).svg](https://en.wikipedia.org/wiki/Support_vector_machine#/media/File:Svm_separating_hyperplanes_(SVG).svg)

[Εικόνα 2.14]

https://en.wikipedia.org/wiki/Principal_component_analysis#/media/File:Gaussian_ScatterPCA.svg

[Εικόνα 2.17]

<http://legacydirs.umiacs.umd.edu/~knkim/paper/bgs-ICIP2004.pdf>

[Εικόνα 2.19]

https://el.wikipedia.org/wiki/%CE%9D%CE%B5%CF%85%CF%81%CF%89%CE%BD%CE%B9%CE%BA%CF%8C_%CE%B4%CE%AF%CE%BA%CF%84%CF%85%CE%BF#/media/File:Neural_network_example.svg

[Εικόνα 2.20]

<https://link.springer.com/article/10.1007/s00521-009-0285-8>

[Εικόνα 3.1]

https://docs.opencv.org/3.3.1/d1/db7/tutorial_py_histogram_begins.html

[Εικόνα 3.2]

<https://openclipart.org/detail/218703/cie-labs-colorspace-as-coordinate-system-a1>