



ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

Τμήμα Μηχανικών Η/Υ, Τηλεπικοινωνιών και Δικτύων

Πολυτεχνική Σχολή

«Adaptive Color Modelling for Background Detection and Tracking»

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Νικόλαος Α. Καπετάνιος

e-mail : nikapeta@inf.uth.gr

Επιβλέπουσα Καθηγήτρια : Αλεξία Μπριασούλη

Επιτροπή : Ηλίας Χούστης, Αλεξία Μπριασούλη

Βόλος, Οκτώβριος 2007



**ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ
ΒΙΒΛΙΟΘΗΚΗ & ΚΕΝΤΡΟ ΠΛΗΡΟΦΟΡΗΣΗΣ
ΕΙΔΙΚΗ ΣΥΛΛΟΓΗ «ΓΚΡΙΖΑ ΒΙΒΛΙΟΓΡΑΦΙΑ»**

Αριθ. Εισ.: 6010/1
Ημερ. Εισ.: 06-11-2007
Δωρεά: Συγγραφέα
Ταξιθετικός Κωδικός: ΠΤ - ΜΗΥΤΔ
2007
ΚΑΠ

στον αδερφό μου,
Δημήτρη

Ευχαριστίες

Κατ' αρχήν, θα ήθελα να ευχαριστήσω την καθηγήτρια μου κ. Αλεξία Μπριασούλη, η οποία στάθηκε αιτία να ασχοληθώ με το αντικείμενο της «Ψηφιακής Επεξεργασίας Εικόνας». Ήταν αυτή που με βοήθησε στην επιλογή του παρόντος θέματος αλλά και με τις πολύτιμες συμβουλές της καθ' όλη τη διάρκεια της προσπάθειάς μου, να φέρω εις πέρας τη διπλωματική μου εργασία.

Επίσης θα ήθελα να ευχαριστήσω τον καθηγητή κ. Ηλία Χούστη για την συμβολή του στην επιτυχή ολοκλήρωση της παρούσας εργασίας.

Ένα μεγάλο ευχαριστώ στην οικογένεια μου για την αμέριστη συμπαράσταση και υποστήριξή τους στο πρόσωπό μου, αλλά και την υπομονή και καθοδήγησή τους σε όλη τη διάρκεια της πενταετούς φοίτησης μου στο πανεπιστήμιο.

Τέλος, θέλω να ευχαριστήσω όλους τους φίλους μου που μου στάθηκαν όσο κανείς άλλος με την ψυχολογική υποστήριξή τους. Καθώς επίσης, ένα μεγάλο ευχαριστώ στην φίλη μου Ειρήνη, τόσο για τα παραπάνω όσο και για την πολύτιμη βοήθεια της στις διορθώσεις του παρακάτω κειμένου.

Κεφάλαιο 1 - Εισαγωγή.....	4
1.1 Γενικά.....	4
1.2 Κίνητρα και Στόχοι.....	5
1.3 Γενικά για τις Μεθόδους, τα Βήματα και τις Ιδιότητες τους.....	7
Κεφάλαιο 2 - Ο Αλγόριθμος Πρόβλεψης – Μεγιστοποίησης.....	11
(Expectation – Maximization Algorithm).....	11
2.1 Ιστορικά.....	11
2.2 Γενικά για τον EM.....	11
2.3 Προδιαγραφή της διαδικασίας του EM.....	12
2.4 Ιδιότητες του EM.....	14
2.5 Παράδειγμα του EM: Mixture of Gaussian.....	15
2.6 Πιο Αναλυτικά για τον EM.....	21
Κεφάλαιο 3 - Παρουσίαση Βασικών Μεθόδων.....	26
3.1 Μοντελοποίηση Προσαρμόσιμου Έγχρωμου Παρασκηνίου για Ροές.....	
Δεδομένων Video Πραγματικού Χρόνου.....	26
3.1.1 Γενικά.....	26
3.1.1.1 Ο Χώρος Χρωμάτων HSV και η Σχέση του με τον RGB.....	28
3.1.2 Μοντελοποίηση του Παρασκηνίου με το Μοντέλο Χρωμάτων HSV.....	31
3.1.3 Ανίχνευση Pixels των Αντικειμένων του Παρασκηνίου.....	32
3.1.4 Ανανέωση του Μοντέλου του Παρασκηνίου και Αποτελέσματα.....	33
3.2 Μοντελοποίηση Προσαρμόσιμου Έγχρωμου Παρασκηνίου για Ροές.....	
Δεδομένων Video Πραγματικού Χρόνου – Μέθοδος NHD	36
3.2.1 Γενικά.....	36
3.2.2 Ο Αλγόριθμος.....	36
3.2.3 Υλοποίηση.....	38
3.2.4 Τελική επεξεργασία.....	40
3.2.5 Πειράματα – Συγκρίσεις - Αποτελέσματα.....	41
3.3 Μοντελοποίηση Προσαρμόσιμου Έγχρωμου Παρασκηνίου για Ροές.....	
Δεδομένων Video Πραγματικού Χρόνου με Χρήση GMM.....	44
3.3.1 Γενικά.....	44
3.3.2 Η Μέθοδος.....	45
3.3.3 Το Συνδυαστικό Μοντέλο.....	46

3.3.4 Μοντέλο Εκτίμησης Παρασκηνίου με Χρήση GMM.....	50
3.3.5 Συνεκτικές Συνιστώσες.....	51
3.3.6 Ιχνηλάτηση Πολλαπλής Υπόθεσης (Multiple Hypothesis Tracking).....	52
3.3.7 Πειράματα και Αποτελέσματα.....	53
3.3.8 Εφαρμοσιμότητα.....	55
3.4 Μοντελοποίηση Προσαρμόσιμου Έγχρωμου Παρασκηνίου για Ροές.....	
Δεδομένων Video Πραγματικού Χρόνου και Ανίχνευση Σκιών.....	57
3.4.1 Γενικά.....	57
3.4.2 Μοντελοποίηση του Background.....	57
3.4.3 Προσαρμόσιμο Συνδυαστικό Μοντέλο Gaussian.....	58
3.4.4 Σειριακοί Αλγόριθμοι EM.....	60
3.4.5 Ανίχνευση Σκιών και Χρωματικό Μοντέλο.....	60
3.4.6 Πειράματα και Αποτελέσματα.....	62
Κεφάλαιο 4 - Πειράματα και Αναλύσεις.....	64
4.1 Δομή και Ακολουθία των Πειραμάτων.....	64
4.2 Προσέγγιση Background με Median Filtering.....	64
4.2.1 Γενικά για το Median Filter.....	64
4.2.2 Εφαρμογή του Median Filter για Background Subtraction.....	66
4.3 Προσαρμοστικό Μοντέλο Συνδυασμού Gaussians για την.....	
Εξαγωγή του Παρασκηνίου.....	69
4.3.1 Γενικά.....	69
4.3.2 Η προσέγγιση.....	70
4.3.3 Υλοποίηση και Πειράματα.....	71
4.4 Προσέγγιση Background με Αλγορίθμους Ελαχίστης Τομής Γράφου και.....	
Μορφολογικών Τελεστών.....	73
4.4.1 Γενικά.....	73
4.4.2 Αλγόριθμος Μορφολογικών Τελεστών.....	74
4.4.3 Τομές Γράφων για την τμηματοποίηση του Προσκηνίου.....	75
4.4.4 Πειράματα και Αποτελέσματα.....	77
4.4.5 Συμπεράσματα.....	79
Κεφάλαιο 5 – Συμπεράσματα.....	81
5.1 Παρατηρήσεις και Συμπεράσματα.....	81
5.2 Μελλοντική Έρευνα.....	82
Βιβλιογραφία – Αναφορές.....	85

Κεφάλαιο 1 - Εισαγωγή



1.1 Γενικά

Στα πλαίσια αυτής της διπλωματικής εργασίας μελετάται η μοντελοποίηση του παρασκηνίου (background) χρησιμοποιώντας μια σειρά συνεχόμενων χρονικά εικόνων (frames). Το αποτέλεσμα που παράγεται είναι ένα μοντέλο, το οποίο μπορεί να χρησιμοποιηθεί με διάφορους τρόπους. Είναι μια τεχνική που χρησιμοποιείται ευρέως στην παρακολούθηση εικόνας-video μέσω υπολογιστών. Αυτή έχει ως σκοπό την ανίχνευση αντικειμένων στο προσκήνιο (foreground objects) ή την παρακολούθηση κινούμενων αντικείμενων σε μια σειρά από tracking frames.

Αυτό σημαίνει ότι όσο πιο ακριβής είναι η μοντελοποίηση του παρασκηνίου τόσο πιο σωστή θα είναι η ανίχνευση και η περαιτέρω επεξεργασία των διαφόρων αντικειμένων στο προσκήνιο. Επίσης, γίνεται χρήση της μεθόδου σε αυτόματα συστήματα οπτικής παρακολούθησης, διεπαφές μεταξύ ανθρώπου και υπολογιστή καθώς επίσης και σε τηλεπικοινωνίες με πολύ χαμηλό ρυθμό μετάδοσης δεδομένων.

Η ανίχνευση αντικειμένων-στόχων στο προσκήνιο, με ψηφιακά δεδομένα τύπου video, έχουν πολλές εφαρμογές τα τελευταία χρόνια. Σε αυτές τις εφαρμογές οι στόχοι είναι άνθρωποι, οχήματα ή άλλα διάφορα αντικείμενα. Η κοινή ιδιότητα των στόχων αυτών είναι πως αργά ή γρήγορα θα παρουσιάσουν κάποια κίνηση, η οποία είναι απόδειξη που τα ξεχωρίζει από τα αντικείμενα του παρασκηνίου.

Η έννοια του adaptive background modelling περιλαμβάνει το γεγονός μιας χρονικά μεταβαλλόμενης κατάστασης του παρασκηνίου. Νέα αντικείμενα μπορεί να προστεθούν σε έναν χώρο ή περιβάλλον και να παραμείνουν εκεί για αρκετό χρονικό διάστημα, ώστε να πρέπει να ενσωματωθούν στο παρασκήνιο. Αντικείμενα που ήδη αποτελούν μέρος του παρασκηνίου μπορεί να παρουσιάζουν κάποια κίνηση (φύλλα δέντρου κινούμενα από τον αέρα, κίνηση νερού, κλπ), κάτι το οποίο δεν πρέπει να αλλάζει την σύσταση του background, θεωρώντας αυτά τα αντικείμενα ως

αντικείμενα του προσκηνίου. Κάποια φυσικά καιρικά φαινόμενα (βροχή, ομίχλη, κλπ) ή αλλαγές στο φωτισμό (φυσικό ή τεχνητό) δεν πρέπει να επηρεάζουν, ή τουλάχιστον όχι σε σημαντικό βαθμό, την μοντελοποίηση και τα αποτελέσματα της.

Όλα αυτά είναι δυνατόν να συμβούν μόνο με μεθόδους προσαρμοστικής μοντελοποίησης παρασκήνιου (adaptive background modelling). Αν το παρασκήνιο θεωρηθεί στατικό (έστω ότι είμαστε σε εσωτερικό κλειστό χώρο με σταθερό φωτισμό), η παραμικρή μετακίνηση ενός αντικειμένου ή η τοποθέτηση ενός νέου, θα θεωρηθούν λανθασμένα αντικείμενα του προσκηνίου μέχρι να επέμβει ο ανθρώπινος παράγοντας και να επανεκκινήσει το σύστημα, ώστε να δημιουργηθεί ένα νέο παρασκήνιο με τα νέα δεδομένα.

Το ζήτημα που προκύπτει από τα παραπάνω είναι η ακρίβεια των αποτελεσμάτων που επιδιώκεται και οι διαθέσιμοι υπολογιστικοί πόροι. Είναι λογικό το adaptive background modelling να απαιτεί πολλαπλάσιους πόρους σε σχέση τη στατική μοντελοποίηση παρασκήνιου, ως αντίτιμο μιας, μεγάλης ακρίβειας, εκτίμηση/ανίχνευση παρασκήνιου με σαφώς μεγαλύτερα χρονικά διαστήματα παρακολούθησης και με ελάχιστη ανάμιξη του ανθρώπινου παράγοντα κατά την διάρκεια της λειτουργίας του. Επιπλέον, η ακρίβεια αυτή χαρίζει την δυνατότητα του εύκολου object tracking. Όπως προαναφέρθηκε δηλαδή είναι πολύ πιο εύκολο έτσι να αναγνωριστούν και να παρακολουθηθούν σωστά, αντικείμενα του προσκηνίου.

Οι αλγόριθμοι του adaptive background removal estimation (προσαρμοστική εκτίμηση/ανίχνευση και εξαγωγή του παρασκήνιου) ανιχνεύουν τα αντικείμενα-στόχους στο προσκήνιο σαν τα τμήματα εικόνας που διαφέρουν από αυτήν που έχουν ήδη «μάθει» σαν background, αλλά εξαρτώνται σε μεγάλο βαθμό από την κίνηση αυτών των αντικειμένων. Αν αυτά είναι, για αρκετά μεγάλο χρονικό διάστημα, στάσιμα τότε η διαδικασία «μάθησης» του παρασκήνιου τα ενσωματώνει και αυτά στο background.

1.2 Κίνητρα και Στόχοι

Στόχος είναι η υλοποίηση ενός σταθερού και αξιόπιστου μηχανισμού παρακολούθησης εικόνας και video το οποίο θα μπορεί να χειρίζεται καταστάσεις

αλλαγής φωτισμού, κινήσεις αντικειμένων στο προσκήνιο και στο παρασκήνιο· καθώς επίσης, άλλες αυθαίρετες αλλαγές σε παρατηρούμενες σκηνές.

Στο παρελθόν υπήρχαν σοβαροί περιορισμοί στην επεξεργασία real-time video εφαρμογών. Αυτοί οφείλονταν στις υπολογιστικές δυνάμεις του εξοπλισμού που οι ερευνητές είχαν στα χέρια τους. Με την πάροδο των χρόνων και την εκθετική αύξηση της υπολογιστικής δύναμης των υπολογιστών δόθηκε η δυνατότητα για περισσότερο πολύπλοκες μοντελοποιήσεις και επεξεργασία των δεδομένων, μέχρι και πραγματικού χρόνου.

Οι ψηφιακές κάμερες το video (και άλλα είδη multimedia content) και η χρήση αυτών είναι πλέον ευρέως διαδεδομένα. Θεωρούνται απαραίτητα εργαλεία σε επαγγελματικούς χώρους για την παρακολούθησή τους, για τηλεπικοινωνιακές χρήσεις αλλά και για την διασκέδαση, τα οποία οδηγούν σε ανυπολόγιστες ποσότητες καταγεγραμμένου υλικού video που πρέπει να επεξεργαστούν και να αποθηκευτούν μέρα με την μέρα.

Η εξέλιξη της τεχνολογίας έχει βοηθήσει σημαντικά στην ανάπτυξη συστημάτων που αναλύουν αυτόματα το ψηφιακό υλικό που καταγράφεται. Θεμελιακό βήμα στην επεξεργασία αυτή θεωρείται η ανάγκη του διαχωρισμού κινούμενων αντικειμένων από το παρασκήνιο. Ιστορικά αναφέρουμε ότι οι υπολογιστικές πολυπλοκότητες των διαδικασιών αυτών είχαν περιορίσει την αυτόματη τμηματοποίηση του background σε μικρού μεγέθους frames (ανάλυση εικόνας), χαμηλά frame rates και offline διαδικασίες για την επεξεργασία. Το μεγάλο πλήθος των προσεγγίσεων του προβλήματος αυτού διαφέρουν βασικά στον τύπο του background μοντέλου που χρησιμοποιούν και στην διαδικασία ανανέωσης του σε συνθήκες πραγματικού χρόνου.

Η έννοια του μη προσαρμόσιμου (non-adaptive) background έχει απορριφθεί από τους ειδικούς, εξ' αιτίας της ανάγκης που δημιουργείται για συνεχείς και «χειροκίνητες» αρχικοποιήσεις του background, το οποίο με τη σειρά του οφείλεται σε αλλαγές του σκηνικού που πρέπει να ενσωματωθούν από το προσκήνιο στο παρασκήνιο. Οι μέθοδοι με μη προσαρμόσιμα background μπορούν να χρησιμοποιηθούν με σχετική επιτυχία μόνο σε περιβάλλοντα με υψηλή παρακολούθηση από ανθρώπινο δυναμικό, για δεδομένα μικρών χρονικών περιόδων και χωρίς σημαντικές αλλαγές του σκηνικού στις περιόδους αυτές.

Υπάρχουν πολλά βασικά προβλήματα στις μεθόδους του adaptive background modelling. Ενδεικτικά αναφέρονται περιπτώσεις, στις οποίες πολλά αντικείμενα κινούνται μαζί στο προσκήνιο αποκρύπτοντας έτσι ένα μεγάλο μέρος του background για κάποιο χρόνο ικανό να αλλοιώσει την πληροφορία για την σωστή μοντελοποίηση του. Επιπρόσθετα, αλλαγές στο φωτισμό του σκηνικού ή διάφορα καιρικά φαινόμενα (π.χ. χαλάζι, βροχή) είναι δυνατόν να δημιουργήσουν προβλήματα στην χρήση αυτών των μεθόδων. Τέλος, άλλη μια δύσκολη περίπτωση είναι να έχουμε κίνηση σε αντικείμενα του background, όπως κίνηση φύλλων και κλαδιών δέντρων ή κυματισμοί σε επιφάνειες με νερό από τον αέρα.

Αντιλαμβανόμαστε ότι ένα από τα πιο σημαντικά βήματα μιας μεθόδου μοντελοποίησης του παρασκήνιου προϋποθέτει τον σωστό διαχωρισμό των αντικειμένων που βρίσκονται στο προσκήνιο από αυτές που είναι στο παρασκήνιο (τμηματοποίηση-segmentation). Για τον σκοπό αυτό, έχει αναπτυχθεί μια πληθώρα μεθόδων, πολλές από τις οποίες θα αναλυθούν διεξοδικά σε αυτήν την εργασία.

1.3 Γενικά για τις Μεθόδους, τα Βήματα και τις Ιδιότητες τους

Σε αυτήν την ενότητα παρουσιάζεται μια κατηγορία μεθόδων για να διαχωρίσουμε αντικείμενα του background από αυτά του foreground είναι χρησιμοποιώντας το optical flow¹. Τυπικά, η κίνηση αντιπροσωπεύεται από διανύσματα που δημιουργούνται ή τερματίζονται σε pixels, ως αποτέλεσμα του όλου frame sequence. Οι μέθοδοι αυτοί εκμεταλλεύονται την συνοχή και την συνάφεια που έχει το optical flow σε ένα μικρό χρονικό διάστημα. Έχει παρατηρηθεί ότι αυτός ο τρόπος αποδίδει ικανοποιητικά αποτελέσματα όσον αφορά αντικείμενα του προσκήνιου. Υπάρχουν όμως δύο σοβαρά μειονεκτήματα που αξίζει να σημειωθούν.

Καταρχάς, είναι δύσκολο να παράγουμε το optical flow σε περιοχές της εικόνας με λίγα χαρακτηριστικά γνωρίσματα «υφής» (texture), ή στα όρια των ασυνεχειών της εικόνας. Τέτοια λάθη και ασάφειες μπορούν να οδηγήσουν σε προβλήματα στην ανίχνευση αντικειμένων του προσκήνιου. Το δεύτερο βασικό μειονέκτημα είναι η χρονική καθυστέρηση που εισάγεται κατά την ανίχνευση της

¹ Optical flow είναι η έννοια που προσεγγίζει την κίνηση των αντικειμένων μέσα από μία οπτική αναπαράσταση.

συνοχής του optical flow, το οποίο με τη σειρά του οδηγεί σε καθυστέρηση στην ανίχνευση foreground αντικειμένων.

Μια τυπική μέθοδος, η οποία αφορά την real-time τμηματοποίηση (segmentation) κινούμενων περιοχών σε μια σειρά από frames, αποτελεί το color-based background subtraction. Μία καλή υλοποίηση της μεθόδου αυτής αποτελεί το multi-colour background model για κάθε pixel. Ένα σημείο το οποίο αξίζει να αναφερθεί εδώ είναι ότι αυτή η μέθοδος έχει αργή «εκμάθηση» στην αρχή και ειδικά όταν πρόκειται για περιβάλλοντα με συχνή και πυκνή κίνηση των αντικειμένων στο προσκήνιο. Επίσης, δεν μπορεί να διακρίνει μεταξύ κινούμενων foreground objects και κινούμενων σκιών.

Γι' αυτούς τους λόγους, θα πρέπει να ληφθούν σοβαρά υπ' όψιν και να εξεταστούν οι συνθήκες με βάση τις οποίες πραγματοποιείται το update του μοντέλου και πως αυτές πρέπει να αλλάζουν σε διαφορετικές φάσεις. Αυτό θα επιτρέψει στο σύστημα να «εκπαιδευτεί» ταχύτερα και ακριβέστερα στα δεδομένα μας, αλλά και να γίνει προσαρμόσιμο στις διάφορες αλλαγές του περιβάλλοντος.

Το background subtraction προϋποθέτει τον υπολογισμό μιας εικόνας αναφοράς (reference image), από την οποία θα «αφαιρείται» κάθε νέο frame και έπειτα θα «φιλτράρεται» με βάση κάποιο κατώφλι για να μας δώσει την δυαδική τμηματοποίηση της εικόνας, η οποία ξεχωρίζει τις περιοχές με μη στάσιμα αντικείμενα. Αυτό είναι και το βασικό μειονέκτημα του μη προσαρμόσιμου background. Το κατώφλι είναι μια σταθερή παράμετρος και δεν υποστηρίζεται από προσαρμοστικές μεθόδους. Τα διάφορα videos απαιτούν διαφορετικά κατώφλια. Το παρασκήνιο αλλάζει και μάλιστα με διαφορετικούς ρυθμούς κάθε φορά. Οι αλλαγές αφορούν τα αντικείμενα, τον φωτισμό αλλά και τον θόρυβο (π.χ. από την κάμερα λήψης, jitter, κλπ.). Μια απλοϊκή μορφή υπολογισμού της εικόνας αναφοράς του παρασκηνίου είναι ο μέσος όρος όλων των εικόνων-frames στη πάροδο ενός συγκεκριμένου χρονικού διαστήματος. Όμως και αυτή η μέθοδος έχει διάφορα προβλήματα και απαιτεί μια περίοδο «εκπαίδευσης» υπό την απουσία αντικειμένων στο προσκήνιο. Η κίνηση αντικειμένων στο παρασκήνιο μετά την περίοδο της εκπαίδευσης και η στασιμότητα αντικειμένων στο προσκήνιο κατά τη διάρκεια της περιόδου εκπαίδευσης θα θεωρηθούν από το σύστημα ως μόνιμα foreground αντικείμενα. Επιπροσθέτως, η προσέγγιση αυτή δεν μπορεί να χειριστεί κατάλληλα βαθμιαίες αλλαγές στον φωτισμό του σκηNIKού. Όλα αυτά οδηγούν στην απαίτηση

της συνεχούς επανεκτίμησης του background μοντέλου και έτσι αναπτύσσονται συνεχώς νέες μέθοδοι προσαρμόσιμων background models.

Μία πολλά υποσχόμενη μέθοδος, είναι αυτή των πολλαπλών κατανομών Gauss (Mixture of Gaussians). Στην μέθοδο αυτή, το χρώμα του κάθε pixel (τιμή) του background αναπαρίσταται από πολλαπλές κατανομές Gauss. Ωστόσο, έχει παρατηρηθεί πειραματικά, πως η χρήση περισσότερων από δύο κατανομών Gauss μπορεί να αλλοιώσει την απόδοση του συστήματος στην επιτυχή ανίχνευση αντικείμενων του προσκηνίου. Αυτές οι παράμετροι θα αναλυθούν παρακάτω.

Μία κοινή τακτική βελτιστοποίησης της χρήσης του πολλαπλού μοντέλου Gauss είναι ο αλγόριθμος Expectation – Maximization (EM algorithm). Ο αλγόριθμος αυτός είναι μια επαναληπτική μέθοδος, η οποία εγγυάται την σύγκλιση στο τοπικό μέγιστο δεδομένου ενός χώρου – πεδίου ορισμού. Εξ' αιτίας των χρονικών αλλά και των αποθηκευτικών περιορισμών που θέτει αυτή η μέθοδος, η μοντελοποίηση κάθε pixel απαιτεί την υλοποίηση online παραλλαγών του EM αλγορίθμου.

Στην επιστήμη των υπολογιστών όταν αναφέρουμε την έννοια του online αλγορίθμου εννοούμε αυτόν που μπορεί να επεξεργαστεί την είσοδό του σειριακά (κομμάτι-κομμάτι), χωρίς να έχει ολόκληρα τα δεδομένα εισόδου διαθέσιμα από την αρχή. Έχουν ήδη υλοποιηθεί πολλές τέτοιες παραλλαγές, οι οποίες μπορούν να κατηγοριοποιηθούν σε δύο βασικές κατηγορίες. Η πρώτη αφορά τους αλγορίθμους, οι οποίοι χρησιμοποιούν παραμετρική εκτίμηση των συναρτήσεων πυκνότητας πιθανότητας (pdf's). Δηλαδή χρησιμοποιούν νέα δεδομένα για να ανανεώσουν την προηγούμενη εκτίμηση, χωρίς όμως να τροποποιούν την δομή του προηγούμενου μοντέλου. Αυτή η προσέγγιση βασίζεται σε πολύ μεγάλο βαθμό στην αρχική εκτίμηση του μοντέλου. Η δεύτερη βασική κατηγορία αφορά τις μη παραμετρικές προσεγγίσεις οι οποίες κάνουν χρήση στοχαστικών διαδικασιών σε συνδυασμό με τιμές κατωφλίου για την παραγωγή πυρήνων Gauss (Gaussian kernels) στο ήδη υπάρχον mixture model. Αυτή η κατηγορία δεν θα αναφερθεί ιδιαίτερα σε αυτήν την εργασία. Περισσότερα για τον αλγόριθμο EM και την λειτουργία του αναφέρονται σε επόμενα πεδία.

Εφόσον το παρασκήνιο και η εμφάνιση που αυτό παρουσιάζει αλλάζει με την πάροδο του χρόνου, τόσο γεωμετρικά (σχηματικά) όσο και στον φωτισμό του σκηνικού, είναι πλέον αναγκαίο να γίνεται προσαρμοστική επανεκτίμηση του background. Η ανανέωση της εκτίμησης όμως, ιδανικά θα πρέπει να γίνεται χωρίς την συμβολή του ανθρώπινου παράγοντα. Κατ' αυτό τον τρόπο, οι περισσότερες

προσεγγίσεις προτείνουν την εκτίμηση του background απ' ευθείας από το frame sequence. Αυτή η εκτίμηση είναι το background model. Σε κάθε pixel το μοντέλο του παρασκηνίου μπορεί να αναπαρίσταται από μία μόνο αριθμητική τιμή. Ωστόσο, πολλές φορές τίθεται η ανάγκη να αναπαρίσταται από μία ή και περισσότερες κατανομές, με σκοπό να αντικατοπτρίζεται έτσι η τάση αλλαγής του σκηνικού του background. Με αυτήν την λογική το πρόβλημα ανάγεται στην ακριβή εκτίμηση της συνάρτησης πυκνότητας πιθανότητας σε κάθε pixel, ποσοτικοποιώντας την πιθανότητα ένα συγκεκριμένο pixel να είναι στο παρασκήνιο. Η λύση δίνεται, όπως θα δούμε και παρακάτω, από τον αλγόριθμο EM και τα Gaussian Mixture Models.

Μια άλλη μέθοδος για background modelling βασίζεται στην διαδικασία mean-shift. Το mean shift παρέχει ιδιότητες σύγκλισης που οδηγούν το σύστημα να επιτύχει αξιόπιστη μοντελοποίηση του παρασκηνίου. Η μέθοδος αυτή προσομοιάζει με άλλες βασικές μεθόδους, εκτός κάποιων μικρών διαφορών, και δεν θα αναλυθεί περαιτέρω.

Μια διαφορετική και πιο απλή λύση είναι η χρήση του ιστογράμματος των τιμών των pixels στα τελευταία N frames. Αν αυτό το N είναι αρκετά μεγάλο, τότε είναι πολύ πιθανόν το ιστόγραμμα να αναπαριστά τις τιμές του background και όχι του foreground. Η χρήση των N πιο πρόσφατων frames μας επιτρέπει την προσαρμογή του παρασκηνίου στα πιο πρόσφατα δεδομένα. Από την άλλη πλευρά, όμως, είναι γνωστό πως το ιστόγραμμα δεν είναι αρκετά ικανοποιητικό στην εκτίμηση μιας pdf, εξ' αιτίας, της διακριτής φύσης του, και πρέπει να εφαρμοστούν συναρτήσεις εξομάλυνσης (smoothing) για βελτίωσή του.

Άλλη μία εναλλακτική λύση στην γνώση και μοντελοποίηση του background είναι αυτή κατά την οποία ο αλγόριθμος αναπαριστά κάθε pixel στο frame με μία ομάδα συστάδων (group of clusters)². Τα clusters αυτά είναι ταξινομημένα σύμφωνα με την πιθανότητα, ότι μοντελοποιούν το background και είναι προσαρμοσμένα να χειρίζονται αλλαγές στο παρασκήνιο και στον φωτισμό του. Τα pixels που έπονται στην διαδικασία αυτή, ελέγχονται σύμφωνα με το αντίστοιχο cluster group και κατηγοριοποιούνται σύμφωνα με το εάν το cluster αυτό θεωρείται μέρος του background (αν δηλαδή η πιθανότητα του αντίστοιχου cluster είναι πάνω από ένα κατώφλι).

² Clustering είναι η κατηγοριοποίηση αντικειμένων σε διαφορετικές ομάδες, δηλαδή η διαμέριση ενός συνόλου δεδομένων σε υποσύνολα (clusters), ώστε τα δεδομένα σε κάθε υποσύνολο να έχουν κάποιο κοινό γνώρισμα – συνήθως ορισμένο από κάποιο μέτρο ομοιότητας [19].

Κεφάλαιο 2 - Ο Αλγόριθμος Πρόβλεψης – Μεγιστοποίησης (Expectation – Maximization Algorithm)



2.1 Ιστορικά

Ο αλγόριθμος expectation-maximization (EM) είχε παρουσιαστεί και πάρει το όνομα του το 1977 στο δημοσιευμένο paper των Arthur Dempster, Nan Laird και Donald Rubin σε επίσημο επιστημονικό περιοδικό της Royal Statistic Society. Οι προαναφερθέντες έδειξαν πως αν και η συγκεκριμένη μέθοδος είχε ήδη προταθεί και συζητηθεί πολλές φορές από άλλους συναδέλφους τους, η δική τους δημοσίευση γενίκευσε τον αλγόριθμο και υλοποίησε την όλη θεωρία που είχε αναπτυχθεί πάνω σε αυτόν [16].

2.2 Γενικά για τον EM

Ο EM αλγόριθμος έχει πολλές και διάφορες εφαρμογές. Καταρχάς, χρησιμοποιείται κατά κόρον στο data clustering αλλά και σε πρακτικές machine learning και computer vision της τεχνητής νοημοσύνης. Στην ψυχομετρία, ο EM είναι σχεδόν αναπόσπαστο κομμάτι στην εκτίμηση αντικειμενικών παραμέτρων αλλά και κρυφών ικανοτήτων στην μοντελοποίηση δεδομένων. Ο αλγόριθμος EM είναι επίσης ευρέως διαδεδομένος στην βιοπληροφορική με την έννοια της ανάλυσης και μοντελοποίησης εικόνων που αναπαριστούν βιολογικές δομές. Τα οφέλη που παρέχει ο EM στο background modelling είναι πολύ σημαντικά και γι' αυτό παρουσιάζεται αναλυτικά στο κεφάλαιο αυτό.

Με την ικανότητα που έχει να χειρίζεται δεδομένα και παραμέτρους χωρίς να έχουν οριστεί από τα δεδομένα μας, ο EM έχει γίνει ένα πολύ χρήσιμο εργαλείο ώστε

να εκτιμά και να χειρίζεται μη προφανείς για τον αναλυτή καταστάσεις που προκύπτουν.

Ένας αλγόριθμος expectation-maximization χρησιμοποιείται στην στατιστική για να υπολογίζει μέγιστες πιθανοφανείς εκτιμήσεις των παραμέτρων σε πιθανοτικά μοντέλα, τα οποία εξαρτώνται από απαρατήρητες κρυφές μεταβλητές. Ο EM εναλλάσσει την λειτουργία του μεταξύ του βήματος εκτέλεσης και πρόβλεψης (E-step), το οποίο υπολογίζει μια πρόβλεψη της πιθανότητας ενσωματώνοντας τις κρυφές μεταβλητές σαν να ήταν δοσμένες από την αρχή και του βήματος μεγιστοποίησης (M-step), το οποίο με τη σειρά του υπολογίζει την μέγιστη πιθανή εκτίμηση των παραμέτρων, μεγιστοποιώντας την προσδοκώμενη πιθανότητα που βρέθηκε στο E-step. Οι παράμετροι που βρέθηκαν στο M-step χρησιμοποιούνται έπειτα σαν δεδομένα εισόδου για να εκκινηθεί το επόμενο E-step και η διαδικασία επαναλαμβάνεται.

2.3 Προδιαγραφή της διαδικασίας του EM

Ας θεωρήσουμε σαν y τα μη πλήρη δεδομένα τα οποία αποτελούνται από παρατηρούμενες μεταβλητές και έστω z τα δεδομένα που λείπουν από την είσοδο. Μαζί τα z και y αποτελούν όλα τα δεδομένα. Το z μπορεί να είναι πραγματικές μετρήσεις που λείπουν από τα δεδομένα μας ή κρυμμένες μεταβλητές των οποίων οι τιμές θα έλυναν πολύ εύκολα το πρόβλημα, αν ήταν γνωστές. Για παράδειγμα, στα συνδυαστικά μοντέλα (mixture models) το αποτέλεσμα της πιθανοφάνειας θα ήταν πολύ πιο πειστικό αν τα συστατικά που «παρήγαγαν» τα δείγματα ήταν γνωστά από την αρχή.

Έστω p η υπό συνθήκη συνάρτηση πυκνότητας πιθανότητας (για συνεχή περίπτωση) ή η υπό συνθήκη συνάρτηση μάζας πιθανότητας (για διακριτή περίπτωση) των ολοκληρωμένων δεδομένων τις παραμέτρους που δίνονται από το διάλυσμα $\theta : p(z|y, \theta)$. Αυτή η συνάρτηση μπορεί επίσης να θεωρηθεί σαν την πιθανοφάνεια των συνολικών δεδομένων, γιατί μπορεί να ληφθεί υπ' όψιν σαν συνάρτηση του θ . Επιπλέον, ας σημειωθεί ότι η υπό συνθήκη κατανομή των δεδομένων που λείπουν από την παραπάνω πιθανότητα μπορεί να γραφτεί ως :

$$p(\mathbf{z}|\mathbf{y}, \theta) = \frac{p(\mathbf{y}, \mathbf{z}|\theta)}{p(\mathbf{y}|\theta)} = \frac{p(\mathbf{y}|\mathbf{z}, \theta)p(\mathbf{z}|\theta)}{\int p(\mathbf{y}|\hat{\mathbf{z}}, \theta)p(\hat{\mathbf{z}}|\theta)d\hat{\mathbf{z}}}$$

χρησιμοποιώντας τον κανόνα του Bayes και αυτόν της ολικής πιθανότητας. Το παραπάνω απαιτεί μόνο την γνώση της πιθανότητας των παρατηρούμενων δεδομένων y δεδομένου των δεδομένων που λείπουν $p(y|z, \theta)$ αλλά και της πιθανότητας των δεδομένων που λείπουν $p(z|\theta)$.

Ο EM αλγόριθμος επαναληπτικά βελτιώνει μια αρχική εκτίμηση θ_0 υπολογίζοντας συνεχώς νέες θ_1 , θ_2 κλπ. Ο υπολογισμός μιας επόμενης εκτίμησης από την προηγούμενη έχει την εξής μορφή :

$$\theta_{n+1} = \arg \max_{\theta} Q(\theta)$$

όπου το $Q(\theta)$ είναι η αναμενόμενη τιμή της λογαριθμικής πιθανότητας που δίνεται από την (1) (βλ. παρακάτω).

Με άλλα λόγια, αν δεν είναι γνωστά τα συνολικά δεδομένα, ώστε να μην μπορούμε να ξέρουμε την ακριβή τιμή της πιθανοφάνειας, μπορούμε να υπολογίσουμε μια εκ των υστέρων εκτίμηση των τιμών των πιθανοτήτων των άγνωστων z , βάσει των δεδομένων που είναι γνωστά (y). Για κάθε ομάδα τιμών του z , υπάρχει μια τιμή πιθανοφάνειας για το θ , οπότε μπορούμε να υπολογίσουμε την προσδοκώμενη τιμή της πιθανοφάνειας [expected value of likelihood $Q(\theta)$] δοσμένων των τιμών y (οι οποίες εξαρτώνται από την προηγούμενη υποθετική τιμή του θ , επειδή αυτή επηρέασε τις πιθανότητες των z).

Το Q δίνεται από τον τύπο:

$$Q(\theta) = \sum_z p(z | y, \theta_n) \ln p(y, z | \theta) \quad (1)$$

ή πιο γενικά:

$$Q(\theta) = E_{\mathbf{z}} \left[\ln p(\mathbf{y}, \mathbf{z} | \theta) \mid \mathbf{y} \right]$$

που δηλώνει την υπό συνθήκη πρόβλεψη του $\ln[p(y,z|\theta)]^3$.

Με άλλα λόγια, το θ_{n+1} είναι η τιμή που μεγιστοποιεί (M) την υπό συνθήκη πρόβλεψη (E) της πιθανότητας των συνολικών δεδομένων, έχοντας σαν είσοδο μόνο κάποιες ήδη καταγεγραμμένες τιμές μεταβλητών. Η πρόβλεψη $Q(\theta)$ στην συνεχή της περίπτωση θα μπορούσε να περιγραφεί ως εξής :

$$Q(\theta) = E_z \left[\ln p(y, z | \theta) | y \right] = \int_{-\infty}^{\infty} p(z | y, \theta_n) \ln p(y, z | \theta) dz$$

Ο όρος «βήμα πρόβλεψης», δεν είναι απόλυτα σωστός σε αυτή τη φάση. Αυτό το οποίο υπολογίζεται στο πρώτο βήμα είναι οι σταθερές, εξαρτώμενες από τα δεδομένα, παράμετροι θ της συνάρτησης Q . Μόλις αυτές γίνουν γνωστές, το Q είναι πλήρως προσδιορισμένο και μεγιστοποιείται στο επόμενο βήμα (M) του EM αλγορίθμου.

Η ονομασία EM στην πραγματικότητα, προέρχεται από τους συγγραφείς του συγκεκριμένου paper που αναφέρθηκαν προηγουμένως, οι οποίοι είχαν περιγράψει ένα λιγότερο γενικό πρόβλημα στο οποίο η πιθανοτική κατανομή είχε εκθετική μορφή (e^x) και έτσι δημιουργήθηκε το E-step. Όμως, και σε αυτή την περίπτωση το E-step είχε την ευθύνη του υπολογισμού των αναμενόμενων τιμών στατιστικών δεδομένων για τα πλήρη δεδομένα, χωρίς αυτά να είναι δοσμένα από την αρχή [16].

2.4 Ιδιότητες του EM

Σημαντικός λόγος που η χρήση του αλγορίθμου EM εξαπλώθηκε πολύ γρήγορα στους ερευνητές είναι, όπως μπορεί να δειχθεί, ότι κάθε μία επανάληψη του δεν μειώνει την συνάρτηση πιθανοφάνειας των δεδομένων που έχουμε στα χέρια από την πρώτη στιγμή. Ωστόσο, πρέπει να αναφερθεί ότι δεν υπάρχει εγγύηση, ότι η ακολουθία συγκλίνει στην μέγιστη δυνατή πιθανοφάνεια. Αυτό σημαίνει ότι σε πολυτροπικές κατανομές (multimodal distributions), ο EM αλγόριθμος μπορεί να

³ Η χρήση της λογαριθμικής πιθανότητας είναι πιο συχνή από αυτή της πραγματικής γιατί οδηγεί σε ευκολότερους τύπους και επιπλέον, η χρήση αυτών πετυχαίνει την μεγιστοποίηση της στο ίδιο σημείο με αυτό της πραγματικής.

συγκλίνει σε κάποιο τοπικό μέγιστο ή σε κάποιο saddle-point⁴ της συνάρτησης πιθανοφάνειας των γνωστών δεδομένων. Αυτό εξαρτάται από τις αρχικές τιμές.

Γι' αυτόν τον σκοπό, έχει δημιουργηθεί μια πληθώρα ευριστικών προσεγγίσεων, ώστε να αποφεύγονται τοπικά μέγιστα όπως για παράδειγμα να δοκιμάζονται διάφορες αρχικές εκτιμήσεις θ_0 ή άλλες μέθοδοι εύρεσης ολικών βέλτιστων [16].

Γενικά, Expectation – Maximization είναι μια περιγραφή μιας κλάσης αλγορίθμων, και όχι ενός συγκεκριμένου. Μπορούμε να πούμε ότι είναι μια συνταγή ή ένας μετα-αλγόριθμος που χρησιμοποιείται για να σχεδιαστούν πάνω σε αυτόν πολλοί αλγόριθμοι. Ο αλγόριθμος Baum-Welch είναι ένα παράδειγμα τέτοιου αλγορίθμου, ο οποίος εφαρμόζεται πάνω σε Κρυφά Μαρκοβιαννά Μοντέλα (Hidden Markov Models).

Ένας EM αλγόριθμος μπορεί επίσης να υπολογίσει την μέγιστη εκ των υστέρων (maximum a posteriori - MAP) πιθανότητα εφαρμόζοντας εκτίμηση MAP στο M-step, αντί για μέγιστη πιθανοφάνεια.

Φυσικά, υπάρχουν και άλλες μέθοδοι για τον υπολογισμό της μέγιστης πιθανοφάνειας, όπως οι μέθοδοι gradient descent, conjugate gradient ή παραλλαγές της μεθόδου του Gauss-Newton. Σε αντίθεση όμως με τον EM, αυτές οι μέθοδοι συνήθως απαιτούν τον εξ' αρχής υπολογισμό της πρώτης ή και δεύτερης παραγώγου της συνάρτησης πιθανότητας.

2.5 Παράδειγμα του EM: Mixture of Gaussian

Υποθέτουμε ότι (στη μονοδιάστατη περίπτωση) έχουμε ένα δείγμα από διανύσματα y_1, y_2, \dots, y_m , όπου κάθε $y_j \in \mathcal{R}^l$, έχει δημιουργηθεί από μία gaussian κατανομή διαστάσεως l . Συνολικά υπάρχουν n gaussian κατανομές. Έστω ότι το z_k δηλώνει από ποιά gaussian κατανομή προήλθε το y_j , δηλαδή το k παίρνει τιμές από

⁴ Όταν αναφερόμαστε σε saddle-point εννοούμε ένα σημείο, για παράδειγμα δύο διαστάσεων $w_0 \in W$, $z_0 \in Z$, το οποίο ικανοποιεί την εξής ανισότητα για μια συγκεκριμένη συνάρτηση

$$f(\cdot): f(w_0, z) \leq f(w_0, z_0) \leq f(w, z_0), \forall w \in W, z \in Z.$$

το 1 μέχρι το n . Η πιθανότητα να προέρχεται ένα μονοδιάστατο y_j από την k -οστή gaussian κατανομή ($1 \leq k \leq n$) είναι:

$$p(y|z=i, \theta) = (2\pi)^{-1/2} |\sigma_i|^{-1/2} \exp\left(-\frac{1}{2} (y - \mu_i)^T \sigma_i^{-1} (y - \mu_i)\right)$$

για $y, \mu_i, \sigma_i \in \mathbb{R}$

και γενικά:

Σκοπός μας είναι να εκτιμηθούν οι άγνωστοι παράμετροι $\theta = \{(\mu_1, \dots, \mu_n), (\sigma_1, \dots, \sigma_n), P(z=1 \dots n)\}$, οι οποίοι είναι οι μέσοι και οι σταθερές αποκλίσεις⁵ για κάθε gaussian κατανομή, όπως επίσης και η πιθανότητες χρήσης της κάθε gaussian κατανομής για κάθε σημείο.

Βήμα της Πρόβλεψης (E-step):

Η εκτίμηση των μη δεδομένων z (για το ποια gaussian κατανομή i με $i=1, \dots, n$ έχει χρησιμοποιηθεί), υπό συνθήκη των δεδομένων y_j , θ_l όπου $j=1, \dots, n$ χρησιμοποιώντας τις τιμές τους από το προηγούμενο βήμα μεγιστοποίησης (M-step) l είναι :

$$p(z_j = i | y_j, \theta_l) = \frac{p(z_j = i, y_j | \theta_l)}{p(y_j | \theta_l)} = \frac{p(y_j | z_j = i, \theta_l) p(z_j = i | \theta_l)}{\sum_{k=1}^n p(y_j | z_j = k, \theta_l) p(z_j = k | \theta_l)}$$

$$1 \leq i, j \leq n.$$

Βήμα της Μεγιστοποίησης (M-step):

Σε αυτό το βήμα πρέπει να μεγιστοποιήσουμε την προσδοκώμενη λογαριθμική πιθανοφάνεια του από κοινού γεγονότος:

⁵ Στην πραγματικότητα, δεν είναι ξεκάθαρο αν θα πρέπει να επιτρέπουμε στις αποκλίσεις να πάρουν οποιαδήποτε τιμή, γιατί τότε η μέγιστη πιθανοφάνεια μπορεί να μη φράσσεται, καθώς επικεντρωνόμαστε σε μια συγκεκριμένη gaussian κατανομή σε ένα συγκεκριμένο σημείο και μειώνεται η τιμή της απόκλισης προς το 0.

$$\begin{aligned}
Q(\theta) &= E_z \left[\ln \prod_{j=1}^m p(\mathbf{y}_j, \mathbf{z} | \theta) \mid \mathbf{y}_j \right] \\
&= E_z \left[\sum_{j=1}^m \ln p(\mathbf{y}_j, \mathbf{z} | \theta) \mid \mathbf{y}_j \right] \\
&= \sum_{j=1}^m E_z \left[\ln p(\mathbf{y}_j, \mathbf{z} | \theta) \mid \mathbf{y}_j \right] \\
&= \sum_{j=1}^m \sum_{i=1}^n p(z_j = i | \mathbf{y}_j, \theta) \ln p(z_j = i, \mathbf{y}_j | \theta)
\end{aligned}$$

και αφού $p(\mathbf{z}, \mathbf{y} | \theta) = p(\mathbf{y} | \mathbf{z}, \theta) p(\mathbf{z} | \theta)$ ο παραπάνω τύπος γίνεται:

$$Q(\theta) = \sum_{j=1}^m \sum_{i=1}^n p(z_j = i | \mathbf{y}_j, \theta) \ln (p(\mathbf{y}_j | z_j = i, \theta) p(z_j = i | \theta))$$

Επίσης, έχουμε τον περιορισμό:

$$\sum_{i=1}^n p(z_j = i | \theta) = 1$$

Αν προσθέσουμε τώρα έναν πολλαπλασιαστή Lagrange για τον περιορισμό και αναπτύξουμε την συνάρτηση πυκνότητας πιθανότητας για μονοδιάστατες gaussian μεταβλητές έχουμε:

$$\begin{aligned}
\mathcal{L}(\theta) &= \left(\sum_{j=1}^m \sum_{i=1}^n p(z_j = i | \mathbf{y}_j, \theta) \left(-\frac{1}{2} \ln(2\pi) - \frac{1}{2} \ln |\sigma_i| - \frac{1}{2} (\mathbf{y}_j - \mu_i) \right. \right. \\
&\quad \left. \left. \sigma_i^{-1} (\mathbf{y}_j - \mu_i) + \ln p(z_j = i | \theta) \right) \right) - \lambda \left(\sum_{i=1}^n p(z_j = i | \theta) - 1 \right)
\end{aligned}$$

Για να βρούμε την νέα εκτίμηση θ_{t+1} , αρκεί να βρούμε το μέγιστο της $\mathcal{L}(\theta)$, όπου η παράγωγος της Λαπλασιανής μηδενίζεται:

:

$$\frac{\partial \mathcal{L}(\theta)}{\partial \theta} = \mathbf{0}$$

όπου η παράμετρος θ περιλαμβάνει τους μέσους μ_i την τυπική απόκλιση σ_i και την $p(z|\theta)$ πιθανότητα κατηγοριοποίησης. Η νέα εκτίμηση ως προς τους μέσους χρησιμοποιώντας κανόνες διαφορίσης είναι:

$$\begin{aligned} \frac{\partial \mathcal{L}(\theta)}{\partial \mu_i} &= \sum_{j=1}^m p(z_j = i | \mathbf{y}_j, \theta_t) \left(-\frac{\partial}{\partial \mu_i} \frac{1}{2} (\mathbf{y}_j - \mu_i)^T \sigma_i^{-1} (\mathbf{y}_j - \mu_i) \right) \\ &= \sum_{j=1}^m p(z_j = i | \mathbf{y}_j, \theta_t) \left(-\frac{1}{2} (\sigma_i^{-1} + \sigma_i^{-T}) (\mathbf{y}_j - \mu_i) (-1) \right) \\ &= \sum_{j=1}^m p(z_j = i | \mathbf{y}_j, \theta_t) (\sigma_i^{-1} (\mathbf{y}_j - \mu_i)) \\ &= 0 \end{aligned}$$

⇓

$$\sum_{j=1}^m p(z_j = i | \mathbf{y}_j, \theta_t) \sigma_i^{-1} \mu_i = \sum_{j=1}^m p(z_j = i | \mathbf{y}_j, \theta_t) \sigma_i^{-1} \mathbf{y}_j$$

⇓

$$\mu_i \sum_{j=1}^m p(z_j = i | \mathbf{y}_j, \theta_t) = \sum_{j=1}^m p(z_j = i | \mathbf{y}_j, \theta_t) \mathbf{y}_j$$

⇓

$$\mu_i = \frac{\sum_{j=1}^m p(z_j = i | \mathbf{y}_j, \theta_t) \mathbf{y}_j}{\sum_{j=1}^m p(z_j = i | \mathbf{y}_j, \theta_t)}$$

Η νέα εκτίμηση ως προς την απόκλιση είναι:

$$\begin{aligned}
\frac{\partial \mathcal{L}(\theta)}{\partial \sigma_i} &= \sum_{j=1}^m p(z_j = i | \mathbf{y}_j, \theta_t) \\
&\quad \left(-\frac{\partial}{\partial \sigma_i} \frac{1}{2} \ln |\sigma_i| - \frac{\partial}{\partial \sigma_i} \frac{1}{2} (\mathbf{y}_j - \mu_i)^T \sigma_i^{-1} (\mathbf{y}_j - \mu_i) \right) \\
&= \sum_{j=1}^m p(z_j = i | \mathbf{y}_j, \theta_t) \\
&\quad \left(-\frac{1}{2} \sigma_i^{-T} + \frac{1}{2} \sigma_i^{-T} (\mathbf{y}_j - \mu_i) (\mathbf{y}_j - \mu_i)^T \sigma_i^{-T} \right) \\
&= 0 \\
&\Downarrow
\end{aligned}$$

$$\begin{aligned}
&\sum_{j=1}^m p(z_j = i | \mathbf{y}_j, \theta_t) \sigma_i^{-1} \\
&= \sum_{j=1}^m p(z_j = i | \mathbf{y}_j, \theta_t) \sigma_i^{-1} (\mathbf{y}_j - \mu_i) (\mathbf{y}_j - \mu_i)^T \sigma_i^{-1}
\end{aligned}$$

\Downarrow

$$\begin{aligned}
&\sum_{j=1}^m p(z_j = i | \mathbf{y}_j, \theta_t) \\
&= \sum_{j=1}^m p(z_j = i | \mathbf{y}_j, \theta_t) \sigma_i^{-1} (\mathbf{y}_j - \mu_i) (\mathbf{y}_j - \mu_i)^T
\end{aligned}$$

\Downarrow

$$\sigma_i = \frac{\sum_{j=1}^m p(z_j = i | \mathbf{y}_j, \theta_t) (\mathbf{y}_j - \mu_i) (\mathbf{y}_j - \mu_i)^T}{\sum_{j=1}^m p(z_j = i | \mathbf{y}_j, \theta_t)}$$

Η νέα εκτίμηση ως προς την πιθανότητα της κατηγοριοποίησης (για το ποια gaussian κατανομή χρησιμοποιήθηκε) είναι:

$$\begin{aligned} \frac{\partial \mathcal{L}(\theta)}{\partial p(z_j = i|\theta)} &= \left(\sum_{j=1}^m p(z_j = i|\mathbf{y}_j, \theta_t) \frac{\partial \ln p(z_j = i|\theta)}{\partial p(z_j = i|\theta)} \right) \\ &\quad - \lambda \left(\frac{\partial p(z_j = i|\theta)}{\partial p(z_j = i|\theta)} \right) \\ &= \left(\sum_{j=1}^m p(z_j = i|\mathbf{y}_j, \theta_t) \frac{1}{p(z_j = i|\theta)} \right) - \lambda \\ &= 0 \end{aligned}$$

↓

$$\sum_{j=1}^m p(z_j = i|\mathbf{y}_j, \theta_t) \frac{1}{p(z_j = i|\theta)} = \lambda$$

↓

$$p(z_j = i|\theta) = \frac{1}{\lambda} \sum_{j=1}^m p(z_j = i|\mathbf{y}_j, \theta_t)$$

Αθροίζοντας για όλα τα i , έχουμε τον περιορισμό της μονάδας, οπότε:

$$\begin{aligned} \sum_{i=1}^n p(z_j = i|\theta) &= \sum_{i=1}^n \frac{1}{\lambda} \sum_{j=1}^m p(z_j = i|\mathbf{y}_j, \theta_t) \\ &= 1 \end{aligned}$$

↓

$$\lambda = \sum_{i=1}^n \sum_{j=1}^m p(z_j = i|\mathbf{y}_j, \theta_t)$$

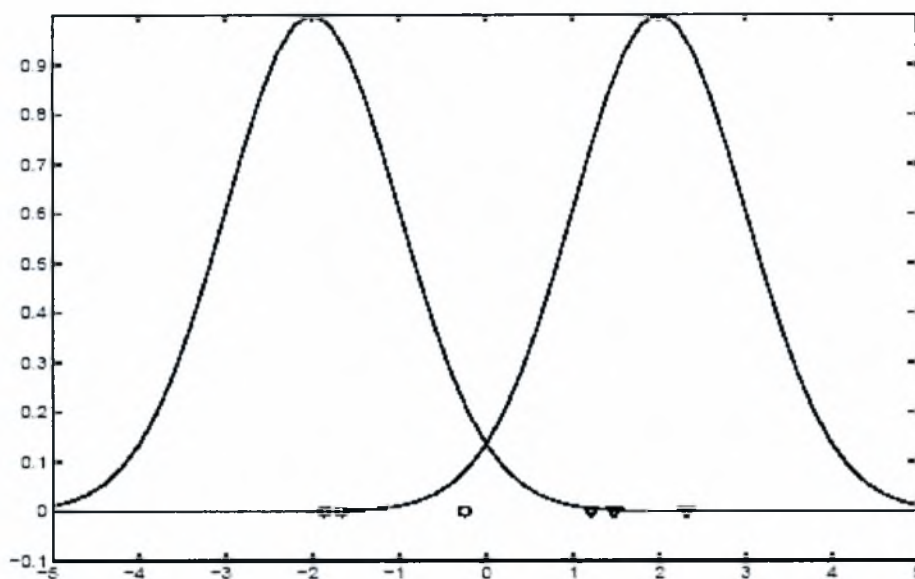
και αντικαθιστώντας το λ στην εκτίμηση μας παίρνουμε:

$$\begin{aligned} p(z_j = i|\theta) &= \frac{1}{\lambda} \sum_{j=1}^m p(z_j = i|\mathbf{y}_j, \theta_t) \\ &= \frac{\sum_{j=1}^m p(z_j = i|\mathbf{y}_j, \theta_t)}{\sum_{k=1}^n \sum_{j=1}^m p(z_j = k|\mathbf{y}_j, \theta_t)} \end{aligned}$$

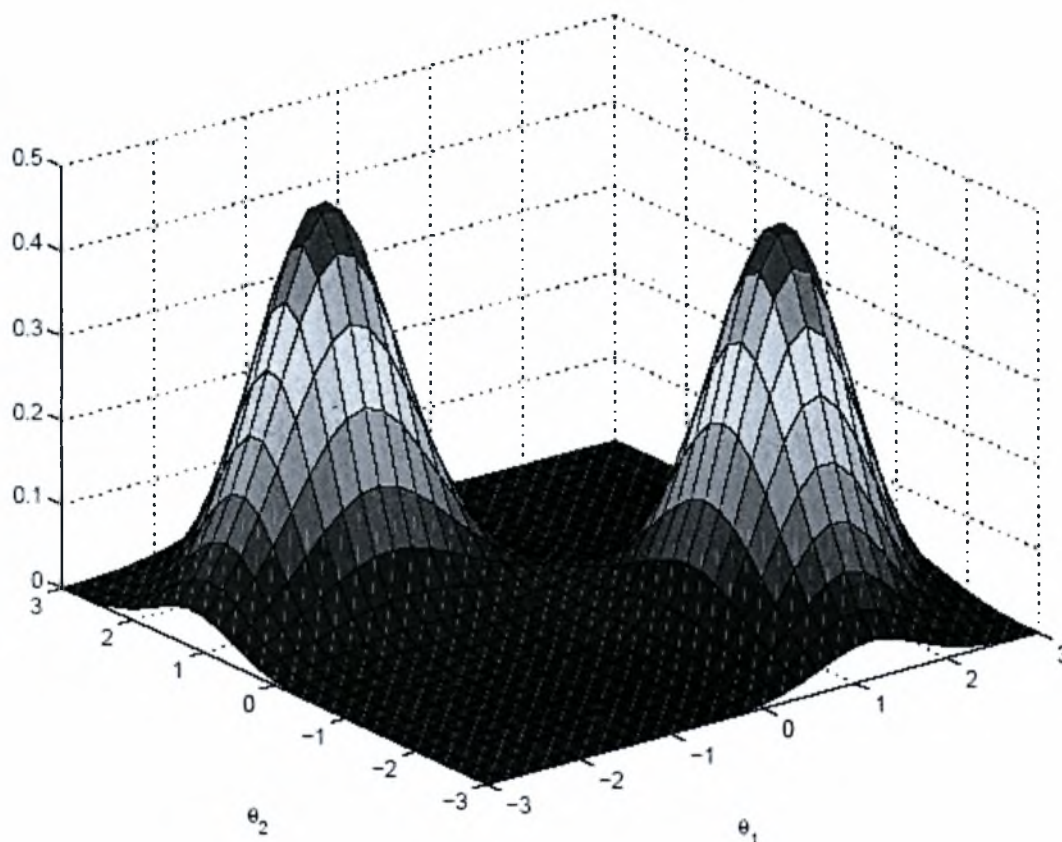
Αυτές οι εκτιμήσεις πλέον είναι το νέο μας θ_{t+1} το οποίο θα χρησιμοποιηθεί στο επόμενο E-step.

2.6 Πιο Αναλυτικά για τον EM

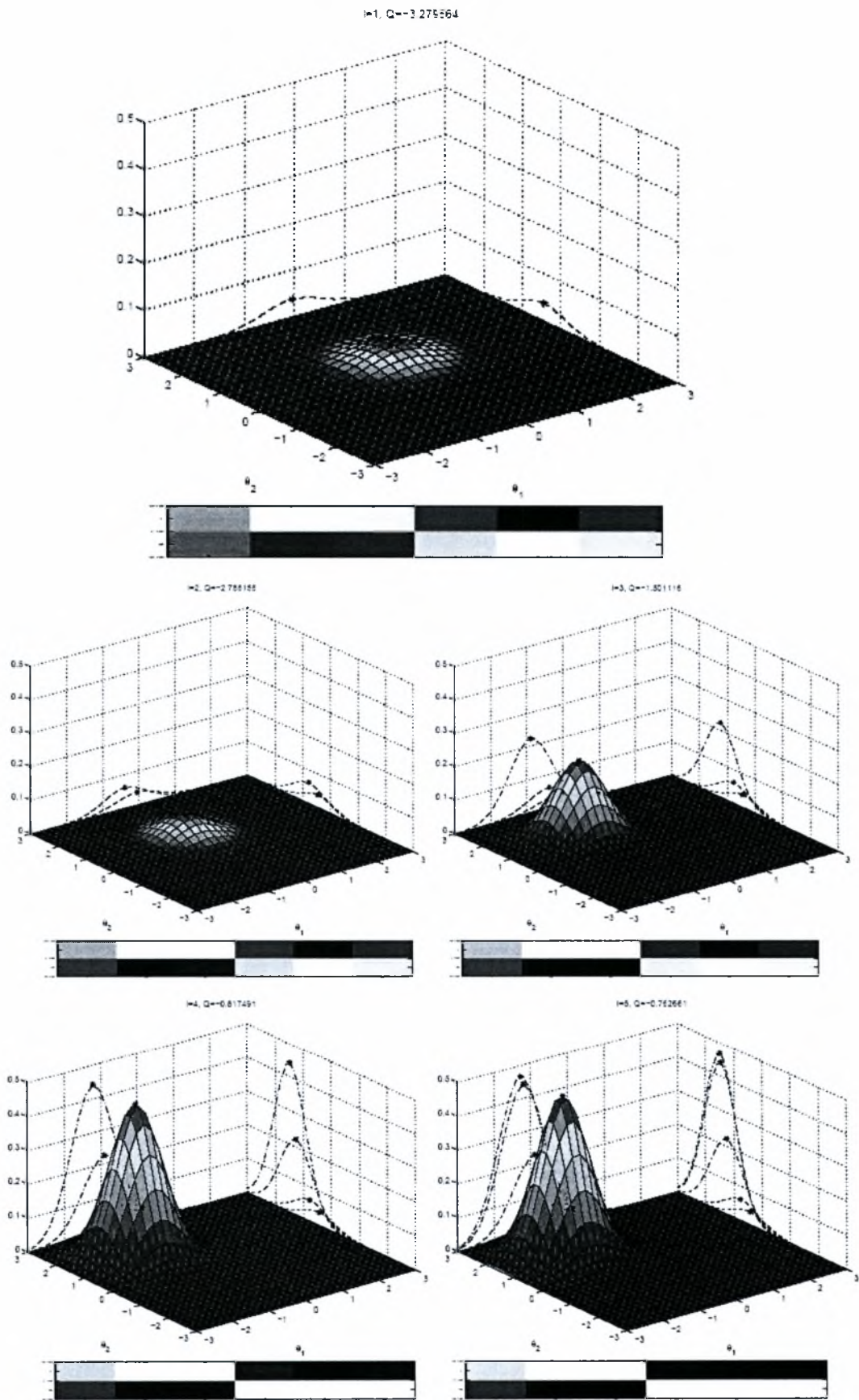
Μιλώντας για μεγιστοποίηση των κάτω ορίων μπορεί να δοθεί μία βαθύτερη εξήγηση του αλγορίθμου του EM, η οποία παρέχει σε μεγαλύτερο επίπεδο την κατανόηση της λειτουργίας του, από ότι αντιλαμβάνεται κανείς σκεπτόμενος απλά και διαισθητικά την εναλλαγή μεταξύ μεταβλητών. Σε αυτήν την λογική, το E-step μπορεί να ερμηνευτεί σαν μια διαδικασία κατασκευής ενός τοπικού κάτω ορίου της εκ των υστέρων κατανομής (με το M-step να βελτιστοποιεί αυτό το όριο). Κατ' αυτόν τον τρόπο, βελτιώνεται η εκτίμηση των αγνώστων παραμέτρων. Αυτό δείχνεται παρακάτω με ένα απλό παράδειγμα.



Εικόνα 1: Παράδειγμα EM : Πολλαπλές κατανομές και δεδομένα. Τα δεδομένα αποτελούνται από τρία δείγματα για κάθε κατανομή, τα οποία φαίνονται από τους κύκλους και τα τρίγωνα αντίστοιχα σε κάθε μία. Οι μέσοι των συνιστωσών είναι αντίστοιχα -2 και 2 [10].



Εικόνα 2: Η πραγματική συνάρτηση πιθανότητας των 2 συνιστωσών μέσω θ_1 και θ_2 , με δεδομένα εισόδου της Εικόνας 1 [10].

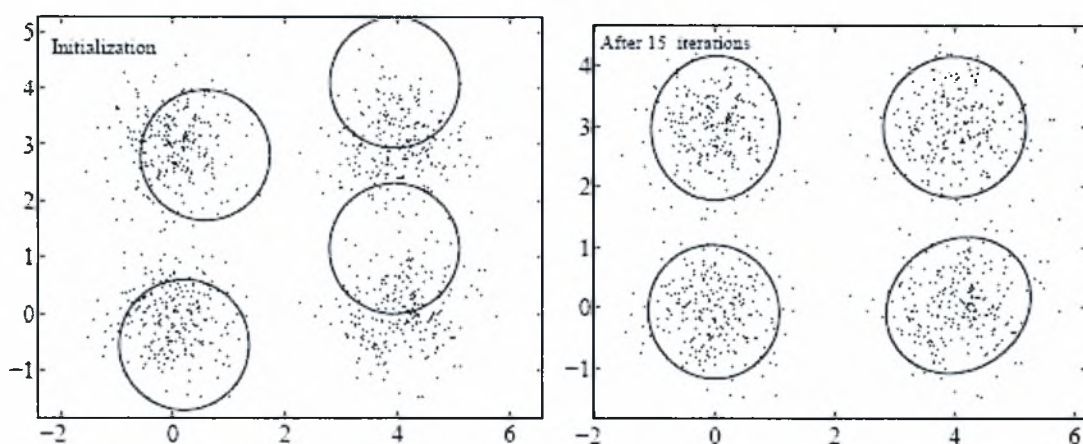


Εικόνα 3: Κάτω όρια [10].

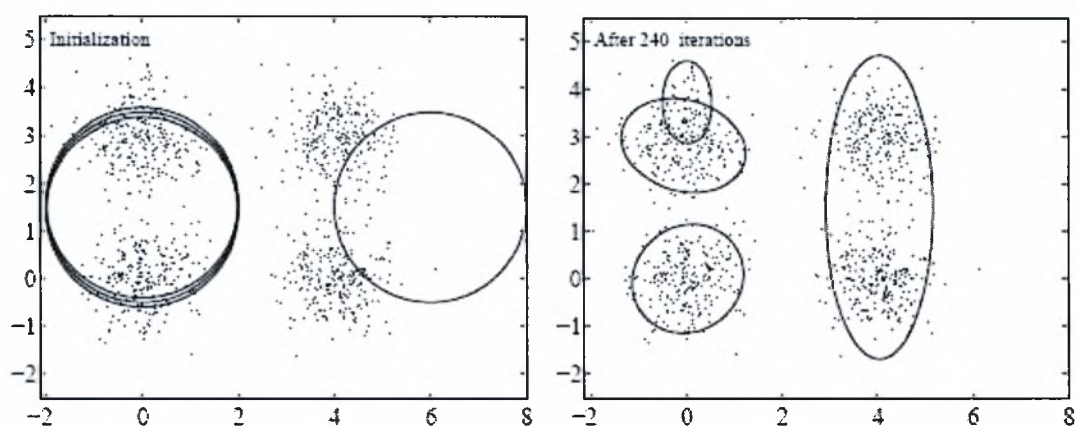
Ας θεωρήσουμε το πρόβλημα εκτίμησης που παρουσιάζεται στην Εικόνα 1. Ο στόχος είναι να εκτιμήσουμε τους δύο μέσους συνιστωσών θ_1 και θ_2 δεδομένων των έξι δειγμάτων που παρατηρούμε, χωρίς όμως να ξέρουμε από ποια κατανομή έχει προέλθει το κάθε δείγμα. Το πεδίο ορισμού είναι δύο διαστάσεων και η πραγματική συνάρτηση πιθανοφάνειας φαίνεται στην Εικόνα 2. Παρατηρούμε, ότι υπάρχουν δύο κορυφές στα σημεία $(-2,2)$ και $(2,-2)$ αντίστοιχα. Αυτό είναι απόλυτα λογικό αφού μπορούμε να εναλλάσσουμε τις συνιστώσες χωρίς αυτό να έχει την παραμικρή επίδραση στην ποιότητα της λύσης. Αξίζει να σημειωθεί επίσης, πως η πραγματική πιθανοφάνεια υπολογίζεται ολοκληρώνοντας πάνω σε όλες τις πιθανές συσχετίσεις των δεδομένων (y) και έτσι μπορούμε να καταλήξουμε σε μία λύση μέγιστης πιθανοφάνειας. Ωστόσο, ακόμα και μόνο για έξι δείγματα, δημιουργείται ένας χώρος από 64 πιθανές συσχετίσεις μεταξύ τους.

Ο EM ακολουθεί την εξής διαδικασία σε αυτό το παράδειγμα. Στο E-step, υπολογίζεται μια πρώτη, και όχι τόσο ακριβής, εκ των υστέρων πιθανότητα $[p(z|y,\theta)]$ για κάθε πιθανή συσχέτιση, για κάθε δείγμα που έχουμε. Στο συγκεκριμένο παράδειγμα, έχουμε 2 συνιστώσες και 6 δείγματα, άρα οι υπολογιζόμενες πιθανότητες $[p(z_{1,2} | y_{1,\dots,6}, \theta)]$ μπορούν να αναπαρασταθούν σε έναν πίνακα 2×6 . Έχοντας υπολογίσει αυτές τις πιθανότητες, ο EM έπειτα υπολογίζει ένα αυστηρό κάτω όριο της πραγματικής συνάρτησης πιθανοφάνειας $[Q(\theta)]$, που φαίνεται στην Εικόνα 2. Το όριο αυτό είναι τέτοιο ώστε να «αγγίζει» την συνάρτηση πιθανοφάνειας στην τρέχουσα εκτίμηση και είναι κοντά στην συνάρτηση αυτή μόνο σε μια γειτονιά αυτής της εκτίμησης. Το όριο και ο αντίστοιχος πίνακας πιθανοτήτων υπολογίζεται σε κάθε επανάληψη της μεθόδου όπως φαίνεται στην Εικόνα 3. Στην συγκεκριμένη περίπτωση ο EM έτρεξε για πέντε επαναλήψεις. Στο M-step το κάτω όριο μεγιστοποιείται (στην Εικόνα 3 αυτό φαίνεται με μαύρο αστερίσκο) και η αντίστοιχη νέα εκτίμηση, (θ_1, θ_2) , εγγυημένα θα βρεθεί πιο κοντά στην περιοχή του κοντινότερου τοπικού μεγίστου της πιθανοφάνειας. Κάθε επόμενο όριο είναι συνεχώς μια βελτίωση της προσέγγισης της κορυφής της πιθανοφάνειας μέχρι να επέλθει η σύγκλιση και το όριο να την αγγίζει στο τοπικό μέγιστο, όπου και δεν μπορεί να γίνει περαιτέρω βελτίωση. Αυτό φαίνεται στα τελευταία δύο σχήματα της Εικόνας 3.

Ένα από τα βασικά μειονεκτήματα του EM είναι, ότι εξαρτάται πολύ από τις τιμές αρχικοποίησης του. Αυτό γίνεται πιο φανερό όταν αναφερόμαστε σε μοντέλα πολλών συνιστωσών, των οποίων η συνάρτηση πιθανοφάνειας έχει πολλά τοπικά μέγιστα. Αυτά εμφανίζονται όταν έχουμε πολλά συστατικά σε μία συγκεκριμένη περιοχή στον χώρο της συνάρτησης και πολύ λιγότερα αλλού. Ο EM δεν είναι ικανός να μετακινεί συστατικά από την μία περιοχή σε άλλη περνώντας διαμέσου περιοχών με πολύ μικρές πιθανοφάνειες. Αυτό φαίνεται στις Εικόνες 4 και 5, στις οποίες απεικονίζεται ένα παράδειγμα επιτυχούς αρχικοποίησης και ένα παράδειγμα ανεπιτυχούς αρχικοποίησης αντίστοιχα.



Εικόνα 4: Παράδειγμα επιτυχής εκτέλεσης του EM [8].



Εικόνα 5: Παράδειγμα ανεπιτυχής εκτέλεσης του EM λόγω λανθασμένης αρχικοποίησης του [8].

Κεφάλαιο 3 - Παρουσίαση Βασικών Μεθόδων



3.1 Μοντελοποίηση Προσαρμόσιμου Έγχρωμου Παρασκήνιου για Ροές Δεδομένων Video Πραγματικού Χρόνου

3.1.1 Γενικά

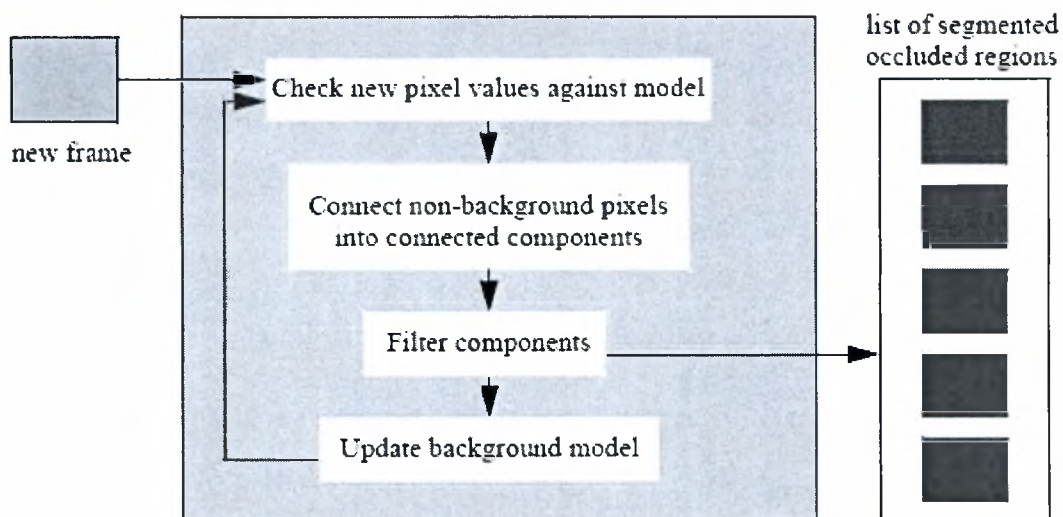
Σε αυτό το πεδίο θα παρουσιαστεί μία συγκεκριμένη, χαρακτηριστική μέθοδος εφαρμογής του adaptive background modelling για την τμηματοποίηση-κατάτμηση (segmentation) δεδομένων video stream. Η μέθοδος αυτή αφορά ένα σύστημα που χρησιμοποιεί ένα στατιστικό μοντέλο για το παρασκήνιο για να εντοπίζει αυτόματα τις αλλαγές του σκηνικού.

Αυτό που συμβαίνει είναι ότι τα frames που καταγράφονται από μια σταθερά τοποθετημένη βιντεοκάμερα μοντελοποιούνται, δεδομένου ότι έχουμε ένα σταθερό ή αργά μεταβαλλόμενο παρασκήνιο. Κάθε frame αποτελείται από το background και από κάποια αντικείμενα στο προσκήνιο που παρεμβάλλονται συχνά και θεωρούνται φραγμένες (συνεκτικές) περιοχές που πρέπει να εντοπιστούν και να διαχωριστούν από αυτό. Η διαδικασία ανίχνευσης περιλαμβάνει την δυναμική εκμάθηση του background αλλά και την τμηματοποίηση των «κλειστών» (συνεκτικών) περιοχών. Έχει επιλεγεί να χρησιμοποιηθεί ο χώρος χρωμάτων HSV (Hue, Saturation, Value) ή όπως αλλιώς αναφέρεται HSB (Hue, Saturation, Brightness) αντί του συνηθισμένου χώρου RGB (Red-Green-Blue), γιατί παρέχει μια καλύτερη χρήση της πληροφορίας των χρωμάτων όπως εξηγείται σε επόμενο πεδίο.

Όσον αφορά τα τυχαία αντικείμενα στο προσκήνιο δεν γίνεται καμία υπόθεση για το μοντέλο αναπαράστασης τους.

Οι τιμές των pixels του background μοντελοποιούνται ως πολυδιάστατες gaussian κατανομές στον χώρο του HSV. Όταν ένα νέο frame έρχεται στα δεδομένα, οι τιμές των pixels συγκρίνονται με τις τρέχουσες αντίστοιχες κατανομές για να αποφασιστεί αν αποτελούν μέρος του background ή μέρος ενός αντικειμένου στο προσκήνιο. Στην δεύτερη φάση, τα pixels αυτά συγκεντρώνονται και ενώνονται σε συνεκτικές συνιστώσες. Το αποτέλεσμα είναι μια λίστα από περιοχές που αποτελούν τα τμήματα του προσκηνίου. Το μέγεθος των περιοχών αυτών βέβαια ελέγχονται από μια τιμή κατωφλίου για την περίπτωση που εσφαλμένα έχουν θεωρηθεί αντικείμενα του προσκηνίου. Οι κατανομή κάθε pixel του παρασκηίου βέβαια, ανανεώνεται με τις τελευταίες παρατηρήσεις για να ληφθούν υπ' όψη τυχόν αλλαγές σε αυτό.

Οι τμηματοποιημένες περιοχές που αναπαριστούν τα αντικείμενα του προσκηνίου μπορούν έπειτα να χρησιμοποιηθούν για ιχνηλάτηση (tracking) μετά από ενδεχόμενη περαιτέρω τμηματοποίηση σε συνιστώσες βασιζόμενες σε ανάλογα μοντέλα. Άλλες εφαρμογές είναι η συμπίεση, τροποποίηση και βελτίωση της ποιότητας του video. Παρακάτω φαίνεται η όλη διαδικασία που πρόκειται να περιγραφεί αναλυτικά.



Εικόνα 6: Μία επανάληψη του συστήματος ανίχνευσης αντικειμένων στο προσκήνιο [11].

3.1.1.1 Ο Χώρος Χρωμάτων HSV και η Σχέση του με τον RGB

Το μοντέλο HSV, όπως προαναφέρθηκε, ορίζει έναν χώρο χρωμάτων όσον αφορά 3 συνιστώσες:

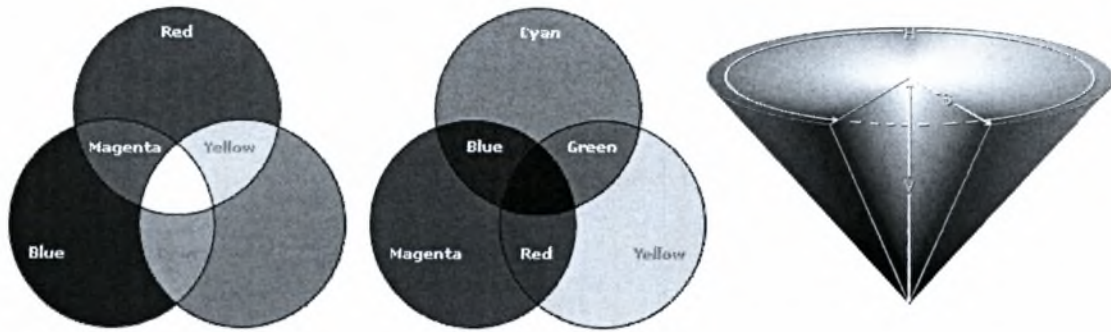
1. Η *απόχρωση (hue)* αντιπροσωπεύει το είδος του χρώματος (μπλε, κίτρινο, κλπ). Οι τιμές της απόχρωσης κυμαίνονται στο εύρος $[0, \dots, 360]$ για τις περισσότερες εφαρμογές, όπου κάθε τιμή αντιστοιχεί σε ένα μοναδικό χρώμα (για παράδειγμα το 0 αντιστοιχεί στο κόκκινο χρώμα).

2. Η *χρωματική καθαρότητα (saturation)* αντιστοιχεί στην ένταση του χρώματος. Είναι ποσοστιαία τιμή (0-100%), όπου το 0 σημαίνει ότι δεν υπάρχει χρώμα.

3. Η *λαμπρότητα (value, brightness)* του χρώματος είναι και αυτή ποσοστιαία τιμή, όπου το 0 αντιστοιχεί πάντα στο μαύρο χρώμα. Όσον αφορά μεγαλύτερες τιμές υπάρχει εξάρτηση από τη χρωματική καθαρότητα. Για παράδειγμα το 100 μπορεί να είναι το άσπρο, αλλά χωρίς πάντα σταθερή ένταση, ανάλογα με το ποσοστό του saturation.

Ο χώρος χρωμάτων του HSV προήλθε από τον RGB. Για την ακρίβεια είναι ένας μη γραμμικός μετασχηματισμός του RGB.

Το μοντέλο HSV προτιμάται συνήθως από άλλα, όπως το RGB ή το CMY(K) (Cyan, Magenta, Yellow, Key-Black), εξ' αιτίας της ομοιότητας του με τον τρόπο που οι άνθρωποι αντιλαμβάνονται τα χρώματα. Τα μοντέλα RGB και CMY(K) (Εικόνα 7) χρησιμοποιούν κάποια πρωτεύοντα και δευτερεύοντα χρώματα αντίστοιχα για να αναπαράγουν όλο το χρωματικό φάσμα, ενώ το HSV ενσωματώνει όρους, όπως η ένταση και η λαμπρότητα, που είναι οικείοι για τους ανθρώπους. Αυτό δίνει μεγάλο πλεονέκτημα στους ερευνητές να αναπτύξουν πιο εύκολα συστήματα που επεξεργάζονται δεδομένα από τον χώρο HSV, αλλά και ευκολία στην ερμηνεία των αποτελεσμάτων τους.



Εικόνα 7: Οι χώροι χρωμάτων RGB, CMY(K) και HSV αντίστοιχα [17].

Παρακάτω φαίνεται η διαδικασία μετατροπής ενός pixel από τον χώρο χρωμάτων RGB στον HSV και αντίστροφα:

- Από RGB σε HSV: Έστω MAX η μέγιστη τιμή από τις [R, G, B] τιμές ενός pixel και MIN η ελάχιστη. Τότε:

$$H = \begin{cases} \text{undefined,} & \text{if } MAX = MIN \\ 60^\circ \times \frac{G-B}{MAX-MIN} + 0^\circ, & \text{if } MAX = R \\ & \text{and } G \geq B \\ 60^\circ \times \frac{G-B}{MAX-MIN} + 360^\circ, & \text{if } MAX = R \\ & \text{and } G < B \\ 60^\circ \times \frac{B-R}{MAX-MIN} + 120^\circ, & \text{if } MAX = G \\ 60^\circ \times \frac{R-G}{MAX-MIN} + 240^\circ, & \text{if } MAX = B \end{cases}$$

$$S = \begin{cases} 0, & \text{if } MAX = 0 \\ 1 - \frac{MIN}{MAX}, & \text{otherwise} \end{cases}$$

$$V = MAX$$

▪ Από HSV σε RGB:

$$H_i = \left\lfloor \frac{H}{60} \right\rfloor \bmod 6$$

$$f = \frac{H}{60} - H_i$$

$$p = V(1 - S)$$

$$q = V(1 - fS)$$

$$t = V(1 - (1 - f)S)$$

$$\text{if } H_i = 0 \rightarrow R = V, G = t, B = p$$

$$\text{if } H_i = 1 \rightarrow R = q, G = V, B = p$$

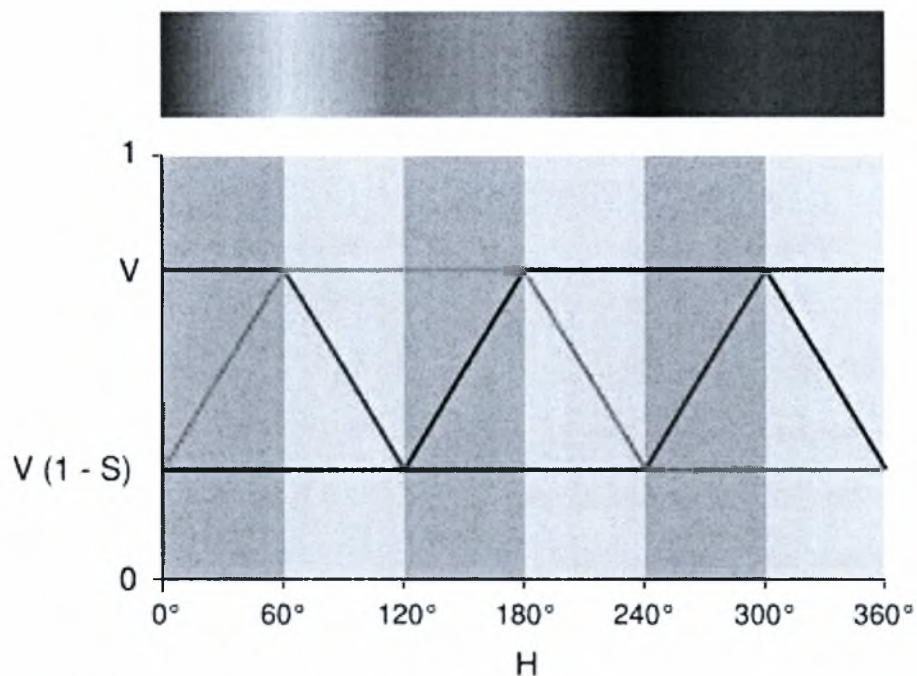
$$\text{if } H_i = 2 \rightarrow R = p, G = V, B = t$$

$$\text{if } H_i = 3 \rightarrow R = p, G = q, B = V$$

$$\text{if } H_i = 4 \rightarrow R = t, G = p, B = V$$

$$\text{if } H_i = 5 \rightarrow R = V, G = p, B = q$$

Έχει διαπιστωθεί πειραματικά σε διάφορες εφαρμογές, ότι το μοντέλο HSV προσθέτει καθυστέρηση στο σύστημα σε σχέση με αυτό του RGB, αλλά αυτό είναι το αντίτιμο της υψηλότερης ακρίβειας των αποτελεσμάτων του πρώτου. Στην παρακάτω εικόνα (Εικόνα 8) φαίνεται η σχέση μεταξύ των δύο χρωματικών χώρων.



Εικόνα 8: Σχέση μεταξύ του HSV και RGB χώρου χρωμάτων [17].

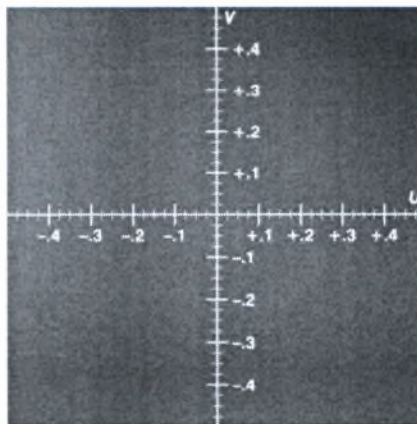
3.1.2 Μοντελοποίηση του Παρασκηνίου με το Μοντέλο Χρωμάτων HSV

Κάθε pixel του background μοντελοποιείται από τον ίδιο τύπο κατανομής. Αυτός ο τύπος είναι μία gaussian κατανομή, η οποία χαρακτηρίζεται από την μέση τιμή μ και την σταθερή απόκλιση σ . Θεωρητικά μπορεί να χρησιμοποιηθεί οποιοσδήποτε χώρος χρωμάτων για ανάλογα πειράματα αλλά όσον αφορά τον χώρο RGB, μπορούμε να πούμε ότι δεν «συμπεριφέρεται» ικανοποιητικά στην χρωματική αντίληψη, αφού η υπολογιζόμενη απόσταση μεταξύ δύο χρωμάτων στον RGB χώρο δεν αντικατοπτρίζει την αντιληπτική ομοιότητα που έχουν.

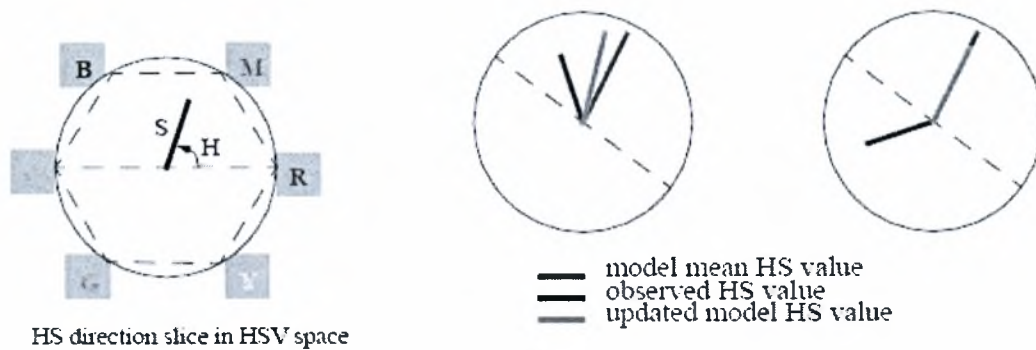
Ένα χρωματικό μοντέλο που καθαρά ξεχωρίζει την χρωματική ένταση από την χρωματική πληροφορία, επιτρέπει να θεωρήσουμε την χρωματικότητα σε μια δομή γενικεύοντας μοντέλα που αναπαριστούν μόνο την φυσική χρωματική ένταση.

Κάποια συστήματα χρησιμοποιούν τον YUV χώρο χρωμάτων (Εικόνα 9) το οποίο διαχωρίζει την ένταση (Y) από την χρωματικότητα (U,V). Ομοίως, το μοντέλο HSV διαχωρίζει την ένταση (V) από τις χρωματικές συνιστώσες (H,S), ωστόσο η αναπαράσταση της χρωματικότητας του υποχώρου UV, η οποία βασίζεται σε γραμμικούς συνδυασμούς των R, G, B συνιστωσών δεν είναι τόσο διαισθητικά αντιληπτή όσο η αναπαράσταση του κυκλικού υποχώρου HS (Εικόνα 10).

Στην υλοποίηση που ακολούθησε οι τιμές των pixels της εικόνας είναι τριπλέτες τιμών HSV. Η κατανομή ενός pixel του παρασκηνίου $B(x,y)$ είναι ένα διάνυσμα τριών μεταβλητών $H(x,y)$, $S(x,y)$ και $V(x,y)$, όπου η καθεμία χαρακτηρίζεται από μία μέση τιμή μ_k και μια σταθερή απόκλιση σ_k .



Εικόνα 9: Παράδειγμα του επιπέδου χρωμάτων YUV, για $Y=0.5$ αναπαριστώμενο στην γκάμα χρωμάτων του RGB [18].



Εικόνα 10: Ανανέωση της χρωματικής πληροφορίας [11].

Είναι φανερό ότι δεν υπάρχει κάποια αναλογία στις τρεις διαστάσεις και δεν μπορούν να επεξεργαστούν με τον ίδιο τρόπο. Η χρήση τιμών HSV απαιτεί περισσότερη επεξεργασία (μετατροπή των RGB μετρήσεων στον χώρο χρωμάτων του HSV και άλλες πολύπλοκες διαδικασίες), αλλά η σημερινή εξέλιξη της επεξεργαστικής δύναμης των υπολογιστών έχει καταστήσει το παραπάνω, πολύ μικρό τίμημα μπροστά στα αποτελέσματα που προσφέρει.

Οι κατανομές των pixels είναι αρχικά άγνωστες. Στην παρούσα υλοποίηση αρχικοποιούμε τις μέσες τιμές από τις τιμές των pixel του πρώτου frame και η σταθερή απόκλιση τίθεται ίση με το 0. Οι πραγματικές κατανομές θα υπολογιστούν από τα επόμενα frames που θα επέλθουν.

Μπορούμε να χωρίσουμε την διαδικασία επεξεργασίας κάθε νέου frame σε δύο βήματα: το πρώτο είναι η ανίχνευση των pixels που αναπαριστούν αντικείμενα του προσκηνίου και το δεύτερο, η ανανέωση του μοντέλου μας.

3.1.3 Ανίχνευση Pixels των Αντικειμένων του Προσκηνίου

Όταν ένα νέο frame έρχεται προς επεξεργασία, κάθε τιμή ενός pixel συγκρίνεται με την αντίστοιχη κατανομή μοντελοποίησης. Το αποτέλεσμα είναι ένα δυαδικό μαρκάρισμα σε κάθε pixel για το αν αποτελεί μέρος του παρασκηνίου ή πρόκειται για αντικείμενο του προσκηνίου.

Η απόφαση αυτή είναι ακόλουθο της έντασης και της χρωματικής πληροφορίας κάθε pixel. Για να θεωρηθεί ένα pixel σαν μέρος της εικόνας του παρασκηνίου πρέπει κάθε μία από τις τιμές των συνιστωσών H, S και V να βρίσκεται σε διάστημα

το πολύ δύο σταθερών αποκλίσεων⁶ της μέσης τιμής της συνιστώσας της αντίστοιχης κατανομής μοντελοποίησης [11].

3.1.4 Ανανέωση του Μοντέλου του Παρασκηνίου και Αποτελέσματα

Αφού έχουν ανιχνευθεί τα pixels που αντιπροσωπεύουν αντικείμενα στο προσκήνιο, οι τρέχουσες κατανομές πρέπει να ανανεωθούν ώστε να ενσωματώσουν τις τελευταίες πληροφορίες που έχουν ληφθεί. Έστω x η τιμή ενός παρατηρούμενου νέου pixel. Οι αντίστοιχη κατανομή του pixel του background ανανεώνεται σύμφωνα με τους παρακάτω γενικούς κανόνες:

$$\mu \leftarrow (1-\alpha)\mu + \alpha x,$$

$$\sigma^2 \leftarrow \max[\sigma_{min}^2, (1-\alpha)\sigma^2 + \alpha(x-\mu)^2]$$

όπου α είναι ο ρυθμός «εκμάθησης» (learning rate). Παρατηρούμε την εισαγωγή μιας ακόμα μεταβλητής, της σ_{min} , σαν ένα κατώφλι θορύβου, ώστε να εμποδίζει την σταθερή απόκλιση να μειωθεί κάτω από μία ελάχιστη τιμή, επειδή οι μετρήσεις για το background πρέπει να παραμένουν αυστηρά σταθερές για κάποια χρονική περίοδο. Αυτές οι παράμετροι εφαρμόζονται ολικά, δηλαδή σε κάθε ξεχωριστή κατανομή κάθε pixel του παρασκηνίου.

Η ανανέωση των διαφορετικών συνισταμένων κατανομών των pixels εξαρτάται από την διαθεσιμότητα και την συσχέτιση της χρωματικής πληροφορίας των παρατηρούμενων τιμών. Ο αλγόριθμος διακρίνει όλους τους συνδυασμούς χρωματικότητας/αχρωματικότητας για όλες τις παρατηρήσεις και τα μοντέλα. Στην περίπτωση όπου οι παρατηρούμενες και οι τιμές του μοντέλου είναι χρωματικές, αν η παρατήρηση και το μοντέλο έχουν πολύ διαφορετικές αποχρώσεις, η απόχρωση του μοντέλου δεν ανανεώνεται. Χρησιμοποιώντας την προβολή της παρατηρούμενης χρωματικής καθαρότητας στην κατεύθυνση της μέσης τιμής της απόχρωσης, βεβαιώνουμε ότι η μέση τιμή της χρωματικής καθαρότητας μειώνεται, αντικατοπτρίζοντας έτσι την αύξηση της αβεβαιότητας της πληροφορίας της απόχρωσης. Στην Εικόνα 11 φαίνεται ένα παράδειγμα της όλης διαδικασίας. Η χρωματική πληροφορία χρησιμοποιήθηκε μόνο στις περιοχές όπου ήταν αξιόπιστη,

⁶ Η τιμή αυτή είναι πειραματικά υπολογισμένη και ανταποκρίνεται στη βέλτιστη κατά μέσο όρο συμπεριφορά του αλγορίθμου.

όπως φαίνεται και στην εικόνα χρωματικότητας. Η διακύμανση της χρωματικής καθαρότητας είναι σε υψηλά επίπεδα σε σκοτεινές χρωματικές περιοχές. Σε μη χρωματικές περιοχές η χρωματική καθαρότητα είναι χαμηλή όπως και η διακύμανσή της. Φυσικά, η διακύμανση της απόχρωσης παραμένει υψηλή σε αχρωματικές περιοχές, όπου εξ' ορισμού η απόχρωση εκεί είναι αναξιόπιστη.

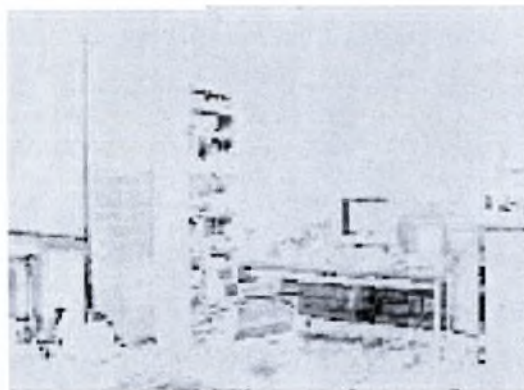
Αποτελέσματα των πειραμάτων φαίνονται επίσης και στην Εικόνα 12 αλλά και στην Εικόνα 13 για την περίπτωση gray-level με φαινόμενα θορύβου. Το σύστημα δίνει πολύ ικανοποιητικά αποτελέσματα ωστόσο, θα πρέπει να δοθεί έμφαση στην βελτιστοποίηση της απόδοσής του (κάποιες διαδικασίες ήταν δραματικά αργές ακόμα και σε αρκετά ισχυρά επεξεργαστικά συστήματα [11]).



(α) μέσες τιμές



(β) χρωματικές (άσπρο) και μη χρωματικές (μαύρο) κατανομές pixel

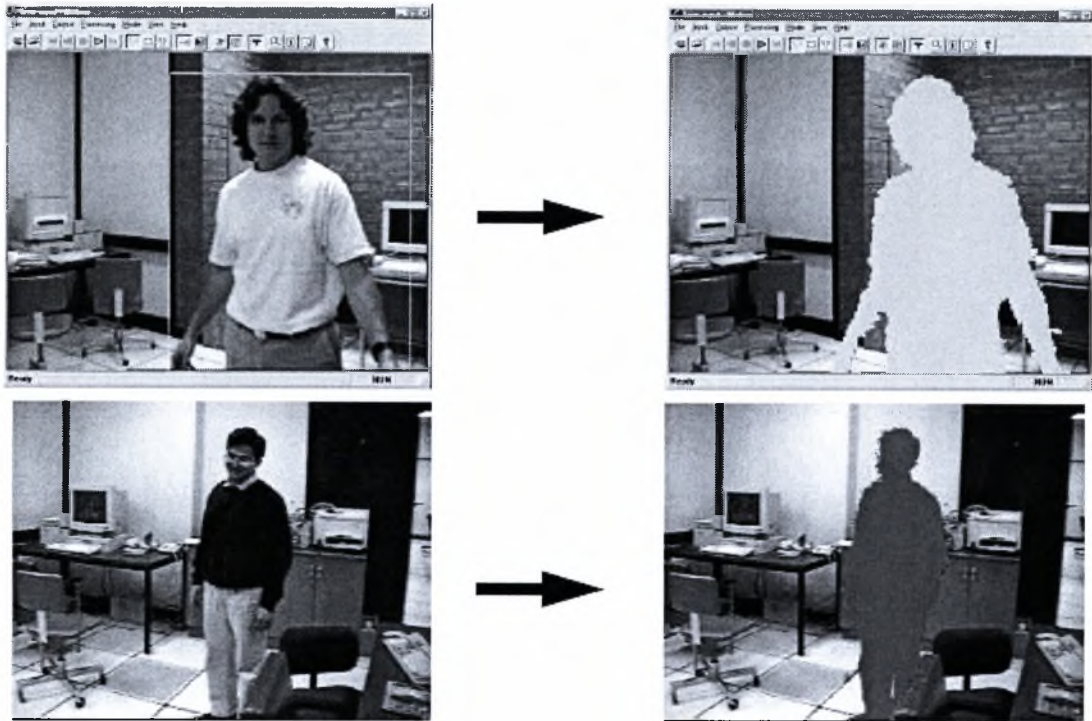


(c) διακύμανση της χρωματικής καθαρότητας (όσο πιο σκοτεινό τόσο πιο υψηλή)

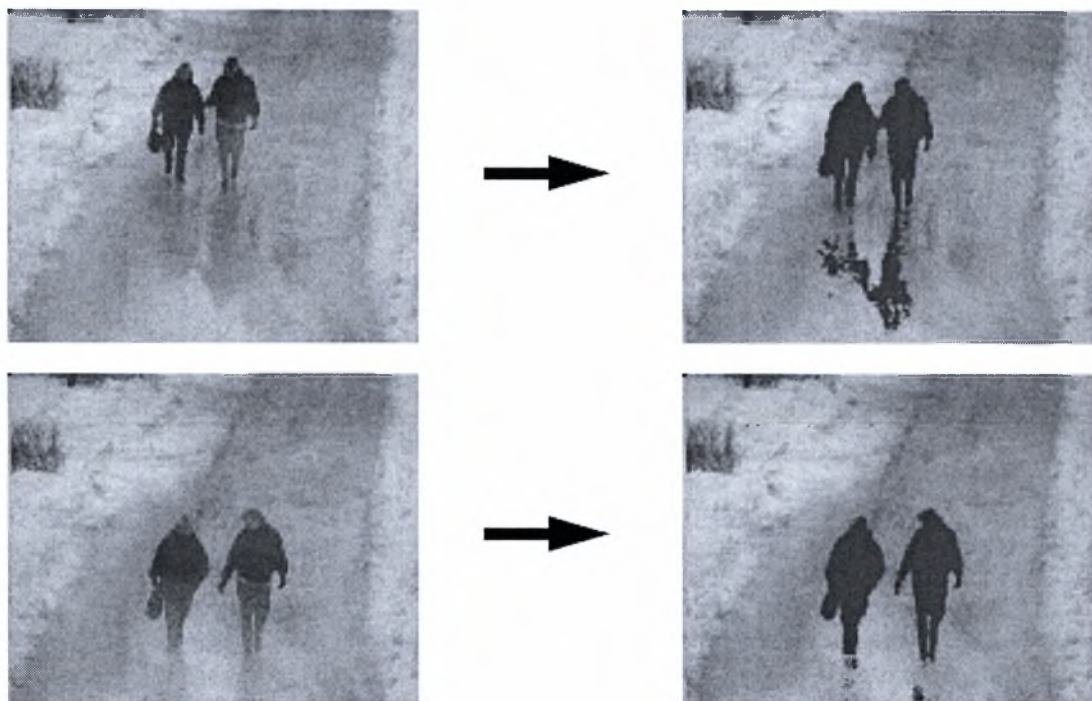


(d) διακύμανση της απόχρωσης (όσο πιο σκοτεινό τόσο πιο υψηλή)

Εικόνα 11: Κατανομές background έπειτα από μάθηση [11].



Εικόνα 12: Εικόνες λειτουργίας του συστήματος και διαχωρισμός περιοχών προσκηνίου [11].



Εικόνα 13: Αποτελέσματα τμηματοποίησης περιοχών σε gray-level [11].

3.2 Μοντελοποίηση Προσαρμόσιμου Έγχρωμου Παρασκηνίου για Ροές Δεδομένων Video Πραγματικού Χρόνου – Μέθοδος NHD

3.2.1 Γενικά

Συχνό πρόβλημα των μεθόδων του adaptive background modelling είναι ότι δεν υπάρχει καμία εκ των προτέρων γνώση για αυτό. Έτσι, το κάθε σύστημα θα πρέπει να το «μάθει» και να το μοντελοποιήσει. Όπως έχει ήδη προαναφερθεί, μια λύση θα ήταν να υπολογιστεί μια «μέση» εικόνα στον χρόνο, και αν είναι δυνατόν, σε περιόδους όπου δεν υπάρχουν κινούμενα αντικείμενα στο προσκήνιο. Έτσι η διαφορά κάθε νέου frame από το υπολογιζόμενο «μέσο» frame θα είναι εύκολο να ανιχνευθεί με την εφαρμογή ενός καταφλίου στη διαφορά αυτή. Ωστόσο, πρέπει να ληφθούν σοβαρά υπ' όψιν οι φυσικές αλλαγές του παρασκηνίου, για να μην γίνουν ικανές να μειώσουν την απόδοση του αλγορίθμου. Αυτό σημαίνει ότι απαιτείται ένα προσαρμόσιμο μοντέλο του παρασκηνίου.

Σε αυτό το πεδίο θα παρουσιαστεί ένας αλγόριθμος ο οποίος χρησιμοποιεί συνδυασμό gaussian μεταβλητών και έχει την δυνατότητα να χειρίζεται πραγματικού χρόνου δεδομένα video.

3.2.2 Ο Αλγόριθμος

Η λογική του αλγορίθμου NHD⁷ είναι ότι όσο πιο συχνά ένα pixel παίρνει μια συγκεκριμένη τιμή (έχει ένα συγκεκριμένο χρώμα), τόσο πιο πιθανό είναι αυτό να αποτελεί μέρος του παρασκηνίου. Γι' αυτόν τον λόγο έχει αναπτυχθεί μια τεχνική, η οποία κρατάει πληροφορία για τις τιμές των pixels στο παρελθόν (σε προηγούμενα frames). Κάθε pixel μοντελοποιείται από ένα group of K clusters όπου κάθε cluster αποτελείται από ένα βάρος w_k και μία μέση τιμή pixel ή centroid c_k . Επίσης, η περιοχή του background θεωρείται στατική.

⁷ Οι ερευνητές δεν έχουν κάνει γνωστή την προέλευση της ονομασίας του αλγορίθμου.

Τα pixels των νέων frames που ακολουθούν συγκρίνονται με το αντίστοιχο group of clusters και αναζητείται το cluster που ταιριάζει με το μεγαλύτερο βάρος w_k . Τα clusters είναι ταξινομημένα σε φθίνουσα σειρά βάρους για ταχύτερη αναζήτηση. Το cluster που ταιριάζει πρέπει να έχει απόσταση Manhattan⁸ μεταξύ της μέσης τιμής του και της τιμής του συγκεκριμένου pixel, κάτω από ένα συγκεκριμένο κατώφλι T . Αν δεν βρέθηκε κανένα ταιρίασμα, τότε το cluster με την μικρότερη τιμή βάρους αντικαθίσταται με ένα νέο cluster που έχει σαν μέση τιμή, την τιμή του pixel αυτού και ένα μικρό αρχικό βάρος. Αν βρέθηκε κάποιο ταιρίασμα, τότε τα βάρη όλων των clusters του συγκεκριμένου group ανανεώνουν την τιμή τους σύμφωνα με τον παρακάτω τύπο:

$$w_k = w_k + \frac{1}{L} (M_k - w_k)$$

όπου το M_k παίρνει την τιμή 1 για το cluster που ταιριάζει και 0 για τα υπόλοιπα. Η παράμετρος L είναι το αντίστροφο του κλασικού ρυθμού «εκμάθησης», α . Είναι αυτό που ελέγχει το πόσο γρήγορα οι αλλαγές του σκηνικού ενσωματώνονται στο μοντέλο του παρασκηνίου. Μικρές τιμές του L αντιστοιχούν σε ταχύτερη προσαρμοστικότητα και αντίστροφα.

Η μέση τιμή του cluster με το οποίο έγινε το ταιρίασμα πρέπει επίσης να ανανεωθεί σύμφωνα με το αντίστοιχο pixel. Αυτό που γίνεται είναι η συγκέντρωση του σφάλματος (διαφοράς) του centroid και του συγκεκριμένου νέου pixel. Όταν αυτό το σφάλμα υπερβαίνει το $L-1$ η τιμή του centroid αυξάνεται και όταν το σφάλμα πέφτει κάτω από το $-L$ μειώνεται. Αυτό σχεδόν ισοδυναμεί με την χρήση του τύπου:

$$c_k = c_k + \frac{1}{L} (x_i - c_k)$$

Μετά την προσαρμογή, τα βάρη όλων των clusters στο συγκεκριμένο group κανονικοποιούνται ώστε:

$$w_k = \frac{w_k}{S}, \text{ για κάθε } k, \text{ όπου } S = \sum_k w_k$$

⁸ Απόσταση Manhattan δύο σημείων είναι το άθροισμα των απολύτων των διαφορών των συντεταγμένων των δύο σημείων. π.χ.: $\text{ManhattanDistance}[P1(x1,y1),P2(x2,y2)] = |x1-x2|+|y1-y2|$.

Τα βάρη απαραίτητως αθροίζουν στην μονάδα και συμπεριφέρονται όπως οι πιθανότητες επειδή αναπαριστούν το ποσοστό του παρασκηνίου που ερμηνεύεται για κάθε cluster.

Τα κανονικοποιημένα clusters έπειτα από αυτό ταξινομούνται σε φθίνουσα σειρά. Προσοχή ωστόσο, πρέπει να δοθεί στην νέα τοποθεσία του cluster με το οποίο έγινε το ταίριασμα. Διατηρώντας τα βάρη ταξινομημένα βοηθάει τόσο την αρχική σύγκριση με τα clusters, αλλά και το τελικό βήμα της κατηγοριοποίησης. Τα pixels κατηγοριοποιούνται αθροίζοντας τα βάρη όλων των clusters, των οποίων το βάρος είναι μεγαλύτερο του cluster με το οποίο έγινε το ταίριασμα. Αυτό σημαίνει:

$$P = \sum_{k > M_k}^K w_k$$

Το αποτέλεσμα P , είναι το συνολικό ποσοστό του παρασκηνίου που ερμηνεύεται από τα υψηλότερου βάρους clusters και είναι μια εκτίμηση της πιθανότητας των επερχόμενων pixels να ανήκουν στο προσκήνιο. Μεγάλες τιμές του P , δείχνουν ότι το pixel ανήκει στο προσκήνιο και χαμηλές τιμές δείχνουν ότι ανήκει στο παρασκήνιο. Η τιμή αυτή μπορεί να περαστεί από ένα κατώφλι, ώστε να έχουμε μια δυαδική απόφαση ή μπορεί να κλιμακωθεί και να δώσει σαν αποτέλεσμα ένα gray-scale χάρτη απεικόνισης.

3.2.3 Υλοποίηση

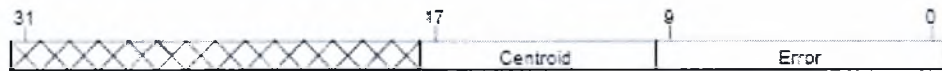
Από πειραματικές αναλύσεις παρατηρήθηκε πως ο χωρισμός των centroids οδηγεί σε βελτιωμένα και πιο ακριβή αποτελέσματα γιατί επιτρέπει τον προσδιορισμό διαφορετικών κατώφλιων. Έτσι τα centroids έγιναν δύο. Το luma centroid (luminance) και το chroma centroid (chromatic). Το luma centroid αποτελείται από δύο γειτονικές luma συνιστώσες (Y_1, Y_2) και το chroma centroid από τις αντίστοιχες chroma συνιστώσες (C_b, C_r) . Έχοντας μόνο ένα κατώφλι στην διαδικασία ήταν φανερό ότι το luma (η φωτεινότητα) θα κυριαρχούσε στις συγκρίσεις και στα ταιριάσματα των pixels, ενώ με αυτόν τον τρόπο, το ταίριασμα επιτελείται μόνο όταν η απόσταση Manhattan και των δύο συνιστωσών είναι κάτω από τα αντίστοιχα κατώφλια.

Όπως προαναφέρθηκε, η τιμή του σφάλματος είναι στο διάστημα $[-L, L-1]$. Ισοδύναμα μπορούμε να μετακινήσουμε αυτό το διάστημα κατά L ώστε να έχουμε το διάστημα $[0, 2L-1]$. Αν επίσης θέσουμε ως περιορισμό το L να είναι της μορφής $L=2^B$, τότε η όλη έκταση του διαστήματος μπορεί να καθοριστεί από $B+1$ bits. Με λίγα λόγια, η προσαρμογή των centroids πλέον, γίνεται μόνο σε περίπτωση που το συνολικό άθροισμα του σφάλματος προκαλεί υπερχειλίση ή υποχειλίση των $B+1$ bits.

Για το συγκεκριμένο τύπο video που χρησιμοποιήθηκε στα πειράματα χρειάστηκαν ακριβώς 8 bits για να περιγραφούν όλες οι συνιστώσες. Με αυτόν τον τρόπο αν χρησιμοποιηθούν ακέραιες τιμές integer 32 bits για τις τιμές των centroids (Y_1, Y_2, C_b, C_r) , τα υπολειπόμενα 24 bits μπορούν να χρησιμοποιηθούν για το συγκεντρωτικό αθροιστικό όρο σφάλματος. Αυτό δίνει στο B ένας εύρος $[0,23]$ ή ισοδύναμα $L \in \{1,2,4,\dots,8388608\}$. Ο χρήστης προσδιορίζει τον αριθμό των bits του B (που αναπαριστούν το L) και γίνεται ένα shift προς τα πάνω στις συνιστώσες των centroids κατά $B+1$ bits ανάλογα. Τα χαμηλότερα bits χρησιμοποιούνται για το άθροισμα του όρου σφάλματος το οποίο αρχικοποιείται στο L (Εικόνα 14). Έτσι, όταν έχουμε υπερχειλίση ή υποχειλίση των bits του όρου σφάλματος, το centroid αυτομάτως προσαρμόζεται συστηματικά.

Η προσεκτική παρατήρηση της εξίσωσης ανανέωσης των βαρών του cluster αποκαλύπτει άλλη μια μορφή βελτιστοποίησης του βασικού αλγορίθμου. Είναι απλό να δειχθεί ότι αν τα βάρη των clusters του group αθροίζουν στην μονάδα, μετά από εφαρμογή της εξίσωσης ανανέωσης τους, πάλι θα αθροίζουν στην μονάδα. Επομένως είναι αρκετό να κανονικοποιούμε τα βάρη μόνο όταν δεν μπορεί να βρεθεί κάποιο cluster να ταιριάζει στο συγκεκριμένο pixel, αναγκάζοντας τον αλγόριθμο να δημιουργήσει ένα νέο.

Άλλη μια ιδιότητα της εξίσωσης ανανέωσης βαρών είναι ότι τα βάρη των μη ταιριασμένων clusters κλιμακώνονται προς τα κάτω από τον ίδιο παράγοντα $\left(\pi. \chi. w_k = \frac{L-1}{L} w_k \right)$. Άρα, μόνο τα clusters που δημιουργήθηκαν και τα ταιριασμένα clusters έχουν την δυνατότητα να μην είναι ταξινομημένα. Επιπροσθέτως, όσο το βάρος ενός ταιριασμένου cluster αυξάνεται, απαιτείται μόνο να γίνεται ταξινόμηση προς την κατεύθυνση των υψηλόβαρων clusters.



Εικόνα 14: Ανάθεση bit centroid : $[B=9 \Leftrightarrow L=512]$ [4].

3.2.4 Τελική επεξεργασία

Υπάρχουν ωστόσο, δύο είδη λανθασμένης κατηγοριοποίησης που μπορεί να λάβουν χώρα στα αποτελέσματα τμηματοποίησης. False positive, όταν περιοχές του background είναι λανθασμένα κατηγοριοποιημένες ως περιοχές του προσκηνίου και αντίστροφα false negative, όταν περιοχές του προσκηνίου έχουν λανθασμένα κατηγοριοποιηθεί ως περιοχές του παρασκηνίου. Η τελική επεξεργασία (μετα-επεξεργασία) έχει ως σκοπό την μείωση αυτών των λανθασμένων κατηγοριοποιήσεων χωρίς βέβαια, όσο αυτό είναι δυνατόν, να μειώσει την ταχύτητα της κατηγοριοποίησης.

Τα λάθη false positive μοιάζουν με τον θόρυβο του «αλατοπίπερου» (salt & pepper noise) σε μια εικόνα και συνήθως οφείλονται σε θόρυβο της βιντεοκάμερας. Αυτός ο θόρυβος είναι μικρού μεγέθους περιοχές (1-2 pixels), λανθασμένα κατηγοριοποιημένες, περιτριγυρισμένες από σωστά κατηγοριοποιημένες περιοχές στο background. Ο συνηθισμένος μορφολογικός τελεστής του «ανοίγματος» (open⁹) μπορεί να χρησιμοποιηθεί για να μειώσει αυτά τα φαινόμενα διατηρώντας ωστόσο τα περιγράμματα των σωστά κατηγοριοποιημένων περιοχών.

Τα λάθη false negative προκύπτουν από την ύπαρξη ομοιοτήτων μεταξύ των χρωμάτων των αντικειμένων του παρασκηνίου και του παρασκηνίου. Σχηματίζουν τρύπες σε σωστά κατηγοριοποιημένες περιοχές του προσκηνίου και πιθανόν να εμφανίζονται και σε μεγάλο μέγεθος. Αυτό οδηγεί στο συμπέρασμα ότι τα σφάλματα false positive είναι πιο δύσκολο να διορθωθούν. Τυπικά, χρησιμοποιείται ένας αλγόριθμος συνεκτικών συνιστωσών για την ανίχνευση συνδεδεμένων κατηγοριοποιήσεων στο προσκήνιο. Έπειτα, οι μικρές περιοχές που θεωρούνται false positive εξαλείφονται και οι τρύπες στις εναπομένουσες περιοχές «γεμίζονται» κατάλληλα.

⁹ Οι μορφολογικοί τελεστές αναφέρονται αναλυτικότερα στην παράγραφο 4.4.2.

Πρώτο βήμα στην μείωση του ρυθμού της λανθασμένης κατηγοριοποίησης είναι η ένωση pixels που χωρίζονται μόνο από ένα κενό pixel. Ύστερα, εξάγονται τα περιγράμματα όλων των περιοχών που έχουν κατηγοριοποιηθεί σαν προσκήνιο¹⁰. Έπειτα, αν η περιοχή που περικλείεται από τα εξαγόμενα περιγράμματα είναι κάτω από ένα συγκεκριμένο κατώφλι, εξαλείφεται. Όλα τα περιγράμματα που απομένουν θεωρούνται πραγματικές περιοχές του προσκηνίου και συγκρατούνται. Η όλη τεχνική αποδείχθηκε ικανή στην αποφυγή των περισσότερων σφαλμάτων false positive και false negative και τα αποτελέσματα φαίνονται στην Εικόνα 15 που ακολουθεί.

3.2.5 Πειράματα – Συγκρίσεις - Αποτελέσματα

Ο αλγόριθμος που περιγράφηκε παραπάνω συγκρίθηκε ποιοτικά με άλλους τρεις αλγόριθμους-μεθόδους τμηματοποίησης του background. Ο πρώτος (VAR), υπολογίζει ένα μέσο και ένα frame διακύμανσης από ένα πλήθος αποθηκευμένων, προηγούμενων frames. Τα frames τμηματοποιούνται με βάση την κανονικοποιημένη διαφορά διακύμανσης μεταξύ αυτών και του μέσου frame. Οι αλγόριθμοι GMM1 και GMM2 είναι αλγόριθμοι που χρησιμοποιούν γκαουσιανά συνδυαστικά μοντέλα (Gaussian Mixture Models) με κάποιες διαφορές μεταξύ τους. Οι δοκιμές και οι συγκρίσεις έχουν γίνει χωρίς κανέναν παραλληλισμό του κώδικα και χρησιμοποιήθηκε μόνο ένας επεξεργαστής. Ο Πίνακας 1 συνοψίζει τις πραγματικού χρόνου αποδόσεις των τεσσάρων αλγορίθμων χωρίς το κομμάτι της τελικής επεξεργασίας και η Εικόνα 15 συγκρίνει τα τελικά αποτελέσματα τμηματοποίησης μετά και της τελικής επεξεργασίας.

Σημείο κλειδί στην χώρο της τμηματοποίησης του παρασκηνίου και στους αλγορίθμους που το επιτυγχάνουν είναι η υπολογιστική τους πολυπλοκότητα. Ο αλγόριθμος NHD όπως φαίνεται και στον Πίνακα 1 είναι 4 φορές πιο απαιτητικός υπολογιστικά από τον απλό αλγόριθμο τμηματοποίησης με βάση την διακύμανση (VAR). Επίσης, πετυχαίνει σχεδόν ίσο ρυθμό επεξεργασίας frames (28 fps), αλλά όπως φαίνεται στην Εικόνα 15 τα ποιοτικά αποτελέσματα είναι πολύ πιο ικανοποιητικά από τον VAR. Επιπλέον, ο NHD προσαρμόζει το background στο πέρασμα κάθε νέου frame ενώ ο VAR το πραγματοποιεί μία φορά κάθε 300 frames.

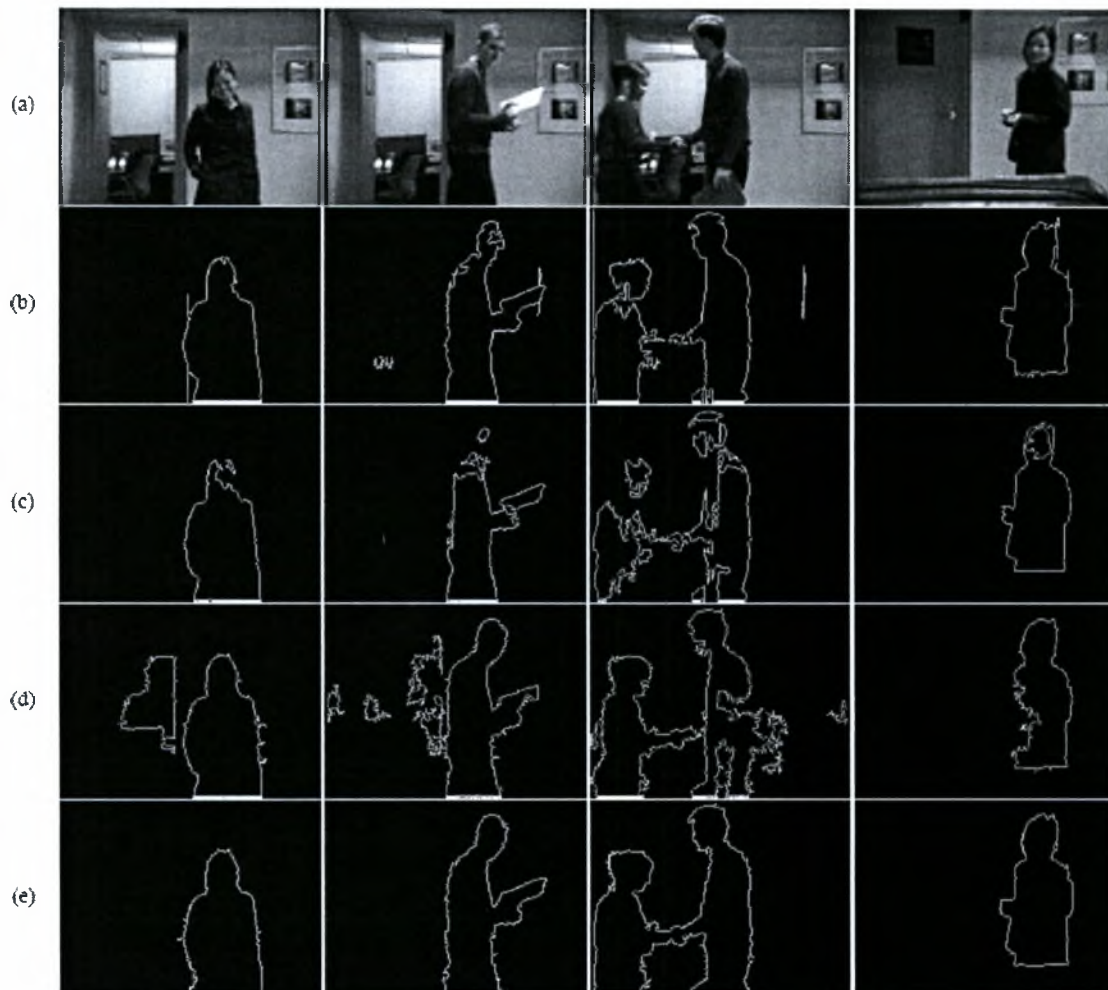
¹⁰ Για τη συγκεκριμένη διαδικασία χρησιμοποιήθηκαν συναρτήσεις της βιβλιοθήκης OpenCV – Open Source Computer Vision Library (<http://www.intel.com/technology/computing/opencv/>)

Αυξάνοντας βέβαια τον ρυθμό προσαρμογής του παρασκηνίου οδηγεί σε αύξηση της χρήσης του επεξεργαστή.

Το στάδιο της τελικής επεξεργασίας και των τεσσάρων αλγορίθμων ήταν πανομοιότυπο. Αν και η τεχνική λειτούργησε καλά σε γενικές γραμμές, σίγουρα δεν είναι η βέλτιστη για κάθε αλγόριθμο και γι' αυτόν τον λόγο τα αποτελέσματα δεν πρέπει να ερμηνευτούν ως τα καλύτερα δυνατά που μπορεί να δώσει κάθε ένας από αυτούς. Παρόλα αυτά, είναι ξεκάθαρο, ότι ο NHD είναι το ίδιο ή και ακόμα καλύτερος από άλλες δοκιμασμένες τεχνικές, ενώ επιτυγχάνει υψηλούς ρυθμούς επεξεργασίας frame.

Algorithm	Video Format	User Params.	Frame Rate (fps)	Mem. Usage (Mb)	CPU Usage (%)
VAR	Y'CbCr	8	29	11.9	25
GMM1 [3]	RGB	2	9	23.9	97
GMM2 [4]	RGB	2	13, 6 ²	25.4	97
NHD	Y'CbCr	4	28	13.0	92

Πίνακας 1: Γνωρίσματα αλγορίθμων και απόδοση [4].



Εικόνα 15: Αποτελέσματα Τμηματοποίησης (a) frame εισόδου (b) VAR (c) GMM1 (d) GMM2 (e) NHD [4].

3.3 Μοντελοποίηση Προσαρμόσιμου Έγχρωμου Παρασκηνίου για Ροές Δεδομένων Video Πραγματικού Χρόνου με Χρήση GMM

3.3.1 Γενικά

Η λογική που έχει ακολουθηθεί σε αυτήν την μέθοδο είναι αυτή που έχει συνοπτικά περιγραφεί σε προηγούμενα πεδία. Επιλέγεται η μοντελοποίηση κάθε τιμής pixel από συνδυασμό πολλών gaussian κατανομών αντί της μοντελοποίησης όλων των τιμών των pixels με μία συγκεκριμένου τύπου κατανομή. Βασιζόμενοι στη μέση τιμή και στην διακύμανση καθεμίας από τις κατανομές του συνδυασμού αποφασίζεται ποιες από τις gaussians αντιπροσωπεύουν χρώματα του παρασκηνίου. Οι υπόλοιπες τιμές θεωρούνται χρώματα αντικείμενων του προσκηνίου μέχρι να βρεθεί μια κατανομή, η οποία τα εμπεριέχει, με ικανά και αρκετά στοιχεία να το υποστηρίξουν.

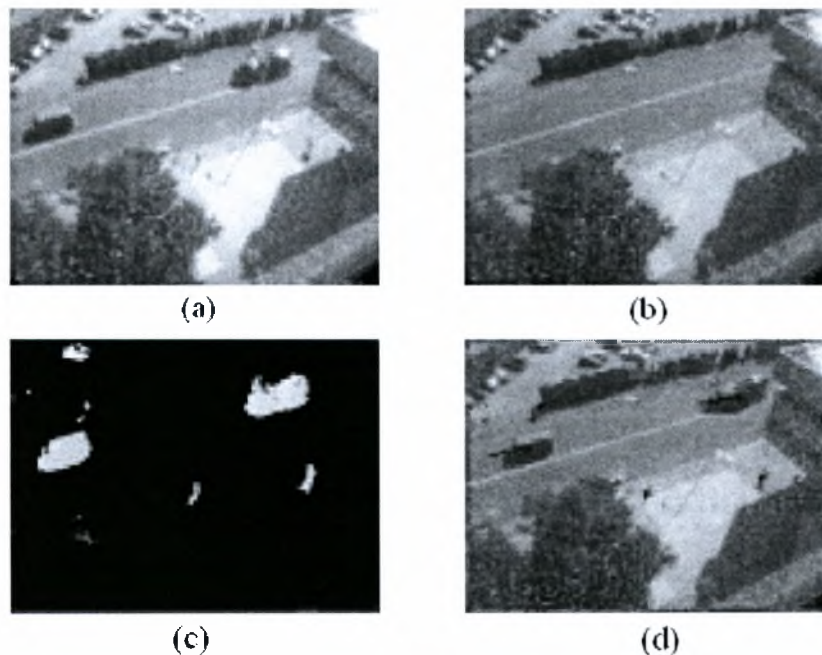
Το σύστημα που έχει υλοποιηθεί με την μέθοδο αυτή χειρίζεται ικανοποιητικά αλλαγές φωτισμού, επαναληπτικές κινήσεις αντικείμενων του σκηνικού, αργά κινούμενα αντικείμενα και εισερχόμενα ή εξερχόμενα αντικείμενα του σκηνικού. Τα αργά κινούμενα αντικείμενα απαιτούν περισσότερο χρόνο να ανιχνευτούν και να ενσωματωθούν στο παρασκήνιο, διότι τα χρώματά τους έχουν μεγαλύτερη διακύμανση από αυτά του background. Επίσης, οι επαναληπτικές αποκλίσεις προσαρμόζονται και το μοντέλο της κατανομής του παρασκηνίου διατηρείται ακόμα και αν παροδικά αντικατασταθεί από άλλη κατανομή, κάτι το οποίο οδηγεί σε ταχύτερη ανάκτηση όταν αντικείμενα του προσκηνίου απομακρύνονται από αυτό.

Για την υλοποίηση της μεθόδου χρειάστηκαν δύο παράμετροι: α , ο ρυθμός «εκμάθησης» και T , το ποσοστό των δεδομένων τα οποία θα μπορούν αν ερμηνεύσουν το παρασκήνιο.

3.3.2 Η Μέθοδος

Μία απλή gaussian κατανομή θα ήταν αρκετή να μοντελοποιήσει την τιμή κάθε pixel, αν βέβαια προερχόταν από μία συγκεκριμένη επιφάνεια κάτω από έναν συγκεκριμένο φωτισμό. Παραδείγματος χάριν, αν άλλαζε μόνο ο φωτισμός κάθε φορά στο σκηνικό, τότε μία και μόνο κατανομή Gauss ανά pixel θα ήταν αρκετή να μοντελοποιήσει κάθε frame. Στην πραγματικότητα όμως, οι επιφάνειες παίρνουν συχνά διαφορετικά σχήματα και μορφή αλλά και οι συνθήκες φωτισμού μεταβάλλονται διαρκώς στο πέρασμα του χρόνου. Γι' αυτούς τους λόγους, η χρήση πολλαπλών προσαρμοστικών gaussian κατανομών κρίνεται απαραίτητη και λαμβάνει χώρα και σε αυτήν την μέθοδο.

Κάθε φορά που οι παράμετροι των gaussians ανανεώνονται, κάθε μία κατανομή αποτιμάται σύμφωνα με κάποιο ευριστικό «μηχανισμό» για το εάν αποτελεί μέρος της «διαδικασίας του παρασκηνίου». Οι τιμές των pixels οι οποίες δεν αντιστοιχούν σε καμία gaussian κατανομή ομαδοποιούνται σε συνεντικές συνιστώσες. Αυτές έπειτα παρακολουθούνται από frame σε frame χρησιμοποιώντας έναν πολλαπλής υπόθεσης ανιχνευτή (multiple hypothesis tracker). Η όλη διαδικασία φαίνεται στην Εικόνα 16.



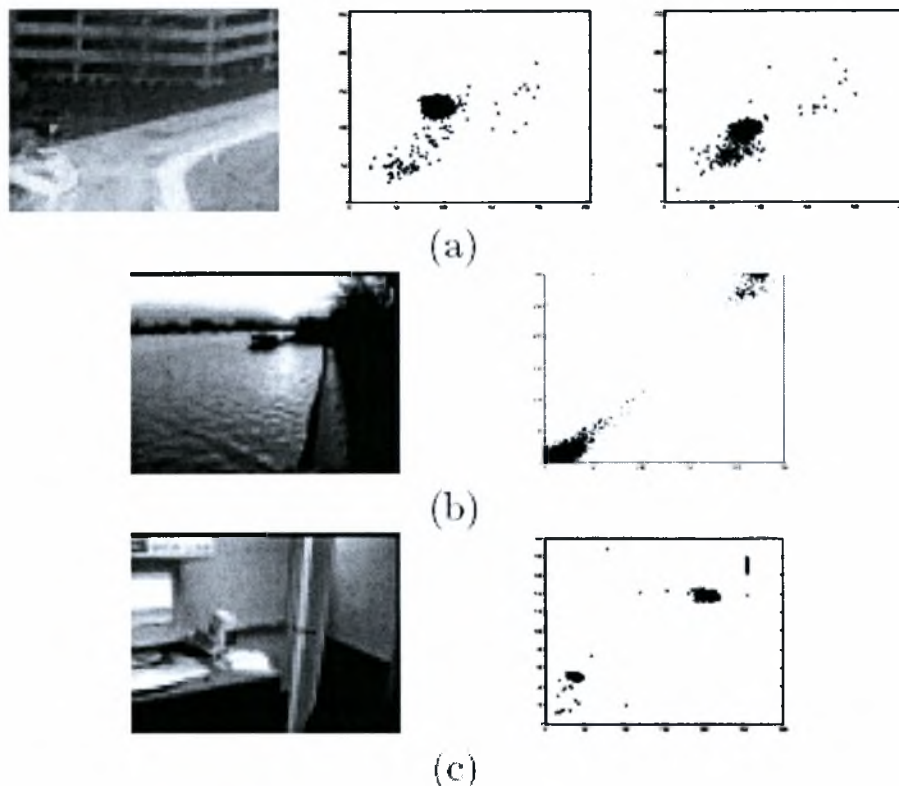
Εικόνα 16: Η εκτέλεση του προγράμματος. (a) αρχική εικόνα (b) εικόνα παραγόμενη από τους μέσους των πιο πιθανών Gaussians του μοντέλου του παρασκηνίου (c) τα pixels του προσκηνίου (d) η εικόνα (b) με έμφαση στα αντικείμενα που έχουν ανιχνευθεί [1].

3.3.3 Το Συνδυαστικό Μοντέλο

Ας θεωρήσουμε σαν «διαδικασία pixels» τις τιμές ενός pixel στην πάροδο του χρόνου. Δηλαδή για κάποια χρονική στιγμή t αυτό που είναι γνωστό για ένα pixel $\{x_0, y_0\}$ είναι η ιστορία της φωτεινότητάς του:

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

όπου I η ακολουθία των εικόνων. Μερικές «διαδικασίες pixels» παρουσιάζονται στην Εικόνα 16 σε μορφή (R,G) scatter plots από τα οποία φαίνεται εμφανώς η ανάγκη για προσαρμόσιμα backgrounds. Οι Εικόνες 17 (b) και (c) επίσης τονίζουν την ανάγκη για πολυτροπικές αναπαραστάσεις.



Εικόνα 17: Εικόνες και scatter plots των χρωμάτων κόκκινο και πράσινο ενός pixel μιας εικόνας στον χρόνο. (a) 2 scatter plots του ίδιου pixels με διαφορά 2 λεπτών (b) ένα διπλής κατανομής μοντέλο των τιμών των pixels συνακόλουθο της ανακλαστικότητας της επιφάνειας του νερού (c) άλλο ένα μοντέλο από το flicking της οθόνης του H/Y [1].

Στην περίπτωση, που το παρασκήνιο και ο φωτισμός ήταν στάσιμα, η τιμή κάθε pixel θα ήταν σχετικά σταθερή. Αν υποθέσουμε, ότι παρεμβαλλόταν ανεξάρτητος,

gaussian θόρυβος στην διαδικασία δειγματοληψίας, η πυκνότητα του μπορούσε να περιγραφεί από μία Gaussian κατανομή με κέντρο την μέση τιμή των τιμών των pixels. Δυστυχώς για εμάς, οι ακολουθίες video που μας ενδιαφέρουν εμπεριέχουν αλλαγές στον φωτισμό του σκηνικού, αλλαγές στο ίδιο το σκηνικό αλλά και κινούμενα αντικείμενα.

Οι αλλαγές του φωτισμού είναι απαραίτητο να παρακολουθούνται από την gaussian κατανομή. Αν ένα στάσιμο αντικείμενο εισερχόταν στο σκηνικό και δεν ενσωματωνόταν στο παρασκήνιο, τα αντίστοιχα pixels θα μπορούσαν να θεωρούνται pixels του προσκηνίου για αυθαίρετα μεγάλο χρονικό διάστημα. Αυτό βέβαια, θα οδηγούσε σε πολλαπλά λάθη εκτίμησης του προσκηνίου και σε χαμηλής ποιότητας ιχνηλάτηση από το σύστημα. Οι παραπάνω παράγοντες δείχνουν πως οι τελευταίες παρατηρήσεις μπορεί να είναι πιο σημαντικές στην εκτίμηση των παραμέτρων των gaussians.

Επιπρόσθετα, άλλο ένα σημαντικό ζήτημα είναι τα κινούμενα αντικείμενα. Ένα κινούμενο αντικείμενο αναμένεται να έχει μεγαλύτερες τιμές διακύμανσης από ένα στάσιμο. Γενικά, θα υπάρχουν περισσότερα δεδομένα, τα οποία θα υποστηρίζουν τις κατανομές του background επειδή αυτές επαναλαμβάνονται, όπου οι τιμές των pixels διαφόρων αντικειμένων δεν είναι του ίδιου χρώματος.

Η πρόσφατη ιστορία κάθε pixel $\{X_t, \dots, X_1\}$ μοντελοποιείται από έναν συνδυασμό από K gaussians κατανομές. Η πιθανότητα να παρατηρηθεί το τρέχον pixel είναι:

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

όπου K το πλήθος των κατανομών, $\omega_{i,t}$ είναι μια εκτίμηση του βάρους της i -οστής κατανομής στον συνδυασμό την στιγμή t , $\mu_{i,t}$ είναι η μέση τιμή της i -οστής κατανομής του συνδυασμού στη στιγμή t , $\Sigma_{i,t}$ είναι ο πίνακας covariance της i -οστής κατανομής του συνδυασμού στη στιγμή t , και η είναι η gaussian συνάρτηση πυκνότητας πιθανότητας:

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

Η παράμετρος K καθορίζεται από την διαθέσιμη μνήμη και υπολογιστική δύναμη του εκάστοτε συστήματος. Συνήθως χρησιμοποιούνται τιμές από το 3 έως το 5. Επίσης, για υπολογιστικούς λόγους ο πίνακας covariance υποθέτουμε ότι είναι της μορφής:

$$\Sigma_{k,t} = \sigma_k^2 I$$

Αυτό δείχνει ότι τα χρώματα κόκκινο, πράσινο και μπλε είναι ανεξάρτητα μεταξύ τους και έχουν ίδιες διακυμάνσεις. Επιπλέον, αυτή η υπόθεση μας επιτρέπει την εύκολη αντιστροφή του πίνακα, φυσικά με κάποιο μικρό κόστος στην ποιότητα.

Κατ' αυτόν τον τρόπο, η κατανομή των πιο πρόσφατα παρατηρούμενων τιμών του κάθε pixel στο σκηνικό χαρακτηρίζεται από έναν συνδυασμό από gaussians. Μια νέα τιμή pixel θα αναπαρασταθεί με μία από τις βασικές συνιστώσες του συνδυαστικού μοντέλου και θα χρησιμοποιηθεί για την ανανέωση του.

Μία ιδανική μέθοδος μεγιστοποίησης της πιθανοφάνειας των παρατηρούμενων δεδομένων είναι αυτή του expectation – maximization, αν η «διαδικασία των pixels» μπορεί να θεωρηθεί στατική. Δυστυχώς όμως, αυτή η διαδικασία διαφέρει σε κάθε pixel στο πέρασμα του χρόνου με τις διάφορες αλλαγές στο περιβάλλον. Γι' αυτό, χρησιμοποιείται μια προσεγγιστική μέθοδος, η οποία θεωρεί κάθε νέα παρατήρηση, ως ομάδα δειγμάτων μεγέθους 1 και χρησιμοποιεί ιδανικούς κανόνες εκμάθησης για να ενοποιήσει τα νέα δεδομένα.

Επειδή, το μοντέλο πολλαπλών κατανομών αντιστοιχεί σε ένα συνδυασμό κατανομών διαφορετικό για κάθε pixel, μια επακριβής υλοποίηση του αλγορίθμου του EM θα ήταν αρκετά δαπανηρή υπολογιστικά. Για τον λόγο αυτό προτιμάται η υλοποίηση μιας σειριακής προσέγγισης του αλγορίθμου των K -μέσων (K -means). Κάθε νέα τιμή pixel X_t , ελέγχεται με βάση τις υπάρχουσες K gaussian κατανομές μέχρι να βρεθεί η κατάλληλη που ταιριάζει. Το ταιρίασμα αυτό ορίζεται όταν η τιμή ενός pixel είναι εντός δύο σταθερών αποκλίσεων μιας κατανομής. Η τιμή του κατώφλιου μπορεί να παίζει και αυτή τον ρόλο της στην απόδοση του συστήματος. Αυτό το κατώφλι εφαρμόζεται αποτελεσματικά ανά pixel και ανά κατανομή, το οποίο είναι εξαιρετικά χρήσιμο όταν διαφορετικές περιοχές της εικόνας έχουν διαφορετικό φωτισμό [Εικόνα 17 (a)]. Αυτό γίνεται, επειδή κάποια αντικείμενα που εμφανίζονται σε σκιερές περιοχές δεν δημιουργούν τον ίδιο θόρυβο με άλλα που εμφανίζονται σε πιο φωτισμένες. Ένα ομοιόμορφο κατώφλι θα οδηγούσε στην εξαφάνιση αντικείμενων όταν αυτά θα εισέρχονταν σε σκιερές περιοχές.

Ας υποθέσουμε, ότι καμία από τις K κατανομές δεν ταιριάζει με ένα συγκεκριμένο pixel, τότε η λιγότερη πιθανή κατανομή αντικαθίσταται με μία κατανομή με μέση τιμή την τιμή αυτού του pixel, μια υψηλή αρχική διακύμανση και ένα χαμηλής προτεραιότητας βάρος.

Τα βάρη των K κατανομών την χρονική στιγμή t , ανανεώνονται με βάση τον παρακάτω τύπο:

$$\omega_{k,t} = (1 - \alpha) \omega_{k,t-1} + \alpha (M_{k,t})$$

όπου α ο ρυθμός «εκμάθησης» και $M_{k,t}$ παίρνει την τιμή 1 για την κατανομή που ταιριάζει και 0 για όλες τις υπόλοιπες. Μετά την ανανέωση τα βάρη επανακανονικοποιούνται. Η τιμή $1/\alpha$ ορίζει μια χρονική σταθερά που προσδιορίζει την ταχύτητα με την οποία αλλάζουν οι παράμετροι των κατανομών. Το $\omega_{k,t}$ είναι μία μέση τιμή της εκ των υστέρων πιθανότητας οι τιμές των pixels να ταιριάζουν στην κατανομή k , δοσμένων των παρατηρήσεων από την χρονική στιγμή 1 μέχρι την t , περασμένη από ένα χαμηλοπερατό φίλτρο.

Οι παράμετροι μ και σ για τις κατανομές που δεν αντιστοιχίστηκαν στην τιμή του pixels παραμένουν ως έχουν. Για την αντίστοιχη κατανομή όμως η ανανέωση γίνεται ως εξής:

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

$$\text{όπου } \rho = \alpha \eta (X_t | \mu_k, \sigma_k)$$

όπου παρατηρούμε ότι είναι ο ίδιος τύπος του χαμηλοπερατού φίλτρου, όπως προαναφέρθηκε, με τη διαφορά ότι μόνο τα δεδομένα που ταιριάζουν περιλαμβάνονται σε αυτήν την εκτίμηση.

Ένα από τα εντυπωσιακά πλεονεκτήματα αυτής της μεθόδου παρατηρείται όταν ένα αντικείμενο γίνεται μέλος του παρασκηνίου. Το ήδη υπάρχον μοντέλο του παρασκηνίου δεν καταστρέφεται, αλλά παραμένει στον συνδυασμό έως ότου γίνει το K -οστό πιθανότερο και ένα νέο χρώμα παρατηρηθεί. Έτσι, όταν ένα αντικείμενο είναι στάσιμο για τόσο χρονικό διάστημα όσο ακριβώς χρειάζεται για να γίνει μέρος του background και έπειτα κινείται, η κατανομή που περιέγραφε το προηγούμενο

background παραμένει με τις ίδιες τιμές μ και σ^2 αλλά μικρότερο ω και γρήγορα θα επανενσωματωθεί στο background.

3.3.4 Μοντέλο Εκτίμησης Παρασκηνίου με Χρήση GMM

Είναι επιθυμητό να προσδιοριστούν ποιες από τις gaussians του μοντέλου είναι πιθανότερο να παράγονται από το background, όσο οι παράμετροι του συνδυαστικού μοντέλου ανά pixel αλλάζουν. Μας ενδιαφέρουν οι gaussians κατανομές, οι οποίες αποτελούν απόδειξη για το παραπάνω και έχουν την μικρότερη δυνατή διακύμανση.

Για να γίνει κατανοητό αυτό, ας υποθέσουμε την συγκέντρωση των υποστηρικτικών στοιχείων και τη σχετικά χαμηλή διακύμανση για τις κατανομές του background, όταν ένα στάσιμο αντικείμενο είναι συνεχώς ορατό. Αντίθετα, όταν ένα αντικείμενο του παρασκηνίου παρεμβάλλεται σε ένα αντικείμενο του παρασκηνίου, δεν θα ταιριάζει με κάποια από τις υπάρχουσες κατανομές του παρασκηνίου. Αυτό θα έχει σαν αποτέλεσμα την δημιουργία μιας νέας κατανομής ή την αύξηση της διακύμανσης μιας ήδη υπάρχουσας. Επιπρόσθετα, η διακύμανση του κινούμενου αντικειμένου περιμένουμε να είναι μεγαλύτερη από τα pixels του background μέχρι αυτό να σταματήσει να κινείται. Για να μοντελοποιηθεί αυτό, πρέπει να αποφασιστεί ποιο ποσοστό του συνδυαστικού μοντέλου αναπαριστά καλύτερα τις «διαδικασίες» του background.

Αρχικά, οι gaussians ταξινομούνται με βάση την τιμή ω/σ . Αυτή η τιμή αυξάνει τόσο, όσο όταν η κατανομή «κερδίζει» νέα δεδομένα τα οποία ταιριάζουν σε αυτήν, αλλά όσο και όταν η διακύμανση μειώνεται. Διαισθητικά, όταν υπάρχει μια συνεκτικότητα στις τιμές της κατανομής και δεν «διακυμαίνονται» σε μεγάλο εύρος (μικρό σ), τότε η κατανομή θεωρείται πιο αξιόπιστη και αυξάνεται η προτίμηση των επόμενων pixels σε αυτήν (αύξηση του ω). Είναι λογικό, μετά την επανεκτίμηση των παραμέτρων του μοντέλου οι κατανομές να ταξινομηθούν ξεκινώντας από αυτήν την οποία έγινε η αντιστοίχιση και προς τις πιθανότερες κατανομές του background. Αυτό γίνεται επειδή μόνο οι κατανομές που έχουν αντιστοιχηθεί έχουν αλλάξει τις σχετικές τιμές τους. Κατ' αυτόν τον τρόπο, στην κορυφή της ταξινόμησης υπάρχουν οι πιο πιθανές κατανομές του background. Οι δε παροδικές (λιγότερο πιθανές

κατανομές) κατεβαίνουν στο τέλος της λίστας αυτής και εν τέλει αντικαθίστανται από άλλες, νέες κατανομές.

Μετέπειτα, οι πρώτες B κατανομές επιλέγονται ως το μοντέλο του παρασκηνίου όπου:

$$B = \operatorname{argmin}_b \left(\sum_{k=1}^b \omega_k > T \right)$$

όπου T είναι μια μέτρηση του ελάχιστου ποσοστού δεδομένων που μπορούν να ερμηνευτούν από το παρασκήνιο. Αυτό δηλαδή παίρνει τις «καλύτερες» κατανομές μέχρι ένα συγκεκριμένο ποσοστό (T) των πρόσφατων δεδομένων το οποίο έχει ερμηνευτεί.

Αν η τιμή του T είναι μικρή, τότε συνήθως το μοντέλο του παρασκηνίου είναι μονοτροπικό. Σε αυτήν την περίπτωση επιλέγεται μόνο μία, η πιο πιθανή κατανομή. Αν το T είναι μεγάλο τότε η πολυτροπική κατανομή που δημιουργείται από κίνηση αντικειμένων στο background (κίνηση φύλλων δένδρου, κίνηση σημαίας στον άνεμο, κλπ). Αυτό οδηγεί στην προσμέτρηση περισσότερων του ενός χρώματος στο μοντέλο του background και σαν επακόλουθο επιτρέπει στο παρασκήνιο την αποδοχή δύο ή και περισσότερων διαφορετικών χρωμάτων.

3.3.5 Συνεκτικές Συνιστώσες

Τα παραπάνω βήματα που περιγράφηκαν επιτρέπουν την ανίχνευση pixels του προσκηνίου σε κάθε νέο frame ενώ ανανεώνουν την περιγραφή του παρασκηνίου. Αυτά τα «σημαδεμένα» pixels του προσκηνίου μπορούν να τμηματοποιηθούν σε περιοχές από έναν αλγόριθμο συνεκτικών συνιστωσών δύο-περασμάτων.

Επειδή, αυτή η διαδικασία είναι αποτελεσματική για τον προσδιορισμό ολόκληρου του κινούμενου αντικειμένου, οι κινούμενες περιοχές μπορούν να χαρακτηριστούν όχι μόνο από την θέση τους, αλλά από το μέγεθος και από άλλες σχηματικές πληροφορίες. Αυτά τα χαρακτηριστικά δεν είναι μόνο χρήσιμα για περαιτέρω επεξεργασία και κατηγοριοποίηση αλλά μπορούν να βοηθήσουν σημαντικά στην διαδικασία ιχνηλάτησης.

3.3.6 Ιχνηλάτηση Πολλαπλής Υπόθεσης (*Multiple Hypothesis Tracking*)

Η επίτευξη της αντιστοιχίας των συνεκτικών συνιστωσών μεταξύ των frames επιτυγχάνεται από έναν αλγόριθμο πρόγνωσης, πολλαπλής υπόθεσης ιχνηλάτησης, ο οποίος λαμβάνει υπ' όψιν τα μεγέθη και τις θέσεις. Για αυτόν τον λόγο έχει υλοποιηθεί μια σειριακής επεξεργασίας μέθοδος που δημιουργεί και διατηρεί ομάδες από φίλτρα Kalman [1].

Για κάθε frame, υπάρχει ένας χώρος με αποθηκευμένα μοντέλα Kalman και ένας νέος διαθέσιμος χώρος από συνεκτικές συνιστώσες που μπορούν να ερμηνευτούν από αυτά. Αρχικά, τα μοντέλα είναι πιθανοτικά αντιστοιχισμένα με τις συνεκτικές περιοχές. Έπειτα, οι συνεκτικές περιοχές που δεν ερμηνεύονται αποτελεσματικά από τα μοντέλα ελέγχονται ώστε να βρεθούν νέα μοντέλα Kalman. Τελικά, τα μοντέλα των οποίων η ικανότητα αντιστοίχισης (όπως αυτή ορίζεται από το αντίστροφο της διακύμανσης του σφάλματος πρόγνωσης) πέφτει κάτω από ένα συγκεκριμένο κατώφλι, απομακρύνονται.

Αντιστοιχίζοντας τα μοντέλα με τις συνεκτικές συνιστώσες συμπεριλαμβάνει τον έλεγχο κάθε ήδη υπάρχοντος μοντέλου με τον διαθέσιμο χώρο συνεκτικών συνιστωσών ο οποίος αποτελείται από πολλά pixels και όχι ένα ή δύο. Όλα τα ταιριάσματα χρησιμοποιούνται για να ανανεωθεί το αντίστοιχο μοντέλο. Αν το ανανεωμένο μοντέλο έχει ικανοποιητικές ικανότητες αντιστοίχισης, τότε θα χρησιμοποιηθεί για το επόμενο frame. Σε περίπτωση που δεν βρεθεί κανένα ταιρίασμα τότε υποθέτεται ένα «κενό» ταιρίασμα για το μοντέλο και μειώνεται η ικανότητα αντιστοίχισης που έχει από έναν σταθερό παράγοντα.

Τα μη αντιστοιχισμένα μοντέλα από το τρέχον frame μαζί με τα προηγούμενα δύο frames έπειτα χρησιμοποιούνται για την υπόθεση νέων μοντέλων. Χρησιμοποιώντας ζευγάρια από μη αντιστοιχισμένες συνεκτικές συνιστώσες από τα δύο προηγούμενα frames, υποθέτεται ένα νέο μοντέλο. Αν το τρέχον frame περιλαμβάνει μια ικανή αντιστοίχιση το ανανεωμένο μοντέλο προστίθεται στα υπάρχοντα μοντέλα. Για να αποφευχθούν συνδυαστικές πολυπλοκότητες σε καταστάσεις θορύβου, μπορεί να είναι απαραίτητο να οριστεί ένα όριο στον μέγιστο αριθμό των μοντέλων εξαλείφοντας τα λιγότερα πιθανά μοντέλα. Σε καταστάσεις θορύβου, συχνά είναι χρήσιμο να απαλείφονται μικρά σημάδια που μπορεί να είναι αποτέλεσμα τυχαίων αντιστοιχίσεων.

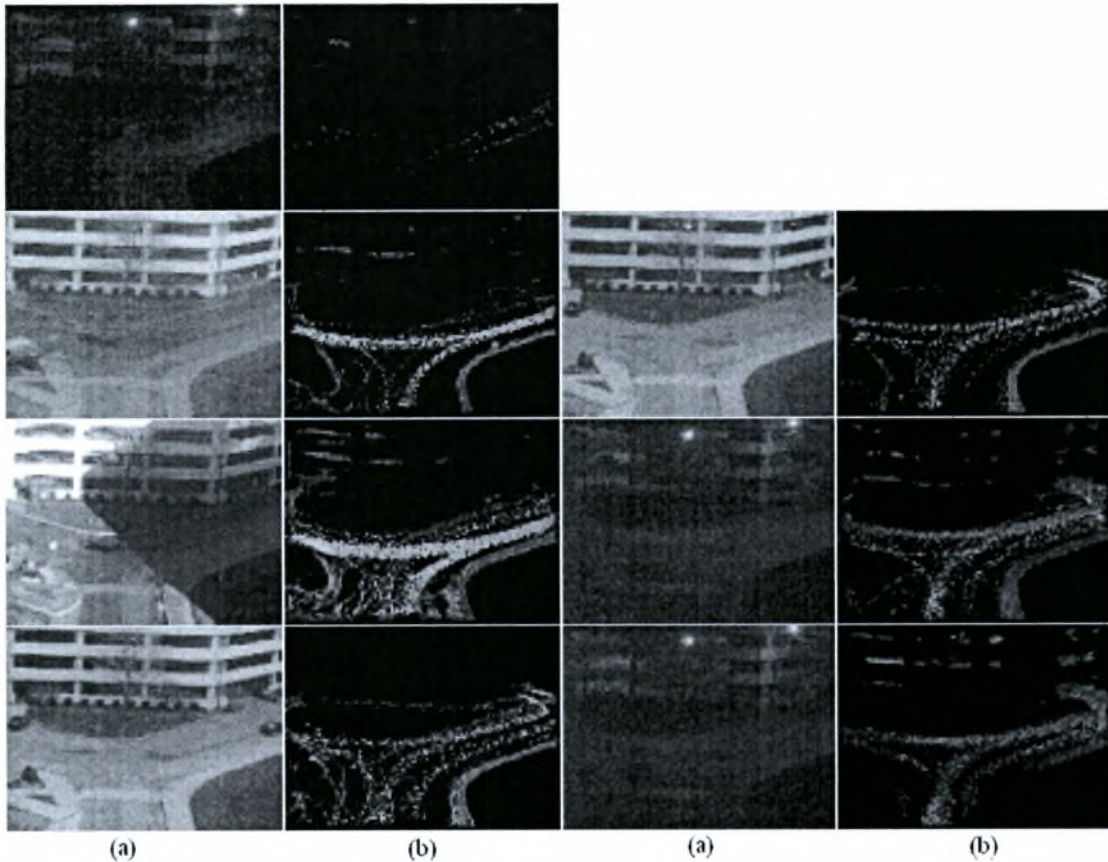
3.3.7 Πειράματα και Αποτελέσματα

Στο συγκεκριμένο σύστημα στο οποίο δοκιμάστηκε η παρούσα μέθοδος, η επεξεργασία είχε έναν ρυθμό από 11 έως 13 frames/second. Αυτή η απόκλιση οφείλεται στην διακύμανση της ποσότητας του παρασκηνίου που ήταν παρούσα κάθε φορά. Η Εικόνα 18 δείχνει τα αθροιστικά ίχνη ενός σκηνικού κατά την πάροδο μιας μέρας. Αν και οι γρήγορες καιρικές αλλαγές, όπως το συννέφιασμα του ουρανού (ανάλογο του ρυθμού «εκμάθησης» α), κάποιες φορές καθιστούσε απαραίτητη την αλλαγή των κατανομών του παρασκηνίου, αυτό σταθεροποιούνταν πάλι σε 10-20 δευτερόλεπτα και η ιχνηλάτηση συνεχιζόταν απρόσκοπτα.

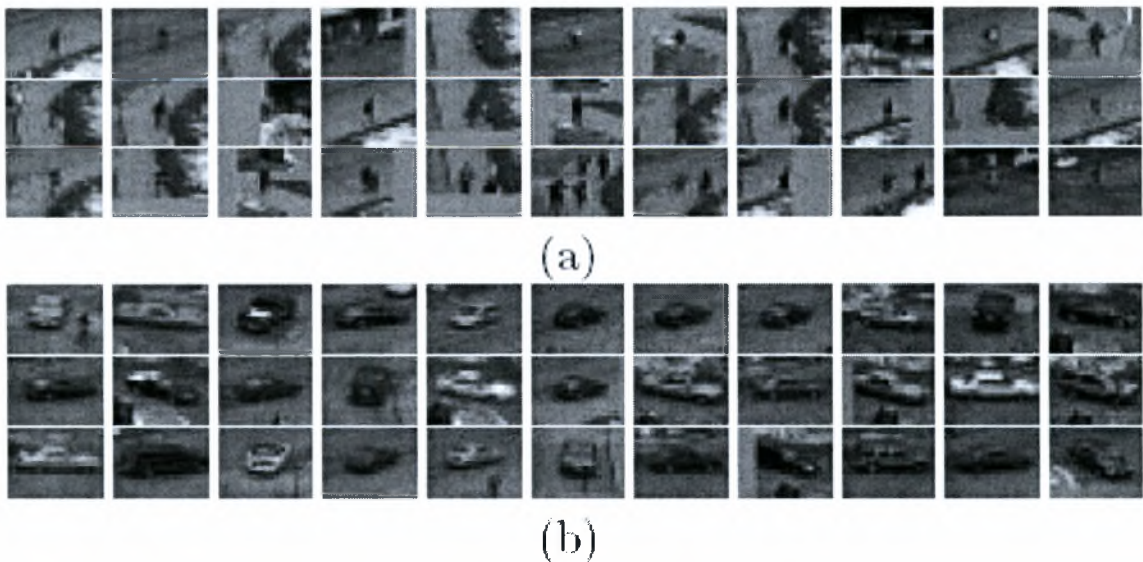
Εξ' αιτίας της σταθερότητας και της πληρότητας της αναπαράστασης είναι πιθανόν να γίνουν κάποιες κατηγοριοποιήσεις. Στην Εικόνα 19 φαίνεται η κατηγοριοποίηση αντικειμένων που εμφανίστηκαν στο σκηνικό για πάνω από δέκα λεπτά χρησιμοποιώντας ένα απλό δυαδικό κατώφλι στον μέσο χρονικά όρο του aspect ratio¹¹ του αντικειμένου. Ίχνη ή σημάδια των οποίων η διάρκεια ήταν λιγότερο από ένα δευτερόλεπτο απομακρύνονταν.

Στο παραπάνω πείραμα, κάθε αντικείμενο που εισέρχονταν στη σκηνή – συνολικά 33 αυτοκίνητα και 34 άτομα – ανιχνεύθηκαν. Το σύστημα πολύ αποτελεσματικά κατηγοριοποίησε κάθε αυτοκίνητο, εκτός μιας περίπτωσης κατά την οποία η κατηγοριοποίηση δύο αυτοκινήτων έγινε σαν ένα, αφού το ένα αυτοκίνητο εισήλθε στο σκηνικό την ίδια στιγμή που ένα άλλο έβγαινε από το ίδιο σημείο. Σε δύο περιπτώσεις υπήρχαν δύο άτομα που περπατούσαν μαζί με φυσική επαφή. Παρόλα αυτά, το σύστημα ανίχνευσε λανθασμένα μόνο το ένα άτομο. Επίσης, μέτρησε διπλά δύο αντικείμενα επειδή τα ίχνη τους δεν αντιστοιχίζονταν κατάλληλα.

¹¹ Aspect ratio ενός αντικείμενου δύο διαστάσεων είναι ο λόγος της μεγαλύτερης διάσταση προς την μικρότερη.



Εικόνα 18: Η εικόνα αυτή δείχνει συνεχόμενες ώρες παρακολούθησης από τις 6π.μ. – 9π.μ. και από τις 3μ.μ. – 7μ.μ. (a) εικόνα την στιγμή που αποθηκεύτηκε και (b) τα αθροιστικά ίχνη των αντικειμένων στην πάροδο του χρόνου. Τα χρώματα δείχνουν την κατεύθυνση και η ένταση δείχνει το μέγεθος. Η συνοχή των χρωμάτων σε συγκεκριμένες περιοχές αντικατοπτρίζει την συνοχή της ταχύτητας, της κατεύθυνσης και του μεγέθους που παρατηρήθηκαν [1].



Εικόνα 19: Η εικόνα δείχνει ποια αντικείμενα στο σκηνικό κατηγοριοποιήθηκαν σαν άνθρωποι ή αυτοκίνητα χρησιμοποιώντας απλούς ευριστικούς κανόνες. Η ακρίβεια αντικατοπτρίζεται στην συνοχή των συνεκτικών περιοχών που ανιχνεύτηκαν [1].

3.3.8 Εφαρμοσιμότητα

Η πιο σημαντική πληροφορία που πρέπει να λάβει υπ' όψη του ένας ερευνητής όταν αποφασίζει να υλοποιήσει έναν ανιχνευτή είναι πού μπορεί αυτός να εφαρμοστεί.

Η μεγαλύτερη δυσκολία αυτών των συστημάτων είναι στην παρουσία αντικειμένων που αλληλοκαλύπτονται οπτικά. Ο ανιχνευτής πολλαπλής υπόθεσης δεν είναι αρκετά «έξυπνος» για τον αξιόπιστο διαχωρισμό αντικειμένων που αντιτίθενται. Αυτό το πρόβλημα μπορεί να γίνει μεγαλύτερο με την παρουσία μεγάλων σκιών. Αλλά για τις συγκεκριμένες εφαρμογές ήταν περισσότερο επιθυμητό να ανιχνευτεί ένα αντικείμενο με τη σκιά του και να αποφευχθούν τα σκοτεινά αντικείμενα, παρά να γίνει προσπάθεια στην αφαίρεση σκιών. Στις καθαρές και φωτεινές μέρες όπου οι σκιές είναι πιο σημαντικές, τόσο οι σκιερές περιοχές όσο και οι σκιασμένες πλευρές σκοτεινών αντικειμένων ήταν μαύρες.

Το σύστημα εμφανίστηκε αρκετά σθεναρό σε όλα, εκτός από τις γρήγορες αλλαγές του φωτισμού. Παρακολούθησε με επιτυχία εξωτερικές σκηνές με βροχή, χιόνι, χαλάζι, συννεφιά αλλά και λιακάδα. Στάθηκε ικανό στην ανίχνευση πουλιών καθισμένα σε κτήρια, ποντίκια την νύχτα, άτομα να εισέρχονται σε κτήρια και αντικείμενα σε εξωτερικά σκηνικά. Επιπλέον, δεν υπήρξε πρόβλημα στους κύκλους της μέρας / νύχτας και σε μεγάλης διάρκειας αλλαγές του σκηνικού .

Βέβαια, έμφαση μπορεί να δοθεί στο μέλλον, στην αύξηση της ταχύτητας του αλγορίθμου, σε μεγαλύτερου μεγέθους frames, και στην χρήση περισσότερων gaussians κατανομών στο συνδυαστικό μοντέλο. Όλα αυτά μπορούν κατά πολύ να αυξήσουν την απόδοση του συστήματος, όπως επίσης η χρήση ενός πλήρη covariance πίνακα ή η πρόσθεση της πρόγνωσης σε κάθε gaussian (προσέγγιση Kalman filter).

Επιπροσθέτως, υπό εξέταση τίθενται και κάποιες άλλες πιθανές βελτιστοποιήσεις. Χαρακτηριστική περίπτωση είναι κάποιες σχέσεις τιμών σε γειτονικά pixels και συσχετίσεις των κατανομών τους. Με την συγκεκριμένη μέθοδο χρησιμοποιήθηκαν τα χρωματικά μοντέλα grayscale, RGB και HSV. Οι υποθέσεις και οι ευριστικοί μηχανισμοί που χρησιμοποιήθηκαν όμως, καθιστούν το σύστημα δυνατό να επεξεργαστεί οποιαδήποτε μορφή εισόδου.

Στο παρελθόν ήταν αναγκαστικός ο χειρισμός σχετικά λίγων δεδομένων αλλά με την υπάρχουσα τεχνολογία και το συγκεκριμένο σύστημα, γίνεται εφικτή η

συλλογή εικόνων κινούμενων αντικειμένων και η παρακολούθησή τους σε πραγματικού χρόνου videos για εβδομάδες.

3.4 Μοντελοποίηση Προσαρμόσιμου Έγχρωμου Παρασκηνίου για Ροές Δεδομένων Video Πραγματικού Χρόνου και Ανίχνευση Σκιών

3.4.1 Γενικά

Σε αυτό το υποκεφάλαιο θα αναλυθεί και θα παρουσιαστεί αναλυτικά μία μέθοδος μοντελοποίησης ενός προσαρμόσιμου παρασκηνίου με πολλές κατανομές. Αυτή θα χρησιμοποιηθεί στην πραγματικού χρόνου ανίχνευση αντικειμένων αλλά και στην ανίχνευση σκιών. Τέλος, θα γίνουν συγκρίσεις και με άλλους ανάλογους αλγορίθμους και μεθόδους.

3.4.2 Μοντελοποίηση του Background

Η διαδικασία μας είναι πλέον γνωστή από τις προηγούμενες αναφορές που έγιναν. Η μέθοδος μοντελοποιεί κάθε ένα *pixel* του παρασκηνίου από έναν συνδυασμό K gaussian κατανομών (K συνήθως από 3 έως 5) και υποθέτουμε ότι διαφορετικές gaussians αναπαριστούν διαφορετικά χρώματα. Οι συνδυασμοί περιέχουν παραμέτρους βαρύτητας που αντιπροσωπεύουν ποσοστά του χρόνου που τα συγκεκριμένα χρώματα παρέμειναν στο σκηνικό. Οι συνιστώσες του παρασκηνίου καθορίζονται υποθέτοντας ότι το background περιέχει B πιο πιθανά χρώματα, δηλαδή αυτά που παραμένουν περισσότερο στο σκηνικό. Στατικά μονόχρωμα αντικείμενα έχουν την μορφή πυκνών και κλειστών συστάδων (clusters) ενώ τα κινούμενα αντικείμενα έχουν την μορφή πιο διευρυμένων συστάδων, εξ' αιτίας των διαφορετικών ανακλώμενων επιφανειών κατά τη διάρκεια της κίνησης. Η μέτρηση αυτή καλείται ως τιμή fitness [2]. Η μέτρηση αυτή δείχνει διαισθητικά την πιθανότητα ένα αντικείμενο να ανήκει στο παρασκήνιο και να ταιριάζει (fit) σε μία κλειστή συστάδα. Ορίζεται ως ο λόγος του βάρους της k -οστής gaussian συνιστώσας w_k προς την τυπική απόκλιση σ_k της ίδιας συνιστώσας. Για να επιτευχθεί η προσαρμοστικότητα του μοντέλου στις αλλαγές του φωτισμού και η λειτουργία του

σε πραγματικού χρόνου δεδομένα, εφαρμόστηκε ένα σχήμα ανανέωσης που βασίζεται πάνω σε επιλεκτική ανανέωση. Κάθε τιμή νέου pixel ελέγχεται με τις συνιστώσες του υπάρχοντος μοντέλου σε σειρά, ανάλογα με την τιμή fitness. Η πρώτη συνιστώσα του μοντέλου που ταιριάζει θα ανανεωθεί. Αν δεν υπάρξει ταιρίασμα, τότε δημιουργείται μια νέα gaussian με μέσο την τιμή του pixel, μια μικρής τιμής παράμετρο βάρους και έναν μεγάλο πίνακα covariance.

3.4.3 Προσαρμόσιμο Συνδυαστικό Μοντέλο Gaussian

Κάθε pixel του σκηνικού, όπως αναφέρθηκε, μοντελοποιείται από ένα μείγμα K gaussian κατανομών. Η πιθανότητα ότι ένα συγκεκριμένο pixel έχει την τιμή x_N την στιγμή N μπορεί να γραφεί ως:

$$p(x_N) = \sum_{j=1}^K w_j \eta(x_N; \theta_j)$$

όπου $\eta(x; \theta_k)$ είναι η κανονική κατανομή της k -οστής συνιστώσας:

$$\eta(x; \theta_k) = \eta(x; \mu_k, \Sigma_k) = \frac{1}{(2\pi)^{\frac{D}{2}} |\Sigma_k|^{\frac{1}{2}}} e^{-\frac{1}{2} (x-\mu_k)^T (\Sigma_k)^{-1} (x-\mu_k)}$$

όπου μ_k είναι ο μέσος και $\Sigma_k = (\sigma_k)^2 I$ είναι ο πίνακας covariance της k -οστής συνιστώσας, όταν οι τιμές των pixels είναι ανεξάρτητες τυχαίες μεταβλητές.

Οι K κατανομές είναι ταξινομημένες με βάση την τιμή fitness w_k/σ_k και οι πρώτες B κατανομές χρησιμοποιούνται σαν μοντέλο του παρασκηνίου όπου το B μπορεί να εκτιμηθεί από:

$$B = \operatorname{argmin}_b \left(\sum_{j=1}^b w_j > T \right)$$

Το κατώφλι T είναι η ελάχιστη εκ των προτέρων πιθανότητα το παρασκήνιο να είναι μέσα στο σκηνικό. Η αφαίρεση του background επιτελείται μαρκάροντας ένα pixel ως pixel του προσκηνίου εφόσον αυτό είναι περισσότερο από 2.5 σταθερές αποκλίσεις [2] μακριά από οποιαδήποτε από τις B κατανομές. Η πρώτη gaussian

συνιστώσα η οποία θα αντιστοιχηθεί, θα ανανεωθεί σύμφωνα με τις ακόλουθες εξισώσεις:

$$\begin{aligned}w_k^{N+1} &= (1-a)w_k^N + a p(\omega_k | x_{N+1}) \\ \mu_k^{N+1} &= (1-a)\mu_k^N + \rho x_{N+1} \\ \Sigma_k^{N+1} &= (1-a)\Sigma_k^N + \rho(x_{N+1} - \mu_k^{N+1})(x_{N+1} - \mu_k^{N+1})^T \\ \rho &= a\eta(x_{N+1}; \mu_k^N, \Sigma_k^N)\end{aligned}$$

και $p(\omega_k | x_{N+1}) = \{1 - \text{αν είναι η πρώτη gaussian συνιστώσα που ταιριάζει, } 0 - \text{αλλού}\}.$

όπου ω_k είναι η k -οστή gaussian συνιστώσα και $1/a$ δηλώνει την χρονική σταθερά που καθορίζει την αλλαγή. Εάν καμία από τις K κατανομές δεν ταιριάζει στην τιμή του pixel, η λιγότερο πιθανή συνιστώσα αντικαθίσταται από μια κατανομή με μέσο την τιμή του pixel, μια αρχικά υψηλή διακύμανση και μια χαμηλή τιμή βαρύτητας.

Η μέθοδος αυτή αν και αρκετά στιβαρή έχει κάποιες ανικανότητες. Καταρχάς, αν η πρώτη τιμή ενός δεδομένου pixel προέρχεται από αντικείμενο του προσκηνίου, υπάρχει μόνο μία gaussian που το βάρος της συμφωνεί. Μόνο με μονόχρωμες διαδοχικές τιμές του παρασκηνίου απαιτείται $\log_{(1-a)}T$ frames μέχρι το αληθινό background να θεωρηθεί σαν background από το σύστημα και $\log_{(1-a)}0.5$ frames μέχρι να γίνει η κυρίαρχη συνιστώσα του background.

Για παράδειγμα, αν υποθέσουμε ότι το λιγότερο 60% του χρόνου το παρασκήνιο είναι παρόν και $a=0.002$ (χρησιμοποιούμε τα 500 πρόσφατα frames), θα χρειαστούν 255 frames και 346 frames αντίστοιχα, ώστε η συνιστώσα να συμπεριληφθεί ως μέρος του background και για να γίνει κυρίαρχη συνιστώσα. Η κατάσταση αυτή μπορεί να γίνει ακόμα χειρότερη σε περιβάλλοντα με συχνή και πυκνή κίνηση, όπου η εμφάνιση ενός καθαρού παρασκηνίου είναι σπάνιο φαινόμενο. Θα παρουσιαστεί μια λύση του προβλήματος αυτού στην επόμενη ενότητα.

Μια δεύτερη περίπτωση είναι η εξής: το ρ είναι πολύ μικρό λόγω του παράγοντα πιθανοφάνειας. Αυτό οδηγεί σε πολύ αργές προσαρμογές στους μέσους και στους πίνακες covariance, και έτσι ο ανιχνευτής μπορεί να αποτύχει μέσα σε λίγα μόλις δευτερόλεπτα από την εκκίνηση. Μια λύση σε αυτό είναι να απαλείψουμε από την πιθανοφάνεια το ρ .

3.4.4 Σειριακοί Αλγόριθμοι EM

Η εκτίμηση του συνδυαστικού gaussian μοντέλου ξεκινά από αναμενόμενες, ικανές, στατιστικές εξισώσεις ανανέωσης και έπειτα μεταβαίνουμε στην μέθοδο του «παραθύρου των L -πρόσφατων» (L -recent window), όταν τα L πρώτα δείγματα επεξεργαστούν. Οι εξισώσεις που αναφέρθηκαν παρέχουν μια καλή αρχική εκτίμηση πριν συλλεχθούν όλα τα L δείγματα. Η αρχική εκτίμηση βελτιώνει την απόδοση του ανιχνευτή επιτρέποντας ταχύτερη σύγκλιση σε ένα ευσταθές μοντέλο παρασκηνίου. Οι εξισώσεις ανανέωσης του L -recent window δίνουν προτεραιότητα στα πρόσφατα δεδομένα παρέχοντας την δυνατότητα στο σύστημα να προσαρμόζεται στις αλλαγές του περιβάλλοντος, και να μην επηρεάζονται αρνητικά οι εκτιμήσεις του από παλαιότερες τιμές που τώρα είναι άκυρες / αταίριαστες.

Παρακάτω, φαίνονται οι εξισώσεις από τον αλγόριθμο του EM πάνω σε ικανές στατιστικές καθώς επίσης και οι εξισώσεις της έκδοσης του L -recent window.

- Update Equations from EM algorithms:

$$w_k^{N+1} = w_k^N + \frac{1}{N+1} (p(\omega_k | x_{N+1}) - w_k^N)$$

$$\mu_k^{N+1} = \mu_k^N + \frac{p(\omega_k | x_{N+1})}{\sum_{i=1}^{N+1} p(\omega_k | x_i)} (x_{N+1} - \mu_k^N)$$

$$\Sigma_k^{N+1} = \Sigma_k^N + \frac{p(\omega_k | x_{N+1})}{\sum_{i=1}^{N+1} p(\omega_k | x_i)} ((x_{N+1} - \mu_k^N)(x_{N+1} - \mu_k^N)^T - \Sigma_k^N)$$

- Update Equations from L -recent window:

$$w_k^{N+1} = w_k^N + \frac{1}{L} (p(\omega_k | x_{N+1}) - w_k^N)$$

$$\mu_k^{N+1} = \mu_k^N + \frac{1}{L} \left(\frac{p(\omega_k | x_N) x_{N+1}}{w_k^{N+1}} - \mu_k^N \right)$$

$$\Sigma_k^{N+1} = \Sigma_k^N + \frac{1}{L} \left(\frac{p(\omega_k | x_{N+1}) (x_{N+1} - \mu_k^N)(x_{N+1} - \mu_k^N)^T}{w_k^{N+1}} - \Sigma_k^N \right)$$

3.4.5 Ανίχνευση Σκιών και Χρωματικό Μοντέλο

Υπάρχουν πολλές μέθοδοι οι οποίες δεν είναι δυνατόν να αναγνωρίσουν κινούμενες σκιές από τα αντικείμενα που τις δημιουργούν. Αυτό οφείλεται στο ότι

δεν υπάρχουν ευριστικοί μηχανισμοί ώστε να μαρκάρουν gaussian συνιστώσες ως κινούμενες σκιές. Μια λύση σε αυτό είναι να χρησιμοποιηθεί μια αναπαράσταση χρωματικού χώρου που μειώνει την ευαισθησία. Πολλά χρωματικά μοντέλα μπορούν να διαχωρίσουν τις συνιστώσες της χρωματικότητας και της φωτεινότητας. Διατηρώντας ένα χρωματικό μοντέλο που δεν λαμβάνει υπ' όψιν του την λάμψη της εικόνας μπορεί να οδηγήσει σε ασταθή μοντέλα ειδικά για πολύ φωτεινά (λαμπερά) ή πολύ σκοτεινά αντικείμενα. Αυτή η μετατροπή επίσης απαιτεί περισσότερους υπολογιστικούς πόρους ανάλογα πάντα και με το μέγεθος της εικόνας.

Η ιδέα της διαφύλαξης των συνιστωσών έντασης και εξοικονόμησης υπολογιστικών πόρων οδηγεί πάλι στο χρωματικό μοντέλο του RGB. Για την απαίτηση της αναγνώρισης κινούμενων σκιών, πρέπει να βρεθεί ένα χρωματικό μοντέλο που να διαχωρίζει τις χρωματικές και τις συνιστώσες λαμπρότητας. Επιπρόσθετα, θα πρέπει να είναι συμβατό και να χρησιμοποιείται από το υπάρχον συνδυαστικό μοντέλο. Αυτό γίνεται συγκρίνοντας ένα pixel του προσκηνίου με τις τρέχουσες συνιστώσες του παρασκηίου. Αν η διαφορά και στις δύο συνιστώσες (χρωματική, λαμπρότητας) είναι μεταξύ κάποιων κατωφλίων, το pixel θεωρείται σκιά. Χρησιμοποιείται ένα αποτελεσματικό, υπολογιστικό, χρωματικό μοντέλο που αποτελείται από ένα διάνυσμα θέσης στον RGB μέσω του pixel του background, E , μια αναμενόμενη γραμμή χρωματικότητας (expected chromaticity line), $||E||$, ένας δείκτης χρωματικής παραποίησης, d , και ένα κατώφλι λαμπρότητας, τ . Για μία δεδομένη τιμή ενός pixel, I , μια παραποίηση λαμπρότητας (brightness distortion), α , και μια παραποίηση χρωμάτων (colour distortion), c , από το μοντέλο του παρασκηίου μπορούν να υπολογιστούν ως:

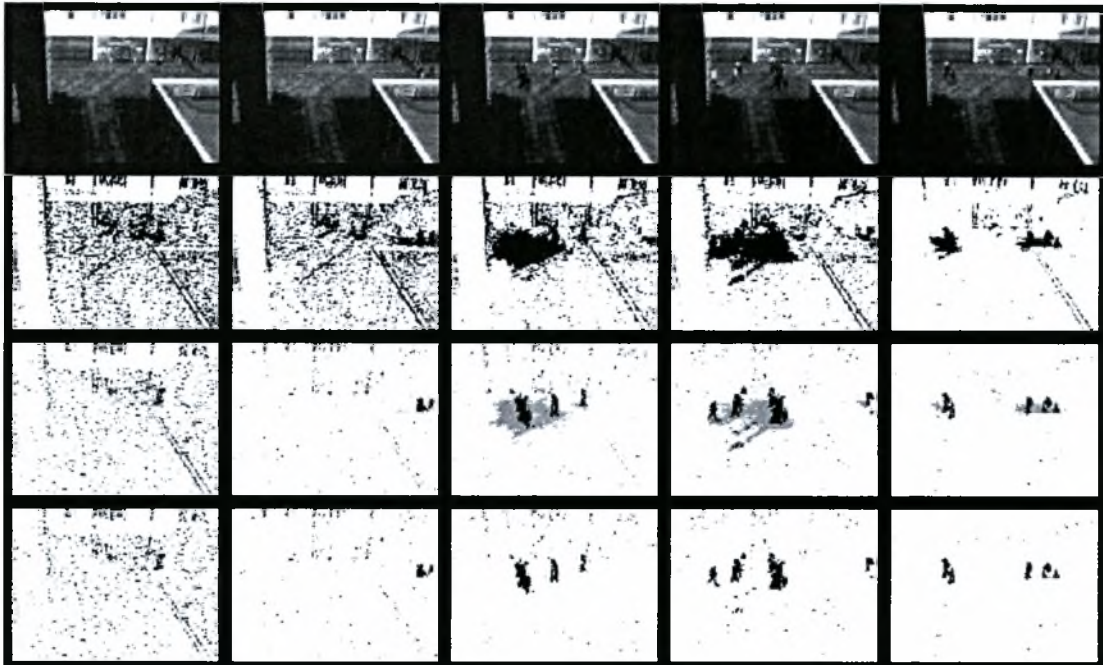
$$\alpha = \operatorname{argmin}_z (I - zE)^2 \text{ και} \\ c = ||I - \alpha E||$$

Με την υπόθεση της σφαιρικής gaussian κατανομής σε κάθε συνδυαστική συνιστώσα, η σταθερή απόκλιση της k -οστής συνιστώσας σ_k μπορεί να εξισωθεί με το d . Ο υπολογισμός των α και c είναι τετριμμένος χρησιμοποιώντας εσωτερικό γινόμενο διανυσμάτων. Ένα δείγμα εκτός του παρασκηίου θεωρείται κινούμενη σκιά αν το α είναι εντός δύο αποκλίσεων και $\tau < c < I$.

3.4.6 Πειράματα και Αποτελέσματα

Σε αυτήν την ενότητα γίνεται λόγος και σύγκριση της απόδοσης του αλγορίθμου που αναφέρθηκε και ενός παραπλήσιου (§3.3 - Grimson [1]). Η ανάλυση της ακολουθίας των εικόνων είναι 192×144 . Χρησιμοποιήθηκε το συνδυαστικό μοντέλο 5 gaussian συνιστωσών. Το L ορίστηκε στα 500 frames ($\alpha=0.002$ για την μέθοδο του Grimson) και το κατώφλι T ίσο με 0.6. Στην μέθοδο της ανίχνευσης σκιών το κατώφλι τ ορίστηκε στο 0.7.

Στα αποτελέσματα εξαγωγής του background δεν χρησιμοποιήθηκαν διαδικασίες υψηλότερου επιπέδου, όπως αφαίρεσης θορύβου ή αλγόριθμοι ανάλυσης συνεκτικών συνιστωσών. Αυτή η διαδικασία έγινε με σκοπό να φανεί η απόδοση των μοντέλων του παρασκήνιου. Η Εικόνα 20 δείχνει μια ακολουθία εξωτερικών σκηνών που περιέχουν ανθρώπους να περπατούν σε δημόσιο δρόμο. Η ακολουθία περιέχει δυνατή ηλιοφάνεια, μεγάλες σκιερές περιοχές, δένδρα, ανακλάσεις από παράθυρα και κινούμενες σκιές. Παρουσιάζονται εικόνες από την αρχικοποίηση της μεθόδου και εικόνες με πυκνή και διαρκή κίνηση. Εξ' αιτίας του μη καθαρισμού των εικόνων από τον θόρυβο στην αρχικοποίηση, ο αλγόριθμος του Grimson είχε κάποια προβλήματα τουλάχιστον για τα πρώτα 100 frames. Η μέθοδος που παρουσιάστηκε εδώ επιτυγχάνει πολύ καλή τμηματοποίηση. Η απόδοση αυξάνεται ακόμα περισσότερο με την μονάδα ανίχνευσης σκιών, η οποία μειώνει το φαινόμενο των μικρών επαναληπτικών κινήσεων στο παρασκήνιο.



Εικόνα 20: Η πρώτη σειρά δείχνει την αρχική ακολουθία των frames 15, 105, 235, 290 και 1200 αντίστοιχα. Η δεύτερη σειρά δείχνει τα αποτελέσματα του αλγορίθμου του Grimson και οι δύο τελευταίες σειρές είναι τα αποτελέσματα της παρουσιαζόμενης μεθόδου με και χωρίς τις κινούμενες σκιές να φαίνονται στις εικόνες. Οι σκιές απεικονίζονται με γκρι χρώμα [2].

Κεφάλαιο 4 - Πειράματα και Αναλύσεις



4.1 Δομή και Ακολουθία των Πειραμάτων

Σε αυτό το κεφάλαιο, θα επικεντρωθούμε στην υλοποίηση και την παρουσίαση μεθόδων και αλγορίθμων που παρουσιάστηκαν μέχρι τώρα. Αρχικά, θα παρουσιαστούν μέθοδοι απλοί στην υλοποίηση και όχι τόσο ακριβείς στα αποτελέσματά τους και σταδιακά θα εμβαθύνουμε σε πιο πολύπλοκες, με περισσότερες απαιτήσεις στην υλοποίηση τους αλλά και σε διαθέσιμους υπολογιστικούς πόρους.

Θα χρησιμοποιηθούν πολλά διαφορετικά είδη δεδομένων εισόδου-video όσον αφορά, τη διάρκειά τους, το μέγεθος των frames, την σταθερότητα της βιντεοκάμερας με την οποία έγινε η λήψη και την κίνηση των αντικειμένων που περιέχουν.

4.2 Προσέγγιση Background με Median Filtering

4.2.1 Γενικά για το Median Filter

Το median filter (φίλτρο μέσου) είναι μια τεχνική ψηφιακού φιλτραρίσματος που χρησιμοποιείται συχνά στην αφαίρεση θορύβου από εικόνες ή από άλλα σήματα. Επίσης, χρησιμοποιείται στην ανίχνευση ακμών (edge detection) σε μια εικόνα. Γενικά, είναι μια μη γραμμική μέθοδος βελτίωσης την απεικόνισης ενός frame που βοηθάει πολύ σε περαιτέρω επεξεργαστικά βήματα της (high-level processing).

Η ιδέα είναι η εξέταση ενός δείγματος δεδομένων κάθε φορά και η απόφαση, αν αυτό είναι αντιπροσωπευτικό της εικόνας. Αυτό γίνεται εφικτό, εφαρμόζοντας ένα παράθυρο (πίνακα ενός ή δύο διαστάσεων - μάσκα) πάνω στα δεδομένα μας (τιμές

των pixels της εικόνας). Οι τιμές που βρίσκονται μέσα στη μάσκα ταξινομούνται αριθμητικά και η τιμή που βρίσκεται στη μέση της ταξινόμησης είναι η median τιμή. Αν υπάρχει μία τιμή στη μάσκα που διαφέρει πολύ από αυτήν, τότε αυτή αντικαθίσταται από την median. Αυτό θα γίνει ξεκάθαρο με τα παρακάτω δύο παραδείγματα [20] [21]. Στο πρώτο, φαίνεται η εύρεση της τιμής median σαν εφαρμογή μιας συνάρτησης πάνω σε κάποιες τιμές και στο δεύτερο έχουμε την εφαρμογή ενός median filter στα pixels μιας εικόνας.

$$\text{Median}[2 \ 80 \ 6] = \text{Median}[2 \ 6 \ 80] = 6$$

$$\text{Median}[80 \ 6 \ 3] = \text{Median}[3 \ 6 \ 80] = 6$$

$$\text{Median}[6 \ 3 \ 3] = \text{Median}[3 \ 3 \ 6] = 3$$

123	125	126	130	140
122	124	126	127	135
118	120	150	125	134
119	115	119	123	133
111	116	110	120	130

Neighbourhood values:

**115, 119, 120, 123, 124,
125, 126, 127, 150**

Median value: 124

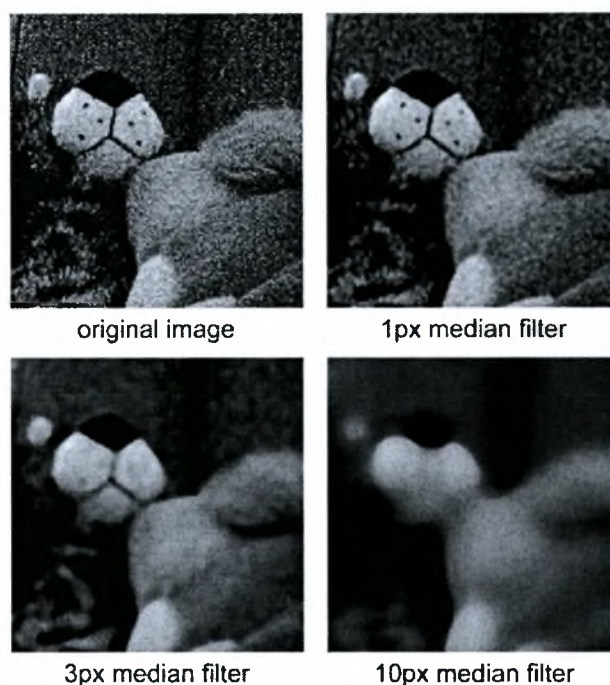
Σε αυτό το παράδειγμα εφαρμόζεται ένα median filter 3×3. Υπολογίζοντας την τιμή median = 124, παρατηρούμε ότι απέχει πολύ από την κεντρική τιμή (pixel) της μάσκας (150) και θα αντικατασταθεί από την τιμή 124.

Με λίγα λόγια, το median filter εξομαλύνει μια εικόνα (smoothing) αφαιρώντας τιμές που διαφέρουν πολύ από την median (outliers). Διαισθητικά, αλλάζει την τιμή ενός pixel (του κεντρικού κάθε φορά στη μάσκα), ώστε να μοιάζει με τα γειτονικά του. Αυτή η τιμή, τις περισσότερες φορές αντιπροσωπεύει θόρυβο, ο οποίος πρέπει να εξαλειφθεί.

Το βασικό πρόβλημα με το φίλτρο αυτό (και με άλλα παρόμοια φίλτρα που χρησιμοποιούν μάσκα), είναι το πώς χειρίζεται τα άκρα της εικόνας. Ένας τρόπος είναι να μην επιτρέψουμε στη μάσκα να εφαρμοστεί σε pixels που βρίσκονται κοντά στα όρια της εικόνας, αλλά τότε η φιλτραρισμένη εικόνα θα είναι πιο μικρή από την αρχική. Ένας άλλος τρόπος είναι να χρησιμοποιήσουμε μόνο τις τιμές της μάσκας

που περιέχονται στην εικόνα. Δηλαδή όταν ένα μέρος της μάσκας «πέφτει» έξω από τα όρια της εικόνας τότε δεν λαμβάνεται υπ' όψιν. Ο τρίτος τρόπος είναι να γίνει μια επέκταση της εικόνας με μηδενικά pixels γύρω της (zero padding) και η μάσκα να εφαρμοστεί κανονικά.

Τέλος, αξίζει να σημειωθεί ότι όσο πιο μεγάλων διαστάσεων είναι το median filter (η μάσκα που εφαρμόζεται), τόσο πιο μεγάλος είναι και ο βαθμός εξομάλυνσης της εικόνας που επιτυγχάνεται και αυτό φαίνεται στην παρακάτω εικόνα.



Εικόνα 21: Παράδειγμα εφαρμογής τριών median filters διαφορετικών διαστάσεων στην ίδια αρχική εικόνα [20].

4.2.2 Εφαρμογή του Median Filter για Background Subtraction

Η ιδέα εφαρμογής του median filter στην εξαγωγή του background είναι απλή και εύκολα υλοποιήσιμη. Ωστόσο, όπως θα δούμε παρακάτω, δεν πρόκειται για την υλοποίηση κάποιου σθεναρού συστήματος, που λειτουργεί κάτω υπό οποιεσδήποτε συνθήκες· αλλά για μία μέθοδο που μπορεί να δώσει ικανοποιητικά αποτελέσματα για συγκεκριμένα δείγματα δεδομένων. Αυτή είναι μια καλή αρχή στην κατανόηση του background subtraction.

Μία χαρακτηριστική περίπτωση, είναι η εφαρμογή της συνάρτησης median του προγραμματιστικού περιβάλλοντος Matlab, σε κάθε μία από τις τρεις συνιστώσες

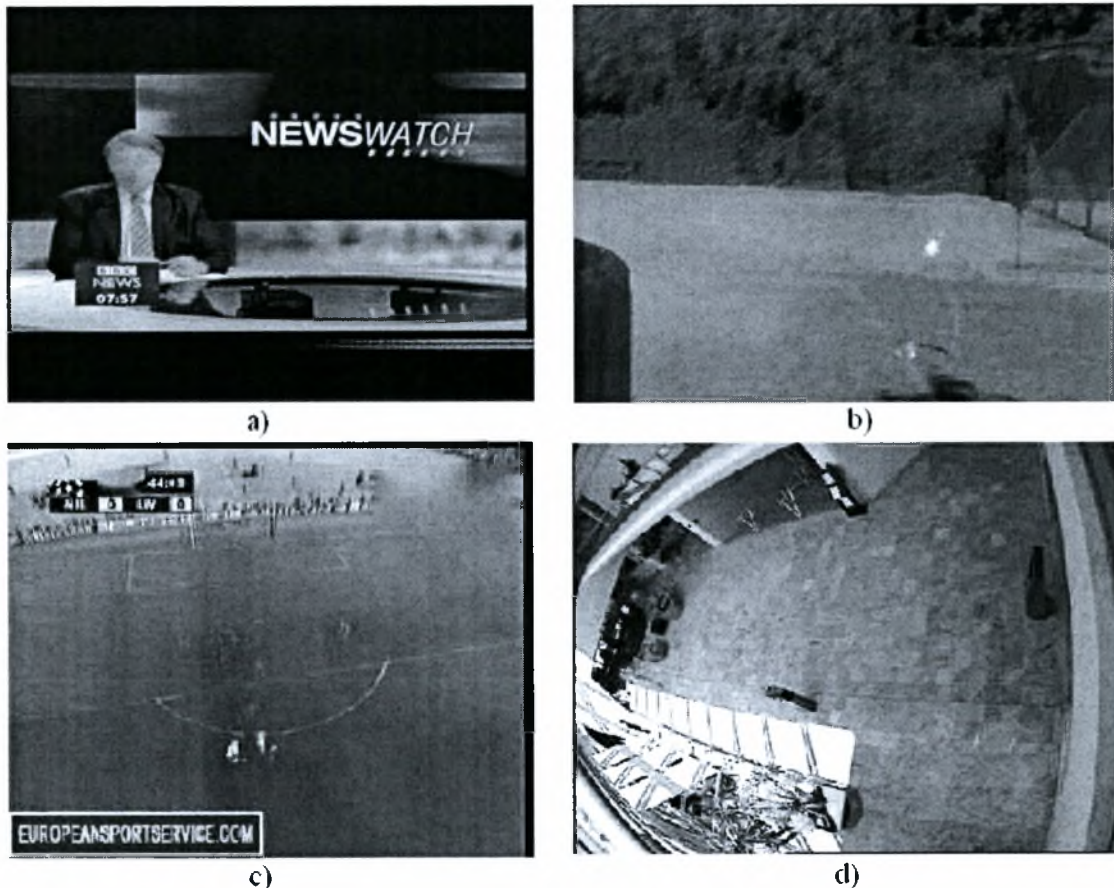
χρωμάτων της εικόνας (R,G,B) για όλα τα frame ενός video. Στην πράξη έχουμε μια ακολουθία από frames ενός video, τα οποία χωρίζονται σε 3 διαφορετικές ακολουθίες frames. Κάθε μία από αυτές περιέχει μία από τις τρεις συνιστώσες χρώματος της πραγματικής ακολουθίας (τρεις 3D πίνακες μεγέθους = [frames × Xsize × Ysize], όπου frames το πλήθος των frames και Xsize, Ysize το μέγεθος της εικόνας στον X και Y άξονα αντίστοιχα). Η συνάρτηση median, αν εφαρμοστεί με μόνο όρισμα έναν πίνακα, τότε εκτελείται ως προς την πρώτη διάσταση. Αν λοιπόν καλέσουμε την median για τους τρεις αυτούς 3D-πίνακες, θα εφαρμοστεί ως προς όλα τα frames και σαν αποτέλεσμα θα πάρουμε τρεις νέους πίνακες διάστασης [1 × Xsize × Ysize]. Αυτοί οι τρεις πίνακες, αν ενωθούν πάλι σε έναν [3 × Xsize × Ysize] πίνακα, θα αποτελούν την εκτίμηση του background για τα δεδομένα μας.

Αυτό συμβαίνει, επειδή η κίνηση που περιέχεται στο frame είναι στην πραγματικότητα κάποια κινούμενα pixels, των οποίων η τιμές διαφέρουν κατά πολύ από τις τιμές του background. Αν το background είναι «παρών» στην μεγαλύτερη διάρκεια του video, τότε είναι δυνατόν η εφαρμογή του median filter να δώσει πολύ ικανοποιητικά αποτελέσματα, γιατί οι τιμές των pixels των κινούμενων αντικειμένων θα αντικατασταθούν από την median τιμή, που αντιπροσωπεύει τιμές του παρασκηνίου.

Ένα pixel για παράδειγμα θα έχει μια σταθερή τιμή (από το σταθερό background) για αρκετά frames, και θα αλλάζει μόνο για ορισμένα (όταν εισέρχεται ένα κινούμενο αντικείμενο). Αν διατάξουμε τις τιμές αυτές αριθμητικά, τότε η median τιμή θα είναι, με μεγάλη πιθανότητα, η τιμή του pixel στο background. Αν το δούμε μαθηματικά, πρέπει το κάθε pixel να έχει την τιμή του background στα μισά συν ένα frames, ώστε η διάταξή τους να έχει ως μεσαία τιμή αυτή του pixel στο background. Στην παρακάτω εικόνα φαίνονται τα αποτελέσματα της διαδικασίας που μόλις περιγράφηκε για τέσσερα διαφορετικά videos (Εικόνα 21).

Το πρώτο video [Εικόνα 22 a)] είναι διάρκειας 21 δευτερολέπτων, όπου φαίνεται ένας παρουσιαστής ειδήσεων να μιλάει προς την κάμερα. Τα είδη των κινήσεων εδώ είναι δύο. Το πρώτο είναι οι κινήσεις του προσώπου του παρουσιαστή (το υπόλοιπο σώμα ήταν σταθερό καθ' όλη τη διάρκεια του video), οι οποίες δημιούργησαν μια νοητή περιοχή μέσα στην οποία βρισκόταν συνεχώς το πρόσωπο του παρουσιαστή. Αυτό είχε σαν αποτέλεσμα το χρώμα του δέρματος να «κυριαρχεί» σε όλη την διάρκεια του video, στην συγκεκριμένη περιοχή. Το δεύτερο είδος

κίνησης που υπήρξε ήταν μια εμφάνιση κυλιόμενων μηνυμάτων στα κάτω (μαύρο πλαίσιο) της εικόνας.



Εικόνα 22: Αποτελέσματα background subtraction από την εφαρμογή του median filtering σε τέσσερα διαφορετικά videos.

Η εμφάνιση αυτή ήταν περιοδική και τα γράμματα κινούνταν γρήγορα πάνω στο πλαίσιο, με αποτέλεσμα το μαύρο χρώμα να είναι, σε πολύ μεγάλο ποσοστό του χρόνου, «παρών» στα συγκεκριμένα pixels. Έτσι σωστά ανιχνεύθηκε το background της περιοχής αυτής.

Το δεύτερο video [Εικόνα 22 b)], διάρκειας μόλις τριών δευτερολέπτων, δείχνει μία παίκτρια του τένις, η οποία χτυπάει τη μπάλα με την ρακέτα της και έπειτα στρίβει και κάνει δύο βήματα προς τα δεξιά. Το background εδώ ανιχνεύθηκε ικανοποιητικά. Ωστόσο, λόγω του μικρού δείγματος, υπάρχουν κάποια λευκά «υπολείμματα» από την ενδυμασία της παίκτριας, επειδή αυτά υπήρχαν εκεί σε μεγάλο ποσοστό του χρόνου. Επιπλέον, παρατηρείται ένα θόλωμα στο δάπεδο και στα δέντρα στο πίσω μέρος, εξ' αιτίας μιας μικρής κίνησης στην κάμερα λήψης.

Στο τρίτο video [Εικόνα 22 c)], διάρκειας 10 δευτερολέπτων, έχουμε παρόμοια αποτελέσματα με την Εικόνα 22 a). Πρόκειται για μία ποδοσφαιρική φάση κατά την οποία γίνεται η προετοιμασία για την εκτέλεση ελεύθερου σουτ. Παίχτες από την μία ομάδα (λευκή εμφάνιση) στήνουν την μπάλα και περιμένουν ενώ παίχτες της αντίπαλης ομάδας στήνονται απέναντι για να εμποδίσουν το πέρασμα της μπάλας. Πάλι έχουμε ένα θάμπωμα της εικόνας λόγω της κίνησης της κάμερας λήψης.

Στο τελευταίο video, διάρκειας 25 δευτερολέπτων, η λήψη γίνεται από μία σταθερή κάμερα παρακολούθησης εσωτερικού χώρου. Συνολικά εμφανίζονται πέντε άτομα στο προσκήνιο να περπατούν, να στέκονται για λίγο και να συνεχίζουν. Τρία από αυτά εμφανίζονται και αποχωρούν μόνα τους από το σκηνικό ενώ δύο άτομα εισέρχονται ταυτόχρονα από το ίδιο σημείο και περπατούν παράλληλα. Κανένα άτομο δεν παρέμεινε στο σκηνικό περισσότερο από 7 δευτερόλεπτα. Η μικρής διάρκειας εμφάνιση των ατόμων, η σταθερότητα της κάμερας λήψης και η μεγάλη διάρκεια του video αποτέλεσαν σημαντικούς παράγοντες για την ακριβέστατη εικόνα του παρασκηνίου που παρουσιάζεται στην Εικόνα 22 d).

4.3 Προσαρμοστικό Μοντέλο Συνδυασμού *Gaussians* για την Εξαγωγή του Παρασκηνίου

4.3.1 Γενικά

Η μέθοδος που παρουσιάζεται σε αυτήν την υποενότητα (Zincovic, Heijden [14]), χρησιμοποιεί την ευρέως διαδεδομένη τεχνική των πολλαπλών κατανομών Gauss, για την υλοποίηση ενός αλγορίθμου προσαρμοστικού παρασκηνίου, όπως έχει περιγραφεί σε προηγούμενα κεφάλαια. Βασίζεται στην προσέγγιση των Grimson και Stauffer [1] που αναλύθηκε στην παράγραφο 3.3, εισάγοντας βέβαια κάποιες σημαντικές βελτιώσεις στην απόδοση και στην ποιότητα των αποτελεσμάτων του.







4.3.2 Η προσέγγιση

Τα κοινά στοιχεία της μεθόδου αυτής με την αντίστοιχη των Grimson και Stauffer είναι πολλά. Κατασκευάζεται μια εκτίμηση του background (μοντέλο παρασκήνιου) με gaussian κατανομές, βασιζόμενο σε αρχικά δεδομένα εκπαίδευσης (training set). Για να επιτευχθεί η προσαρμοστικότητα που επιζητούμε, πρέπει το training set να ανανεώνεται με νέα δείγματα, μέσα σε μία περίοδο προσαρμογής T (adaptation time). Εδώ ισχύουν οι ίδιες [1], περιοδικά επαναλαμβανόμενες, εξισώσεις για την ανανέωση του μοντέλου του background.

Η βασική διαφορά είναι η επιλογή διαφορετικών gaussian κατανομών για κάθε pixel αντί μιας σταθερής παραμέτρου, που υποδεικνύει το πλήθος των κατανομών που θα μοντελοποιήσουν κάθε pixel. Αυτό καθίσταται εφικτό χάρη σε ένα επιπλέον πολύπλοκο Bayesian μοντέλο πιθανοτήτων που χρησιμοποιήθηκε σε συνδυασμό με τις παραμέτρους π_m που αποτελούν τα εκτιμώμενα βάρη (ποσοστό των δεδομένων που αντιστοιχούν στην m -οστή συνιστώσα του Gaussian Mixture Model).

Η πειραματική σύγκριση των δύο αλγορίθμων έδειξε σαφή υπεροχή του νέου, βελτιωμένου αλγορίθμου ως προς τον επεξεργαστικό χρόνο, όπως φαίνεται και στην Εικόνα 23, αλλά και ως προς την ποιότητα των αποτελεσμάτων. Οι γκρι τιμές στις εικόνες απεικονίζουν τον επιλεγμένο αριθμό gaussian συνιστωσών για κάθε pixel. Το μαύρο χρώμα υποδεικνύει μία gaussian συνιστώσα για ένα pixel, ενώ το άσπρο σημαίνει ότι για την μοντελοποίηση ενός pixel έχει χρησιμοποιηθεί ο μέγιστος αριθμός κατανομών (δίνεται ως παράμετρος). Δηλαδή, οι περισσότερες μεταβαλλόμενες περιοχές χρειάζονται περισσότερες συνιστώσες για να μοντελοποιηθούν (άσπρο χρώμα – foreground objects).

Η άσκοπη χρήση πολλών συνιστωσών για την μοντελοποίηση του παρασκήνιου σε περιοχές που κάτι τέτοιο δεν είναι αναγκαίο, αυξάνει τον επεξεργαστικό χρόνο του συστήματος. Για τον λόγο αυτόν, η μέθοδος αυτή επιτυγχάνει ταχύτερη επεξεργασία από ότι η μέθοδος των Grimson και Stauffer, η οποία χρησιμοποιεί τη σταθερή παράμετρο του πλήθους των gaussian συνιστωσών. Αυτό είναι ιδιαίτερα φανερό σε περιβάλλοντα όπου τα παρασκήνιο έχει μεγάλο βαθμό στασιμότητας. Σε άλλα, πιο «δυναμικά» περιβάλλοντα (π.χ. φύλλα που κινούνται από τον αέρα στο προσκήνιο), έχουμε τη δημιουργία νέων συνιστωσών κατά τη διάρκεια της μοντελοποίησης, κάτι που μειώνει την ταχύτητα του αλγορίθμου, αλλά και πάλι ο αλγόριθμος παραμένει ταχύτερος από αυτόν των Grimson και Stauffer.

Test sequences	 Traffic sequence, frame 573	 Laboratory sequence, frame 640	 Trees sequence, frame 550
GMM - the selected number of components and the average processing times	 old: 19.1 ms, improved: 13.0 ms	 old: 19.3 ms, improved: 15.9 ms	 old: 19.7 ms, improved: 19.3 ms

Εικόνα 23: Σύγκριση του αλγορίθμου των Grimson και Stauffer (old) και των Zivcovic και Heijden (improved).

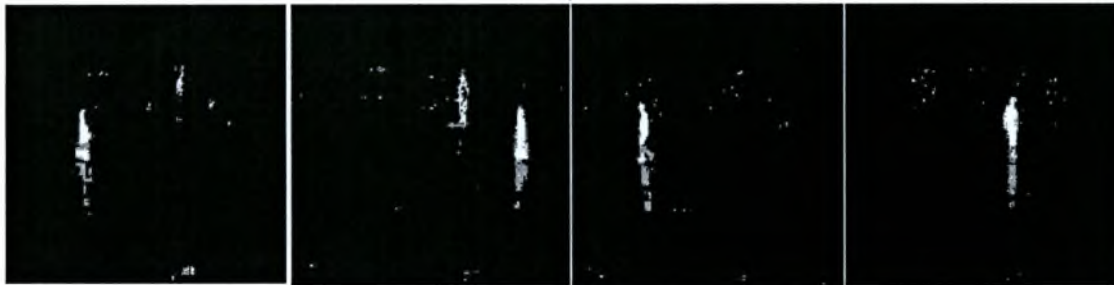
4.3.3 Υλοποίηση και Πειράματα

Οι συναρτήσεις και βιβλιοθήκες που δημιουργήθηκαν για την υλοποίηση της μεθόδου είναι γραμμένες σε κώδικα της γλώσσας προγραμματισμού C++ και η κλήση γίνεται πάνω από το περιβάλλον του Matlab, διαβάζοντας κάθε φορά ολόκληρο το αρχείο δεδομένων (video) σαν είσοδο. Τα αποτελέσματα που πήραμε από την εκτέλεση της μεθόδου ήταν πολύ ικανοποιητικά όσον αφορά το παρασκήνιο, αλλά και για την ανίχνευση αντικειμένων και σκιών στο προσκήνιο.

Στις παρακάτω εικόνες φαίνονται κάποια παραδείγματα εκτέλεσης του αλγορίθμου. Το background έχει ανιχνευθεί και αφαιρεθεί στο τελικό αποτέλεσμα, ώστε να δοθεί έμφαση στην ανίχνευση και ιχνηλάτηση κινούμενων αντικειμένων στο προσκήνιο, αλλά και στις σκιές τους.

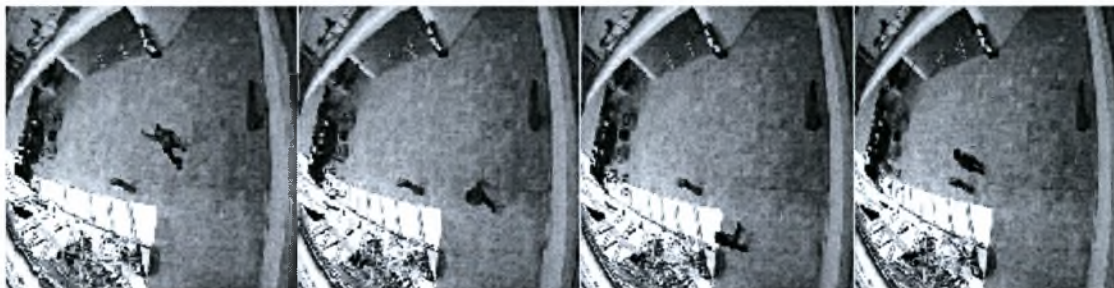


a)



b)

Εικόνα 24: Εξαγωγή του background και ανίχνευση αντικειμένων στο προσκήνιο – άνθρωποι που περπατούν έξω από ένα εμπορικό κατάστημα. a) αρχικά frames του video εισόδου b) frames του video μετά την εκτέλεση της μεθόδου



a)



b)

Εικόνα 25: Εξαγωγή του background και ανίχνευση αντικειμένων στο προσκήνιο – άνθρωποι που περπατούν σε εσωτερικό χώρο. a) αρχικά frames του video εισόδου b) frames του video μετά την εκτέλεση της μεθόδου

Στην Εικόνα 24 a) φαίνονται κάποια frames του video το οποίο δόθηκε ως είσοδος στο πρόγραμμα και στην Εικόνα 24 b) φαίνονται τα αντίστοιχα frames από

τα αποτελέσματα που έδωσε σαν έξοδο ο αλγόριθμος σε αποχρώσεις του γκρι (πλήθος gaussian κατανομών). Είναι ορατή η απομάκρυνση του background, αλλά και η υψηλής ακρίβειας ιχνηλάτηση των ανθρώπων που κινούνται μέσα και έξω από ένα εμπορικό κατάστημα. Επιπροσθέτως, έχουν καταγραφεί σε πολύ ικανοποιητικό βαθμό και οι σκιές των ανθρώπων στο δάπεδο. Κάποιες ατέλειες στο αποτέλεσμα (άσπρες κουκίδες) οφείλονται σε θόρυβο της εικόνας κατά την λήψη και σε μικρές και γρήγορες αλλαγές του φωτισμού, λόγω του αντικατοπτρισμού από το γυαλιστερό δάπεδο.

Στην Εικόνα 25 έχουμε εξίσου καλά αποτελέσματα με λίγες διαφορές. Το παράθυρο που βρίσκεται στα αριστερά της εικόνας εισάγει έντονο φωτισμό στο σκηνικό, ο οποίος «παρεμβάλλεται» συνεχώς από τους ανθρώπους που περνούν μπροστά ή πίσω από το παράθυρο. Αυτό σε συνδυασμό και με την ποιότητα λήψης της σταθερής κάμερας παρακολούθησης, δημιουργούν αυτά τα προβλήματα μοντελοποίησης που αντικατοπτρίζονται στην έξοδο του προγράμματος [Εικόνα 25 b)].

4.4 Προσέγγιση Background με Αλγορίθμους Ελαχίστης Τομής Γράφου και Μορφολογικών Τελεστών

4.4.1 Γενικά

Οι δύο μέθοδοι που θα παρουσιαστούν σε αυτήν την ενότητα διαφέρουν σε αρκετά σημεία από αυτές που έχουν παρουσιαστεί μέχρι τώρα. Πρόκειται για τον αλγόριθμο Τομής Γράφου και για τον αλγόριθμο Μορφολογικών Τελεστών [15], οι οποίοι ανιχνεύουν αντικείμενα στο προσκήνιο δεδομένου ενός στατικού παρασκηνίου. Ο υπολογιστής δημιουργεί ένα μοντέλο του στατικού παρασκηνίου είτε off-line, είτε το ανανεώνει δυναμικά έπειτα από κάθε frame στο video και μετά συγκρίνει το επόμενο frame με το μοντέλο του παρασκηνίου ανά pixel. Pixels, τα οποία διαφέρουν πολύ από το παρασκηνιακό μοντέλο, μπορεί να χαρακτηριστούν ως προσκήνιο, τουλάχιστον υπό ιδανικές συνθήκες.

Δυστυχώς, υπάρχει ένα πλήθος παραγόντων (θόρυβος κάμερας, μη στατικά στοιχεία στο παρασκήνιο, αντικατοπτρισμοί, κλπ) που εμπλέκονται στην διαδικασία, με αποτέλεσμα να μην είναι δυνατή ποτέ η τέλεια εξαγωγή του παρασκηνίου. Συνεπώς, η σύγκριση ενός pixel σε ένα frame με το μοντέλο του παρασκηνίου για το συγκεκριμένο pixel δεν μπορεί οριστικά να το κατηγοριοποιήσει ως μέρος του παρασκηνίου ή του προσκηνίου, χωρίς κάποιο ενδεχόμενο λάθους.

Λάθη σε συγκεκριμένα pixels μπορούν να μετριαστούν συναθροίζοντας τα αποτελέσματα πάνω σε μία «γειτονιά» από pixels (pixels που βρίσκονται σε απόσταση ακτίνας r από το κεντρικό pixel). Αυτή η προσέγγιση είναι διαδομένη για τον «καθαρισμό» της εικόνας από λάθη. Ένας συνδυασμός από μορφολογικούς τελεστές στην δυαδική εικόνα του αποτελέσματος απομακρύνει απομονωμένα προσκηνιακά, ή παρασκηνιακά pixels δημιουργώντας μια καλύτερη προσέγγιση στα περιγράμματα των κινούμενων αντικειμένων. Δυστυχώς, αυτή η προσέγγιση μπορεί να απαλείψει επίσης λεπτομέρειες στις άκρες των περιγραμμάτων.

Ο αλγόριθμος τομής γράφου κατασκευάζει έναν γράφο, ο οποίος ενσωματώνει όλες τις διαφορές που έχουν υπολογιστεί μεταξύ του τρέχοντος frame και του μοντέλου του παρασκηνίου. Οι σύνδεσμοι στο γράφο απεικονίζουν την συνεκτικότητα των pixels στην εικόνα, επιτρέποντας έτσι κάθε pixel να επηρεάζει αυτά που βρίσκονται στην «γειτονιά του». Ο γράφος τμηματοποιείται σε παρασκήνιο και προσκήνιο, ώστε να μπορούν να διορθωθούν μεμονωμένα λάθη, χωρίς να εισάγονται νέες καθολικές παραποιήσεις στην εικόνα.

4.4.2 Αλγόριθμος Μορφολογικών Τελεστών

Οι μορφολογικοί τελεστές είναι μια προσέγγιση για την απαλοιφή του θορύβου μετά από το background subtraction. Συγκεκριμένα, η μορφολογική επεξεργασία βασίζεται σε δύο βασικούς τελεστές. Το dilation (διαστολή) και το erosion (διάβρωση). Το dilation επεκτείνει το προσκήνιο της εικόνας, προσθέτοντας για παράδειγμα ένα pixel σε αυτό, αν οποιοδήποτε από τα pixel που ανήκουν σε μια καθορισμένη γειτονιά ακτίνας r είναι ήδη μέρος του προσκηνίου. Το erosion επεκτείνει το background, απομακρύνοντας ένα pixel από το προσκήνιο, εάν έστω και ένα pixel που ανήκει στην γειτονιά του ανήκει και στο background. Οι δύο τελεστές

μπορούν να συνδυαστούν. Ένα dilation ακολουθούμενο από ένα erosion ονομάζεται closing (κλείσιμο) και «γεμίζει κενά» στο προσκήνιο μικρότερα από την διάμετρο της γειτονιάς. Επίσης, ένα erosion ακολουθούμενο από ένα dilation ονομάζεται opening (άνοιγμα) και χρησιμοποιείται για να εξαλείψει απομονωμένα pixels του προσκηνίου.

Ο θόρυβος που παρουσιάζεται στο αποτέλεσμα της εικόνας του παρασκήνιου, πολλές φορές εμφανίζει κάποια pixels του προσκηνίου στο παρασκήνιο και αντίστροφα. Αν εκτελεσθεί closing και μετά opening τα λάθη αυτά μπορούν να ελαττωθούν σημαντικά. Προσοχή ωστόσο πρέπει να δοθεί στην επιλογή της ακτίνας της γειτονιάς. Αν είναι πολύ μικρή, τότε μεγάλες συστάδες από pixel θορύβου θα παραμείνουν ως έχουν, ενώ αν είναι πολύ μεγάλη μπορεί να χαθεί γνήσια πληροφορία του προσκηνίου.

Σε ιδιαιτέρως θορυβώδη παρασκήνια υπάρχει περίπτωση κάποια εσφαλμένως κατηγοριοποιημένα pixels του παρασκήνιου να είναι τόσα πολλά και συγκεντρωμένα σε μια περιοχή, ώστε το αρχικό closing να γεμίσει τα κενά μεταξύ τους. Το φαινόμενο αυτό εμποδίζεται αυξάνοντας το κατώφλι τ της αρχικής τμηματοποίησης background/foreground. Όταν το κατώφλι είναι υψηλό, περισσότερα pixels του προσκηνίου θα κατηγοριοποιηθούν σαν παρασκήνιο και ελάχιστα pixels του παρασκήνιου ως προσκήνιο. Εφαρμόζοντας έπειτα closing, θα γεμίσουν τα κενά μεταξύ των σωστά κατηγοριοποιημένων pixels του foreground.

4.4.3 Τομές Γράφων για την τμηματοποίηση του Προσκηνίου

Ο αλγόριθμος τομής γράφου (graph-cut) δημιουργεί αρχικά έναν γράφο βασισμένο στην εικόνα. Κάθε pixel p_{ij} στην εικόνα παράγει μια αντίστοιχη κορυφή γράφου u_{ij} . Δύο επιπλέον κορυφές σχηματίζουν την πηγή (source) και την δεξαμενή (sink), που αντιστοιχούν στο προσκήνιο και παρασκήνιο αντίστοιχα.

Η Εικόνα 26 δείχνει τον γράφο από ένα μικρό (3×3) κομμάτι εικόνας. Μια τυπική κορυφή στο γράφο συνδέεται ακριβώς με άλλες έξι κορυφές. Το source, το sink και άλλες τέσσερις κορυφές από τους γείτονές του. Κορυφές, οι οποίες αντιστοιχούν σε pixels τα οποία βρίσκονται στις άκρες της εικόνας θα έχουν λιγότερους συνδέσμους και τα source και sink συνδέονται με όλες τις κορυφές που αντιστοιχούν σε pixels της εικόνας. Τα βάρη των ακμών μεταξύ κορυφών των pixels

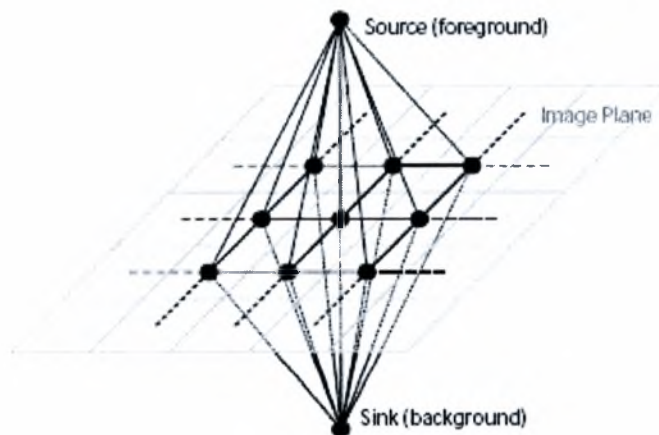
και της πηγής s , ή της δεξαμενής t , προέρχονται απ' ευθείας από την διαφορά μεταξύ του τρέχοντος frame και του μοντέλου του παρασκήνιου στο αντίστοιχο pixel, δ_{ij} :

$$w(s, p_{ij}) = \delta_{ij}$$

$$w(p_{ij}, t) = 2\tau - \delta_{ij}$$

Οι γειτονικές ακμές (μεταξύ κορυφών pixels) έχουν όλες ίδια βάρη, ίσα με τ φορές μια άλλη παράμετρο α . Η παράμετρος τ παίζει ρόλο κατωφλίου που υποδεικνύει αν το pixel σχετίζεται περισσότερο με το background ή με το foreground.

Η τιμή του α ρυθμίζει πόσο ισχυρά είναι μεταξύ τους τα γειτονικά pixels (τείνουν να είναι όλα μαζί στο προσκήνιο ή στο παρασκήνιο). Αν το α είναι μεγάλο, τότε το αποτέλεσμα θα εμφανίσει μεγάλες ομοιογενείς συστάδες. Αν τα δεδομένα εισόδου του αλγορίθμου περιέχουν θόρυβο τότε απαιτείται μεγαλύτερη τιμή στο α , έτσι ώστε να εξομαλυνθούν τα μεγάλα clusters θορύβου.




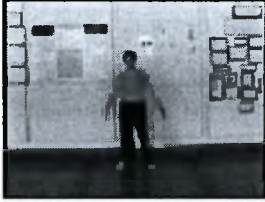
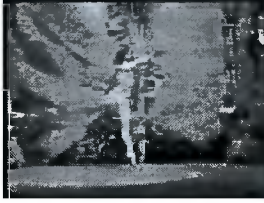
Εικόνα 26: Περιοχή γράφου 3×3. Όλα τα pixels καθώς και τα source, sink αντιστοιχούν σε κόμβους του γράφου [15].

Όταν ο γράφος κατασκευαστεί, υπάρχουν κάποιες κλασικές μέθοδοι βασισμένες στην ροή του γράφου, οι οποίες διαχωρίζουν το sink από το source (το προσκήνιο από το παρασκήνιο). Κάθε κόμβος στο γράφο μετά την βέλτιστη τομή θα συνδέονται μόνο με το source ή με το sink.

4.4.4 Πειράματα και Αποτελέσματα

Το κύριο μέρος των πειραμάτων εξετάζει την απόδοση των δύο αλγορίθμων με είσοδο, video υπό πραγματικές συνθήκες. Τα τρία videos που χρησιμοποιήθηκαν διαφέρουν σε ποιότητα εικόνας. Από έγχρωμα videos σε εσωτερικούς και εξωτερικούς χώρους με σταθερές κάμερες λήψεις μέχρι ασπρόμαυρα videos με ελαφρώς κινούμενη κάμερα. Ο Πίνακας 2 δείχνει περισσότερες λεπτομέρειες για τα videos εισόδου.

Για την εξαγωγή background από ένα video πρέπει να κατασκευαστεί ένα background model χρησιμοποιώντας στατιστικές τεχνικές. Δεν προτιμήθηκαν πολύπλοκα μοντέλα όπως το Mixture of Gaussians, για να μην δημιουργηθεί σύγχυση με την σύγκριση των αποτελεσμάτων.

	Clip #2: Outdoor clip, regular motion, some reflection off glass. Fixed camera. 124 frames.
	Clip #1: Indoor clip with some shadowing, reflections off the floor. Some low contrast portions. Fixed camera. 160 frames.
	Clip #3: Grayscale MPEG video of a dancer. Low contrast. Panning camera. 99 frames.

Πίνακας 2: Λεπτομέρειες για τα videos που χρησιμοποιήθηκαν [15].

Ο Πίνακας 3 δείχνει τους ρυθμούς σφαλμάτων (error rates) του μορφολογικού και του αλγορίθμου τομής γράφου, με δεδομένα εισόδου τα παραπάνω videos, έπειτα από κατάλληλη ρύθμιση των παραμέτρων τους. Εξετάζοντας τα αποτελέσματα μπορεί να δει κάποιος, ότι η ποιότητα του video εισόδου καθορίζει σε μεγάλο βαθμό το αποτέλεσμα των αλγορίθμων. Γι' αυτόν τον λόγο, το τρίτο video έχει τα περισσότερα σφάλματα. Ωστόσο, ο βασιζόμενος στον γράφο αλγόριθμο αποδίδει καλύτερα σε κάθε video που δοκιμάστηκε.

Clip	Morph	Graph
1. Outdoor	0.164	0.161
<i>Params:</i>	$\tau = 20.3, r = 2$	$\tau = 16.2, \alpha = 0.94$
2. Indoor	0.154	0.133
<i>Params:</i>	$\tau = 5.21, r = 1$	$\tau = 4.87, \alpha = 0.81$
3. Dancer	0.541	0.532
<i>Params:</i>	$\tau = 2.26, r = 1$	$\tau = 2.15, \alpha = 0.97$

Πίνακας 3: Σύνοψη των αποτελεσμάτων εξαγωγής του προσκηνίου [15].

Συγκρίνοντας τους αλγορίθμους στην Εικόνα 27, ο αλγόριθμος γράφου (δεξιά στήλη), εμφανίζει πιο «καθαρά» αποτελέσματα από το αποτέλεσμα της εφαρμογής μορφολογικών τελεστών και εμφανίζει πιο πιστά το περίγραμμα των κινούμενων ανθρώπων.

Το πρώτο video (εσωτερικού χώρου), δείχνει καλύτερα τα πλεονεκτήματα του αλγορίθμου τομής γράφου, εμφανίζοντας πιο ομαλά περιγράμματα και λιγότερο συχνή προσμέτρηση του παρασκηνίου. Αν και ο αλγόριθμος αυτός τα πηγαίνει καλύτερα και στα τρία είδη videos, στο δεύτερο (εξωτερικός χώρος), είναι πιο εύκολο να φανούν οι διαφορές στο αποτέλεσμα ενώ στο τρίτο αρκετά δύσκολο. Αυτό οφείλεται στην αντίθεση των χρωμάτων που υπάρχει μεταξύ του παρασκηνίου και του προσκηνίου. Τα εμφανή λάθη στο τρίτο video (Dancer) δείχνουν ότι οποιοσδήποτε αλγόριθμος που βασίζεται στην αντίθεση (contrast-based) θα αποτύχει σε παραπλανητικές εισόδους (χαμηλής αντίθεσης μεταξύ παρασκηνίου και προσκηνίου, «βαριά» σκιά, αντανάκλασεις, θόρυβος, κλπ). Η υψηλή ποιότητα του video εισόδου παραμένει κρίσιμη παράμετρος, ανεξάρτητα ποιος αλγόριθμος πρόκειται να εφαρμοστεί.

Στην Εικόνα 28 εκτελείται ο αλγόριθμος graph-cut με είσοδο ένα video, στο οποίο ένας άνθρωπος περπατά σε έναν εξωτερικό χώρο με στατικό παρασκήνιο. Η παράμετρος `significanceThreshold` στην υλοποίηση, είναι η συσχέτιση ενός pixel με το προσκήνιο (ανάλογο του τ). Αν για παράδειγμα, θεωρούμε ότι κομμάτια του προσκηνίου δεν έχουν προσμετρηθεί σε αυτό, τότε η τιμή του `significanceThreshold` πρέπει να μειωθεί, ενώ εάν στην ιχνηλάτηση του αντικειμένου στο προσκήνιο συμπεριλαμβάνεται και κομμάτι του παρασκηνίου, τότε η τιμή του πρέπει να αυξηθεί. Στο συγκεκριμένο video παρατηρούμε, ότι παίρνουμε πιο ακριβή αποτελέσματα με την μείωση του κατωφλίου (δεύτερη σειρά) και αυτό οφείλεται

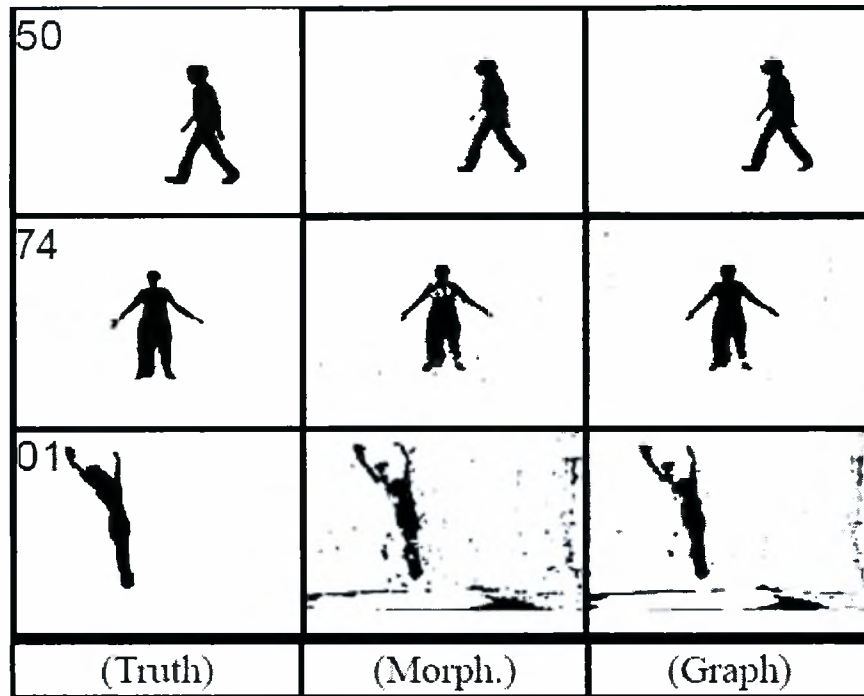
στην αργή κίνηση του αντικειμένου στο προσκήνιο. Η αργή κίνηση έχει σαν αποτέλεσμα την κατηγοριοποίηση μερικών pixels του αντικειμένου ως background.

4.4.5 Συμπεράσματα

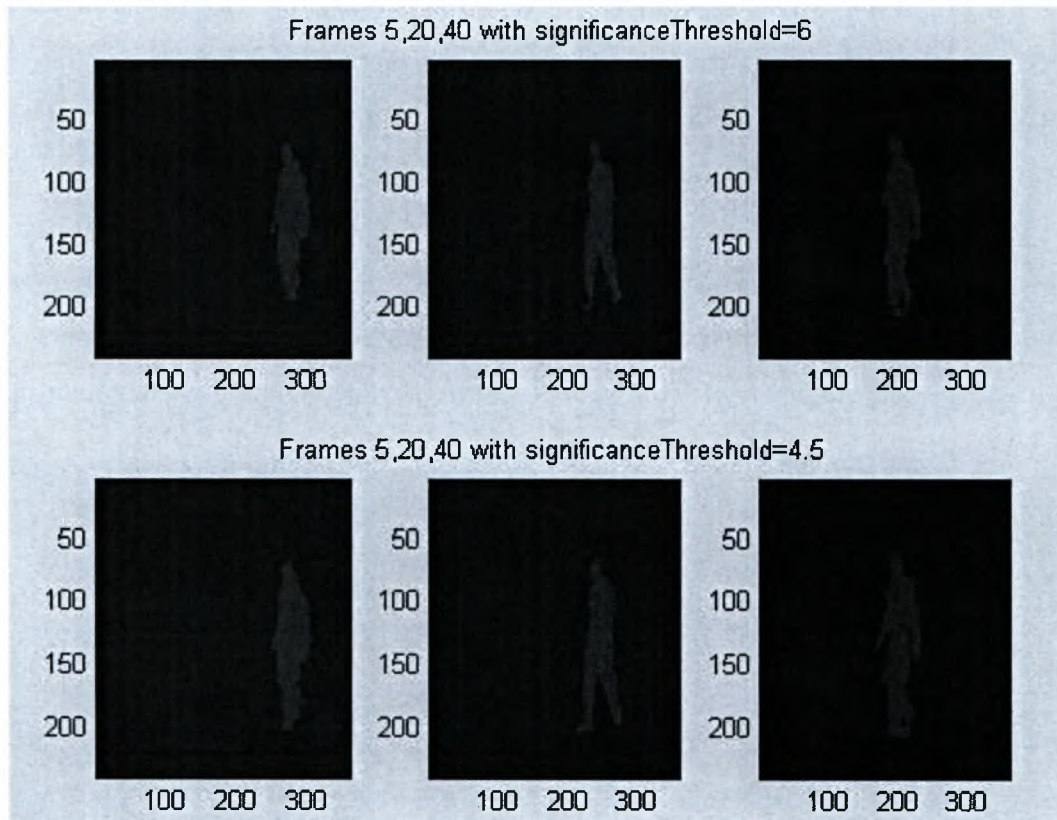
Οι τομές γράφου μπορούν να παράγουν μια «καθαρότερη» εξαγωγή foreground βασισμένο στην ανά frame σύγκριση με ένα μοντέλο παρασκηνίου. Και οι δύο αλγόριθμοι που παρουσιάστηκαν συνδυάζουν κάποιες ερευνητικά δοκιμασμένες τεχνικές, που δεν είχαν ποτέ εφαρμοστεί σε προβλήματα εξαγωγής προσκηνίου. Τα πειραματικά αποτελέσματα έδειξαν, ότι οι τεχνικές που «δανείστηκαν» παράγουν άριστα αποτελέσματα όπως και σε άλλες εφαρμογές.

Οι τομές γράφου, επίσης, παράγουν πολύ καλύτερα αποτελέσματα από αλγόριθμους οι οποίοι βασίζονται σε μορφολογικούς τελεστές. Χειρίζονται καλύτερα τον θόρυβο συναθροίζοντας την πληροφορία που βρίσκεται σε τοπικές «γειτονιές» γύρω από ένα pixel. Σε πειράματα με πραγματικά δεδομένα, η μέθοδος τομής γράφου μειώνει σημαντικά τα σφάλματα αν και δεν μπορεί να εξαλείψει τον θόρυβο που προέρχεται από κακής ποιότητας video εισόδου. Το βασικό μειονέκτημα της μεθόδου αυτής είναι η ταχύτητα του. Εμπειρικά μπορούμε να πούμε, ότι εκτελείται πιο αργά από ότι ο αλγόριθμος με τους μορφολογικούς τελεστές.

Η υιοθέτηση της νέας αυτής τεχνικής δίνει την εντύπωση ότι μπορεί να προσφέρει σημαντικά οφέλη σε πολλές περιπτώσεις, αν αναλογιστούμε την πληθώρα των εφαρμογών που χρησιμοποιούν εξαγωγή του παρασκηνίου. Σε κάθε περίπτωση, οι τομές γράφων για την ιχνηλάτηση του προσκηνίου αξίζουν μία θέση στα πεδία των ερευνητών.



Εικόνα 27: Δείγμα αποτελεσμάτων των δύο αλγορίθμων για κάποιο συγκεκριμένο frame, για τα τρία διαφορετικά videos εισόδου [15].



Εικόνα 28: Εκτέλεση του αλγορίθμου γράφου τομής με το ίδιο video εισόδου, αλλά για διαφορετικό significance threshold.

Κεφάλαιο 5 – Συμπεράσματα



5.1 Παρατηρήσεις και Συμπεράσματα

Τα αποτελέσματα που προκύπτουν έπειτα από διεξοδική διερεύνηση μιας πληθώρας αλγορίθμων και τεχνικών είναι τα αναμενόμενα. Η αναζήτηση της ποιότητας των αποτελεσμάτων σημαίνει πάντα με κάποιο κόστος. Το κόστος αυτό μπορεί να καταμετρηθεί σε υπολογιστικούς πόρους, σε δυσκολία υλοποίησης του πειράματος ή και του αλγορίθμου, σε χρηματικό κόστος (κάμερες λήψης υψηλής ποιότητας) αλλά και σε χρόνο (χρονική διάρκεια διεξαγωγής των πειραμάτων και επεξεργασίας των αποτελεσμάτων).

Παρουσιάστηκαν διάφοροι μέθοδοι, ως προς την ταχύτητα εκτέλεσης, την ποιότητα των αποτελεσμάτων τους και την εφαρμοσιμότητα τους σε διάφορες περιπτώσεις. Αλγόριθμοι απλοί και εύκολα υλοποιήσιμοι, μέχρι πιο πολύπλοκα μαθηματικά μοντέλα και ευριστικούς μηχανισμούς. Αναλύθηκαν αλγόριθμοι μοντελοποίησης και εξαγωγής του background και αλγόριθμοι ιχνηλάτησης foreground αντικειμένων ως άμεσο επακόλουθο των πρώτων.

Σίγουρα δεν είναι δυνατή η ακριβής σύγκριση όλων των αλγορίθμων μεταξύ τους, λόγω της ανομοιογένειας των απαιτήσεων για τις οποίες κατασκευάστηκε ο καθένας και δεδομένων των παρεχόμενων υπολογιστικών πόρων που απαιτήθηκαν. Γενικά όμως, μπορούμε να αναφέρουμε παρατηρώντας μόνο την ποιότητα των αποτελεσμάτων, ότι η μέθοδος των Mixtures of Gaussians παρουσιάστηκε ιδιαίτερα σθεναρή υπό οποιεσδήποτε συνθήκες (εσωτερικών και εξωτερικών χώρων) και κατάφερε να προσαρμόξει το background model στις διάφορες αλλαγές του παρασκηνίου διατηρώντας συνεχώς μια αρκετά πιστή προσέγγιση του. Κατάφερε να δώσει τα καλύτερα αποτελέσματα μοντελοποίησης του παρασκηνίου και αυτό συνεπάγεται άμεσα την βέλτιστη ιχνηλάτηση των αντικειμένων στο προσκήνιο (με

οποιαδήποτε μέθοδο). Το κόστος σε υπολογιστικούς πόρους όμως παραμένει αρκετά υψηλό.

5.2 Μελλοντική Έρευνα

Το πεδίο του “*Adaptive Color Modelling for Background Detection and Tracking*” , αλλά και της *Ψηφιακής Επεξεργασίας Εικόνας και Video* γενικότερα, είναι ένας τομέας της πληροφορικής «ανοιχτός» σε νέες ιδέες, εφαρμογές αλλά και βελτιώσεις των ήδη υπαρχόντων μεθόδων.

Ο βασικός περιορισμός που υπήρχε πάντα στην ψηφιακή επεξεργασία εικόνας και video ήταν οι διαθέσιμοι υπολογιστικοί πόροι. Η διερεύνηση του τομέα έχει αρχίσει από την δεκαετία του '60 αλλά το κόστος της υλοποίησης και λειτουργίας των εφαρμογών τότε, ήταν απαγορευτικό. Η ραγδαία ανάπτυξη των υπολογιστικών συστημάτων είχε σαν αποτέλεσμα, οι υπολογιστικοί φραγμοί που είχαν τεθεί να ξεπερνιούνται ολοένα με την πάροδο των χρόνων. Νέες εφαρμογές έρχονταν στο προσκήνιο, αφού οι απαιτήσεις γίνονταν ολοένα και περισσότερες και νέες ανάγκες δημιουργούνταν για την ποιοτικότερη επεξεργασία εικόνας και την παραγωγή πολύτιμων αποτελεσμάτων.

Η ανάγκη επεξεργασίας real-time δεδομένων ήταν ένας στόχος, που επιτεύχθηκε σε μεγάλο βαθμό, ικανοποιητικά. Η εξέλιξη της βιντεοκάμερας και των υπολογιστικών συστημάτων έχουν καταφέρει αυτή τη στιγμή να υπάρχει μια πληθώρα εφαρμογών της ψηφιακής επεξεργασίας video, η οποία εξελίσσεται συνεχώς.

Είναι κάτι παραπάνω από σίγουρο ότι οι καινοτομίες και οι βελτιώσεις των μεθόδων που περιγράφηκαν σε αυτή την εργασία δεν θα σταματήσουν. Οι έρευνες που γίνονται στον χώρο περιορίζονται αλλά και συμβαδίζουν με την εξέλιξη της τεχνολογίας. Η μέθοδος Mixtures of Gaussians είναι μια απαιτητική σε υπολογιστικούς πόρους μέθοδος αλλά και πολλά υποσχόμενη. Η αύξηση, έστω και κατά ένα, στις gaussian συνιστώσες που χρησιμοποιούνται στην μοντελοποίηση του παρασκηνίου μπορεί να είναι πολύ αποτελεσματική, αλλά το υπολογιστικό κόστος αυξάνεται εκθετικά. Επίσης, υπάρχουν εφαρμογές (κάμερες παρακολούθησης) που

ίσως απαιτούν την επεξεργασία real-time δεδομένων, κάτι που επιβαρύνει κατά πολύ ένα σύστημα.

Παρόλα αυτά, οι έρευνες που γίνονται είναι πολύ αισιόδοξες και τα αποτελέσματα που παράγονται από αυτές πολύ αξιόπιστα. Είναι βέβαιο ότι ο τομέας κερδίζει συνεχώς την προσοχή όλο και περισσότερων ερευνητών και κατά συνέπεια το πεδίο του *“Adaptive Color Modelling for Background Detection and Tracking”* έχει ακόμα να δείξει αρκετά σε όλους μας.

- [1] Chris Stauffer, W.E.L. Grimson, “*Adaptive background mixture models for real-time tracking*”, The Artificial Intelligence Laboratory Massachusetts Institute of Technology, Cambridge, 1999.
- [2] P. KaewTraKulPong, R. Bowden, “*An Improved Adaptive Background Mixture Model for Real-time Tracking with Shadow Detection*”, 2nd European Workshop on Advanced Video Based Surveillance Systems, AVBS01, UK, 2001
- [3] M Piccardi, T. Jan, “*Mean-Shift Background Image Modelling*”, International Conference on Image Processing (ICIP), 2004.
- [4] Darren Butler, Sridha Sridharan, V. Michael Bove, Jr., “*Real-Time Adaptive Background Segmentation*”, USA, 2004.
- [5] Liyuan Li, Weimin Huang, Irene Y.H. Gu, Qi Tian, “*Foreground Object Detection in Changing Background Based on Color Co-Occurrence Statistics*”, Sixth IEEE Workshop on Applications of Computer Vision, 2002.
- [6] Aristodemos Pnevmatikakis, Lazaros Polymenakos, “*2D Person Tracking Using Kalman Filtering and Adaptive Background Learning in a Feedback Loop*”, Athens Information Technology, Greece, 2006
- [7] Aristodemos Pnevmatikakis, Lazaros Polymenakos, “*Kalman Tracking with Target Feedback on Adaptive Background Learning*”, Athens Information Technology, Greece, 2006.
- [8] Mario A.T. Figueiredo, “*Lecture Notes on the EM Algorithm*”, Lisbon, Portugal, 2004.

- [9] Jeff A. Bilmes, “*A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models*”, International Computer Science Institute, Berkeley, USA, 1998.
- [10] Frank Dellaert, “*The Expectation Maximization Algorithm*”, Georgia Institute of Technology, 2002.
- [11] Alexandre R.J. Francois, Gerard G. Medioni, “*Adaptive Color Background Modelling for Real-Time Segmentation of Video Streams*”, Los Angeles, USA, 1999.
- [12] Chengcui Zhang, Shu-Ching Chen, Mei-Ling Shyu, Srinivas Peeta, “*Adaptive Background Learning for Vehicle Detection and Spatio-Temporal Tracking*”, ICICS-PCM, Singapore, 2003.
- [13] Sean Borman, “The Expectation Maximization Algorithm, A Short Tutorial”, 2004.
- [14] Zoran Zivkovic, Ferdinand van der Heijden, “*Efficient Adaptive Density Estimation per Image Pixel for the Task of Background Subtraction*”, Netherlands, PRL 2006.
- [15] Nicholas R. Howe, Alexandra Deschamps, “*Better Foreground Segmentation Through Graph Cuts*”, 1995.
- [16] Expectation-Maximization Algorithm.
http://en.wikipedia.org/wiki/Expectation-maximization_algorithm
- [17] HSV Color Space. http://en.wikipedia.org/wiki/HSV_color_space
- [18] YUV. <http://en.wikipedia.org/wiki/YUV>

[19] Cluster Analysis. http://en.wikipedia.org/wiki/Cluster_analysis

[20] Median Filter http://en.wikipedia.org/wiki/Median_filter

[21] Median Filter <http://www.cee.hw.ac.uk/hipr/html/median.html>



ΠΑΝΕΠΙΣΤΗΜΙΟ
ΘΕΣΣΑΛΙΑΣ



004000091256