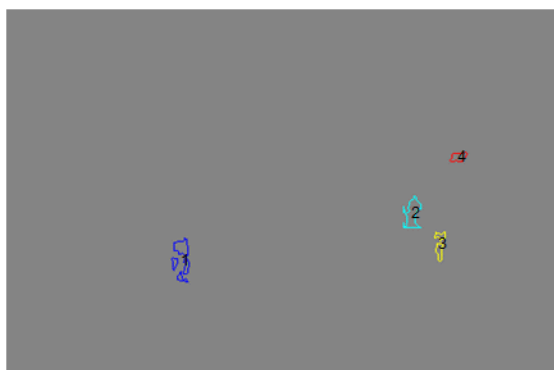




**Πολυτεχνείο Κρήτης Τμήμα Ηλεκτρονικών Μηχανικών και
Μηχανικών Υπολογιστών**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

«ΕΝΤΟΠΙΣΜΟΣ ΑΝΤΙΚΕΙΜΕΝΩΝ ΑΠΟ ΠΟΛΛΑΠΛΕΣ ΚΑΜΕΡΕΣ»



Φοιτήτρια: Αναστασία Πεντάρη

ΑΜ: 2005030009

Επιτροπή: Ζερβάκης Μιχάλης (επιβλέπων καθηγητής)

Πετράκης Ευριπίδης

Μανιά Κατερίνα

Περιεχόμενα

Περίληψη	3
----------------	---

Κεφάλαιο 1: Εισαγωγή

1. Εισαγωγή	4
1.1 Βασικές έννοιες	7
1.2 Εφαρμογές της ανίχνευσης και αναγνώρισης του εντοπισμού κίνησης.....	8
1.3 Ανίχνευση της κίνησης	9
1.3.1 Κατάτμηση της κίνησης	10
1.4 Προβλήματα κατά την ανίχνευση των αντικειμένων	11
1.5 Στατική κάμερα και Δυναμική κάμερα.....	12

Κεφάλαιο 2: Αφαίρεση Φόντου

2. Εισαγωγή	14
2.1 Μείωση θορύβου και φωτισμού	15
2.2 Παρουσίαση μεθόδων αφαίρεσης φόντου.....	17
2.3 Σύγκριση αποτελεσμάτων	18

Κεφάλαιο 3: Εκτίμηση κίνησης

3. Εισαγωγή	24
3.1 Εκτίμηση κίνησης	24
3.2 Υπολογισμός του διανύσματος κίνησης μέσω του Block Matching Algorithm (BMA).....	25
3.3 Χρωματισμός των διανυσμάτων κίνησης βάσει του χρωματικού	

κύκλου	32
3.4 Χρωματικός κύκλος για ομαδοποίηση των διανυσμάτων κίνησης ..	34
Κεφάλαιο 4: Πολυεπίπεδος αλγόριθμος εντοπισμού κίνησης	
4. Εισαγωγή	42
4.1 Διαχωρισμένες περιοχές	43
4.1.1 Επίλυση προβλήματος του χρώματος	49
4.2 Αλγόριθμος εντοπισμού άμορφης μάζας	52
4.2.1 Υπολογισμός πρόβλεψης περιοχής	54
4.3 Αλγόριθμος εντοπισμού αντικειμένων	59
4.4 Εκτεταμένα αποτελέσματα	69
Κεφάλαιο 5: Εντοπισμός αντικειμένων από πολλαπλές κάμερες	
5. Εισαγωγή	78
5.1 Hidden Markov Μοντέλο.....	79
Κεφάλαιο 6: Μελλοντική εργασία	
6. Εισαγωγή	81
6.1 Συμπεράσματα για μελλοντική εργασία	81
ΒΙΒΛΙΟΓΡΑΦΙΑ	83

Περίληψη

Σε αυτή τη διπλωματική εργασία γίνεται μελέτη και ανάλυση του εντοπισμού των αντικειμένων από μία, αρχικά, κι έπειτα από πολλαπλές κάμερες. Στο κεφάλαιο 1 παρουσιάζεται το θεωρητικό υπόβαθρο, περιγράφονται εν συντομία τα μέρη της ανάλυσης μίας ολοκληρωμένης διαδικασίας και ορίζονται οι έννοιες που θα χρησιμοποιηθούν στην συνέχεια. Στο κεφάλαιο 2 παρουσιάζονται οι μέθοδοι που υλοποιήθηκαν για την αφαίρεση του φόντου, μίας βασικής διαδικασίας κατά το στάδιο της προεπεξεργασίας, και γίνεται σύγκριση των αποτελεσμάτων. Επίσης παρουσιάζεται η μέθοδος μείωσης του φωτισμού και του θορύβου. Στο κεφάλαιο 3, παρουσιάζονται οι τεχνικές κατάτμησης κίνησης δηλαδή οι διαδικασίες-αλγόριθμοι που χρησιμοποιήθηκαν για την εξαγωγή της κίνησης. Στο κεφάλαιο 4, αναλύονται οι μέθοδοι που υλοποιήθηκαν έτσι ώστε να γίνει εξαγωγή περιοχών. Τέλος, στο κεφάλαιο 5 γίνεται αναφορά για μελλοντική εργασία και αναλύονται τα προβλήματα που συναντήθηκαν στον εντοπισμό των αντικειμένων από πολλαπλές κάμερες έτσι ώστε η εργασία μας να φτάσει σε άτοπο.

ΚΕΦΑΛΑΙΟ 1: ΕΙΣΑΓΩΓΗ

Η αναγνώριση και η παρατήρηση ενός κινούμενου αντικειμένου αποτελεί ένα σημαντικό κεφάλαιο στο πεδίο της υπολογιστικής όρασης. Η ολοένα και αυξανόμενη ανάγκη για αυτοματοποιημένη ανάλυση βίντεο έχει κεντρίσει το ενδιαφέρον της επιστημονικής κοινότητας για ανάπτυξη αλγορίθμων που θα προσομοιώνουν την ανθρώπινη αντίληψη αναγνώρισης της κίνησης. Η προσομοίωση αυτή δεν είναι ακόμη τόσο ώριμη, βέβαια.

Αν και οι υπολογιστές έχουν σημαντικά αποτελέσματα όσον αφορά την ταχύτητα της επεξεργασίας πολύπλοκων δεδομένων, την αποθήκευση και γρήγορη ανάκτηση μεγάλου όγκου πληροφοριών, αδυνατούν να προσεγγίσουν τα ποιοτικά χαρακτηριστικά που ακολουθεί ο ανθρώπινος εγκέφαλος για να εκτελέσει τη διαδικασία της παρακολούθησης και αναγνώρισης της τροχιάς ενός αντικειμένου.

Στην περίπτωση των υπολογιστικών συστημάτων η αίσθηση της όρασης εξομοιώνεται καταγράφοντας και αναλύοντας ακολουθίες εικόνων μέσω μίας ή περισσότερων καμερών. Η χρήση πολλαπλών καμερών σε αντίθεση με τη μία στατική κάμερα μας προσφέρει βάθος πεδίου, αύξηση οπτικού πεδίου, καλύτερη και ακριβέστερη προσέγγιση της κίνησης. Όπως όμως αναφέρθηκε και παραπάνω, λόγω της διαφοράς της αντίληψης του ανθρώπινου εγκεφάλου με το υπολογιστικό σύστημα, ένα υπολογιστικό σύστημα θα πρέπει να εκτελέσει μία σειρά από σύνθετες λειτουργίες για την πραγματοποίηση του εντοπισμού και της ταυτοποίησης των αντικειμένων από πολλαπλές κάμερες. Άρα ένα πολυσύνθετο σύστημα καμερών εντοπισμού αντικειμένων μπορεί θεωρητικά να έχει κάποια σημαντικά θετικά αποτελέσματα παράλληλα όμως, πρακτικά, υπάρχουν δυσκολίες στην υλοποίησή του.

Ένα υπολογιστικό σύστημα για να πάρει μία απόφαση θα πρέπει να εξάγει κάποια χαρακτηριστικά από τις εικόνες που δέχεται βάσει των οποίων θα καταφέρει να

διακρίνει, να αναγνωρίσει και να χαρακτηρίσει αυτά τα οποία βλέπει. Στην προσπάθεια λοιπόν να μπορέσει ένα σύστημα να κατανοήσει αυτά που καταγράφει μέσω της μηχανικής όρασης, αρχικά δέχεται τα δεδομένα εισόδου, τα οποία πριν το στάδιο της επεξεργασίας δεν είναι τίποτε παραπάνω από συντεταγμένες των εικονοστοιχείων των εικόνων στο χώρο των χρωμάτων. Η μεθοδολογία που ακολουθείται για την επίλυση τέτοιων προβλημάτων είναι υψηλού επιπέδου εργασίες οι οποίες εξαρτώνται από ένα σύνολο παραμέτρων που σχετίζονται με τα επιθυμητά αποτελέσματα εξόδου.

Είναι δύσκολο να συγκρίνουμε τις εργασίες αυτές ποιοτικά καθώς, για παράδειγμα, η διαδικασία της προεπεξεργασίας των εικόνων για μετέπειτα χρήση διαφέρει από τη διαδικασία εξαγωγής της μετατόπισης των αντικειμένων από τη μία εικόνα στην επόμενη. Επίσης διαφορετική διαδικασία ακολουθείται από το υπολογιστικό σύστημα για την ενοποίηση της χρήσιμης πληροφορίας για τον καθορισμό ενός αντικειμένου βάσει κάποιων χαρακτηριστικών (features). Τέλος, άλλη διαδικασία χρειάζεται να πραγματοποιηθεί για τον εντοπισμό και την ταυτοποίηση της τροχιάς κάθε αντικειμένου.

Πολλοί άνθρωποι έχουν ασχοληθεί με τον εντοπισμό των αντικειμένων. Άλλοι έχουν ασχοληθεί με τον εντοπισμό και την ταυτοποίηση των προσώπων [6], άλλοι με την παρακολούθηση της τροχιάς των αντικειμένων και τον εντοπισμό της ανθρώπινης κίνησης [9], η δική μας εργασία βασίστηκε σε μία εργασία για τον εντοπισμό της ανθρώπινης κίνησης, σε μία ακολουθία ταινίας συμπεριλαμβανομένης της κίνησης της κάμερας [2]. Σε αυτήν την εργασία σκοπός ήταν ο εντοπισμός της ανθρώπινης κίνησης μέσω ενός bounding box, δηλαδή ενός κουτιού ουσιαστικά που περιελάμβανε κάθε αντικείμενο σε κάθε καρέ. Χωρίζεται σε δύο μέρη, στην κατάτμηση της κίνησης και στον εντοπισμό των αντικειμένων. Οι μέθοδοι που χρησιμοποιήθηκαν ταιριάζουν με τις δικές μας μεθόδους [2], μέθοδοι που αναλύονται σε αυτή τη διπλωματική εργασία.

Η εργασία αυτή, ακολουθεί τα παρακάτω τέσσερα μέρη. (α) Το αρχικό στάδιο που θα αναλυθεί είναι το στάδιο της προεπεξεργασίας κάθε εικόνας κάθε δείγματος βίντεο που είχαμε να επεξεργαστούμε. Κατά το στάδιο αυτό κάθε τρέχουσα εικόνα

χρειάζεται μια επεξεργασία έτσι ώστε να μπορεί να χρησιμοποιηθεί από το υπολογιστικό σύστημα. (β) Κατά το επόμενο στάδιο θα αναλυθεί η διαδικασία υπολογισμού του motion vector, ενός διανύσματος που περιλαμβάνει τις συντεταγμένες κάθε εικονοστοιχείου (pixel) από τη μία εικόνα στην επόμενη και δείχνει ουσιαστικά τη μετατόπιση που έχει κάθε εικονοστοιχείο. (γ) Στο προτελευταίο στάδιο γίνεται εξαγωγή της χρήσιμης πληροφορίας η οποία συντάσσει ένα αντικείμενο βάσει κάποιων γνωρισμάτων (features), όπως είναι το χρώμα, η περίμετρος κ.α. (δ) Η τελευταία και ουσιαστικότερη μεθοδολογία που ακολουθήθηκε είναι η ταυτοποίηση κάθε αντικειμένου μέσα σε μία βίντεο ακολουθία.

Σε αυτήν την εργασία εργαστήκαμε πάνω σε ήδη υπάρχοντες αλγορίθμους υψηλού επιπέδου. Η καινοτομία της εργασίας αυτής έγκειται στο γεγονός ότι καταφέραμε να προσπεράσουμε κάποιες δυσκολίες που συναντήσαμε κατά την υλοποίηση των διαδικασιών που αναφέρθηκαν παραπάνω. Οι δυσκολίες αυτές χωρίζονται σε δυσκολίες χαμηλού και υψηλού επιπέδου. Χαμηλού επιπέδου δυσκολίες θεωρούνται εκείνες που ήταν σχετικά εύκολη η μέθοδος επίλυσής τους, όπως για παράδειγμα η μείωση του θορύβου και φωτισμού κατά το στάδιο της προεπεξεργασίας. Δυσκολίες υψηλού επιπέδου θεωρούνται εκείνες που για να επιλυθούν χρειαζόνταν εξίσου υψηλού επιπέδου διαδικασίες που αύξησαν την υπολογιστική πολυπλοκότητα της συνολικής εργασίας. Τέτοιου είδους δυσκολίες ήταν η πραγματοποίηση της ομοιομορφίας των διευθύνσεων κατά το στάδιο του υπολογισμού της κατεύθυνσης κάθε άμορφης μάζας κατά το τρίτο στάδιο της εργασίας ή ο εντοπισμός σταθερών αντικειμένων κατά το στάδιο του εντοπισμού και της ταυτοποίησης των αντικειμένων καθώς και άλλα που αναλύονται κατά το πέρας της ανάλυσης της εργασίας αυτής. Πέρα όμως από τις δυσκολίες που έχουμε ξεπεράσει, έχουμε καταφέρει να μειώσουμε και την πολυπλοκότητα ενός τέτοιου μεγάλου προβλήματος, που είναι ο εντοπισμός των αντικειμένων.

Στις επόμενες ενότητες παρουσιάζονται βασικές έννοιες της σχετικής επεξεργασίας στη δική μας μελέτη καθώς και στη βιβλιογραφία. Στο κεφάλαιο 2

αναλύονται οι μέθοδοι αφαίρεσης φόντου. Στο επόμενο κεφάλαιο, 3, παρουσιάζουμε τις μεθόδους εκτίμησης της κίνησης. Στο κεφάλαιο 4 αναπτύσσονται οι μέθοδοι που χρειάστηκαν για τον εντοπισμό των αντικειμένων και τέλος, στο κεφάλαιο 5 αναλύονται τα προβλήματα που παρουσιάστηκαν κατά τον εντοπισμό των αντικειμένων από πολλαπλές κάμερες και παρουσιάζεται η μελλοντική εργασία που μπορεί να γίνει.

1.1 Βασικές Έννοιες

Στο πρώτο στάδιο της ανάλυσης της εργασίας αυτής σκοπός είναι να γίνει αναφορά σε κάποιες βασικές έννοιες παραμέτρων που χρησιμοποιήθηκαν. Αρχικά με τον όρο βασικές έννοιες εννοούμε κάποιους βασικούς όρους που πρώτον θα αναφέρονται συχνά και δεύτερον βοηθούν τη καλύτερη κατανόηση της ανάλυσης.

- *Frame sequence*: Η διαδοχική τομή ενός βίντεο σε πολλαπλές εικόνες.
- *Frame*: Μία εικόνα frame sequence.
- *Image Preprocessing*: Διαδικασίες προεπεξεργασίας για μείωση θορύβου, φωτισμού και μεταφοράς εικόνων σε κατάλληλη μορφή επεξεργασίας από το υπολογιστικό σύστημα.
- *Motion (κίνηση)*: Αποτελεί το πιο βασικό γνώρισμα και εξάγεται από δύο καρέ.
- *Motion Vector*: Συντεταγμένες που καθορίζουν τη θέση ενός pixel ή μιας ομάδας από pixels.

- *Segmented Region*: Μία ομάδα από pixels που καθορίζουν ένα κομμάτι ενός αντικειμένου.
- *Features*: Γνωρίσματα των segmented regions για ομαδοποίησή τους.
- *Περίμετρος*: Ο αριθμός των εικονοστοιχείων (pixels) που καθορίζουν μία άμορφη μάζα (blob). Ανήκει στα χαρακτηριστικά (features).
- *Χρώμα*: Αποτελεί την πιο άμεσα εξαγόμενη πληροφορία μίας εικόνας. Το χρώμα μοντελοποιείται συνήθως από τα βασικά χρώμα (RGB) ή μέσω του χρωματικού κύκλου. Ανήκει στα χαρακτηριστικά.
- *Blob*: Μία ομάδα από segmented regions που έχουν κοινά χαρακτηριστικά και καθορίζουν μέρος ενός αντικειμένου ή και ολόκληρο αντικείμενο.
- *Object*: Πολλά blobs που καθορίζουν ένα ολόκληρο, σε ιδανικές συνθήκες, αντικείμενο.

1.2 Εφαρμογές της ανίχνευσης και αναγνώρισης του εντοπισμού κίνησης

Η μελέτη και αναζήτηση καλύτερων αλγορίθμων για καλύτερα αποτελέσματα εντοπισμού αντικειμένων δε γίνεται πια μόνο για ερευνητικούς αλλά και για πρακτικούς λόγους. Το πλήθος των εφαρμογών είναι τεράστιο και συνεχώς αυξάνεται όσο αυξάνεται η ανάγκη μας για πλήρη αυτοματισμό των διαδικασιών παρατήρησης. Οι κυριότερες εφαρμογές εντοπισμού παρακολούθησης είναι οι εξής:

- **Επικοινωνία ανθρώπου – υπολογιστή**

Πριν την ανάπτυξη της μηχανικής όρασης η επικοινωνία ανθρώπου – υπολογιστή γινόταν με τις κλασικές συσκευές εισόδου – εξόδου (π.χ πληκτρολόγιο, οθόνη, ποντίκι)

. Η μηχανική όραση βοήθησε τους τρόπους επικοινωνίας να φτάσουν πιο κοντά στην ανθρώπινη αντίληψη π.χ επικοινωνία μέσω μίας κάμερας [9].

- ***Υποστήριξη ατόμων με ειδικές ανάγκες***

Η μηχανική όραση μπορεί να φανεί αρκετά χρήσιμη στην υποστήριξη ατόμων με ειδικές ανάγκες. Στο πεδίο της επικοινωνίας ανθρώπου – μηχανής έχουν ήδη αναπτυχθεί συστήματα για την υποστήριξη κατάλληλων τρόπων επικοινωνίας, όπως η κατανόηση της νοηματικής γλώσσας και όχι μόνον [9].

- ***Παρακολούθηση χώρων***

Στα συστήματα παρακολούθησης χώρων μας ενδιαφέρει ο έλεγχος της κίνησης και οι συνθήκες της κίνησης ανάλογα με το είδος της κίνησης, την ταυτότητα του ανθρώπου που κινείται, ο χώρος στον οποίο γίνεται η κίνηση κτλ. Τα συστήματα παρακολούθησης συχνά χρησιμοποιούνται σε χώρους που θεωρούνται ιδιαίτερα κρίσιμοι ως προς το θέμα της ασφάλειας, όπως αεροδρόμια, τράπεζες, σχολεία κ.α [9],[11].

- ***Έλεγχος του κυκλοφοριακού συστήματος***

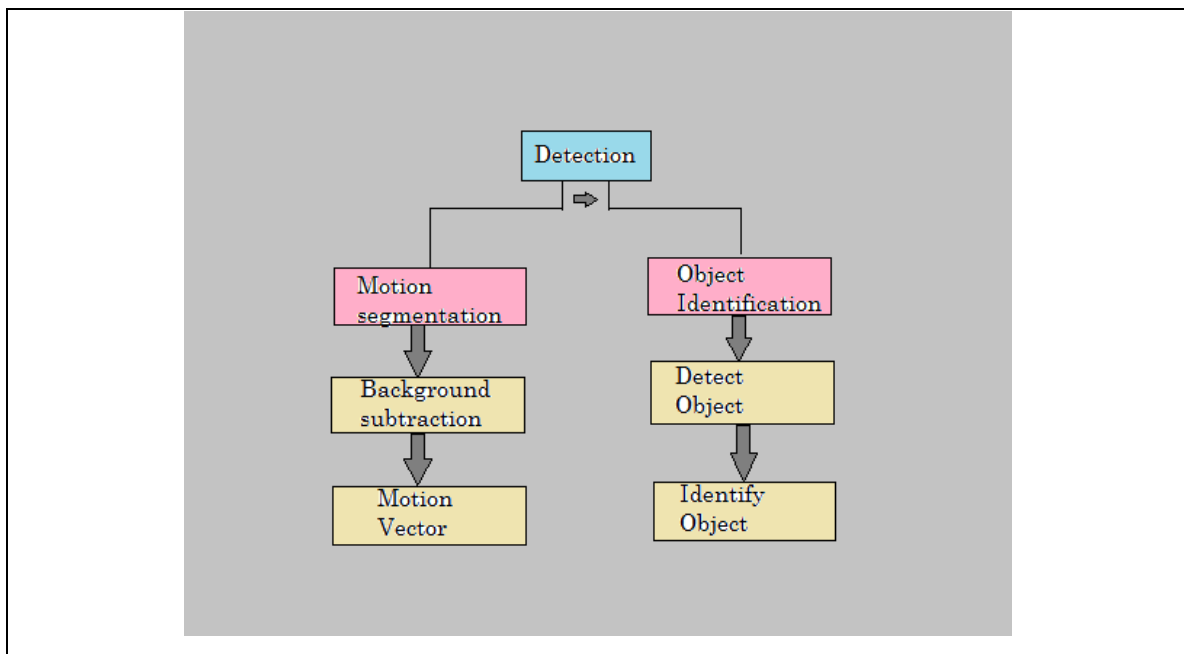
Πολλές φορές ο έλεγχος του κυκλοφοριακού συστήματος γίνεται αυτόματα μέσω συστημάτων παρακολούθησης χώρων. Ήδη χρησιμοποιούνται συστήματα για την καταγραφή παραβιάσεων του κώδικα οδικής κυκλοφορίας σε σημεία υψηλού κινδύνου εκδήλωσης ατυχημάτων ή κατά τη διάρκεια της νύχτας [9].

- ***Ρομποτικές εφαρμογές***

Η χρήση ρομποτικών εργαλείων γίνεται ολοένα και πιο δημοφιλής σε ένα τεράστιο πεδίο εφαρμογών. Ένας από τους βασικότερους στόχους είναι η ανάπτυξη ρομποτικών εργαλείων με δυνατότητες επικοινωνίας με τον άνθρωπο. Η όραση ενός ρομπότ υλοποιείται συνήθως με μία ή περισσότερες κάμερες [9].

1.3 Ανίχνευση της κίνησης

Όταν αναφερόμαστε στο στάδιο της ανίχνευσης της κίνησης από μία σταθερή κάμερα, στην περίπτωση των συγκεκριμένων βίντεο που εξετάστηκαν, εννοούμε τα γενικά βήματα που ακολουθήθηκαν έτσι ώστε να φτάσουμε στον τελικό στόχο που ήταν η ανίχνευση και η ταυτοποίηση των αντικειμένων. Έτσι το στάδιο της ανίχνευσης χωρίζεται στα παρακάτω υποστάδια, όπως φαίνεται στο σχήμα 1.



Σχήμα 1: Γενικό διάγραμμα διαδικασιών εντοπισμού αντικειμένων

Παρακάτω θα γίνει μία γενική ανάλυση των παραπάνω βημάτων του διαγράμματος.

1.3.1 Κατάτμηση κίνησης (Motion Segmentation)

Η κατάτμηση της κίνησης γίνεται για να διαχωριστεί το κινούμενο αντικείμενο, για το οποίο ενδιαφερόμαστε, από την εικόνα του φόντου. Ένας αλγόριθμος κατάτμησης πρέπει να είναι ικανός να ανταπεξέλθει στα πιθανά προβλήματα όπως είναι ο θόρυβος και οι διακυμάνσεις του φωτισμού του φόντου.

- *Αφαίρεση φόντου (Background Subtraction)*: Η αφαίρεση φόντου είναι μια απλή διαδικασία που βοηθά στη λύση της κατάτμησης εικόνας. Μια στατική εικόνα, δηλαδή μία εικόνα σταθερού φόντου ορίζεται ως το βασικό φόντο από το οποίο κάθε τρέχον καρέ (frame) αφαιρείται για να διατηρηθεί μόνο η χρήσιμη πληροφορία, σε ιδανικές συνθήκες.
- *Διανύσματα κίνησης (Motion Vectors)*: Απαραίτητος για την εκτίμηση της κίνησης ενός κινούμενου ή μη αντικειμένου είναι ο υπολογισμός των διανυσμάτων κίνησης. Διάνυσμα κίνησης για τη μηχανική όραση θεωρείται ένα βέλος το οποίο δείχνει τη μετατόπιση και τη φορά της κίνησης κάθε εικονοστοιχείου (pixel) χρήσιμης πληροφορίας.

Στη συνέχεια σκοπός είναι ο εντοπισμός και η ταυτοποίηση των αντικειμένων που χωρίζεται σε δύο υποστάδια, όπως φαίνεται στο παραπάνω διάγραμμα.

- *Εντοπισμός αντικειμένου (Detect Object)*: Κατά το στάδιο του εντοπισμού του αντικειμένου, σκοπός είναι η εξαγωγή, βάσει αλγορίθμου υψηλού επιπέδου, κάθε ομάδας κομματιών (segmented regions) που συντάσσουν κάθε αντικείμενο, βάσει κάποιων κριτηρίων που ονομάζονται χαρακτηριστικά (features).
- *Ταυτοποίηση Αντικειμένου (Identify Object)*: Κατά το στάδιο της ταυτοποίησης του αντικειμένου σκοπός είναι ο καθορισμός κάθε αντικειμένου από τα ήδη εξαγόμενα ως μοναδικό στο σύνολο των καρέ (frames) που συντάσσουν ένα βίντεο. Εν ολίγοις, σε αυτό το στάδιο δίνουμε μία μοναδική ταυτότητα σε κάθε

αντικείμενο, κινούμενο ή σταθερό που το ορίζει ως μοναδικό από την αρχή έως και το τέλος της εμφάνισής του μέσα σε ένα βίντεο.

1.4 Προβλήματα κατά την ανίχνευση των αντικειμένων

Στη μηχανική όραση μία απλή εικόνα, δηλαδή ένα καρέ από μόνο του δε μας δίνει κάποια πληροφορία για τον εντοπισμό ενός αντικειμένου. Ο εντοπισμός ενός αντικειμένου από τον υπολογιστή είναι μία συνολικά δύσκολη διαδικασία και αποτελείται από πολλά υποστάδια, όπως είδαμε στην προηγούμενη ενότητα. Είναι γεγονός ότι δε μπορούμε να καθορίσουμε την τροχιά ενός αντικειμένου παρά μόνο να καταφέρουμε να υποθέσουμε μικρές, σχετικά, μετακινήσεις από καρέ σε καρέ. Κατά τη διαδικασία εντοπισμού και ταυτοποίησης ενός αντικειμένου από μία και σταθερή κάμερα συναντήθηκαν κάποια προβλήματα. Μερικά από αυτά ήταν χαμηλού επιπέδου, δηλαδή οι τρόποι επίλυσής τους ήταν πιο εύκολοι από ότι των προβλημάτων υψηλού επιπέδου, τα οποία για να επιλυθούν χρειάστηκαν δύσκολες και σύνθετες διαδικασίες. Τα βασικότερα προβλήματα που παρουσιάστηκαν αναφέρονται παρακάτω.

Χαμηλού επιπέδου (στάδιο επίλυσης preprocessing):

- *Θόρυβος* που μπορεί να οφείλεται π.χ σε θρόισμα των φύλλων αν και τα δείγματα των βίντεο ήταν σε non-real time με σταθερή κάμερα.
- *Φωτισμός* που οφείλεται σε εναλλαγές του φυσικού ή τεχνητού φωτός και δημιουργεί διακυμάνσεις στην ένταση των εικονοστοιχείων.

Υψηλού επιπέδου (στάδιο επίλυσης processing):

- Δυσκολία ομοιομορφίας των χρωμάτων των εικονοστοιχείων κατά το στάδιο της κατάτμησης κίνησης.
- Δυσκολία εντοπισμού σταθερού αντικειμένου.
- Δυσκολία διαχωρισμού επικαλυπτόμενων αντικειμένων.
- Δυσκολία ταυτοποίησης αντικειμένων στην περίπτωση που τα αντικείμενα μειώνονταν κατά το πέρας του βίντεο, δηλαδή στην περίπτωση που ένα αντικείμενο έφευγε από το χώρο μαγνητοσκόπησης.

Έως τα παραπάνω προβλήματα αφορούν τη μία κάμερα.

- Άτοπος ο εντοπισμός των αντικειμένων από πολλαπλές κάμερες.

Τα παραπάνω προβλήματα για να επιλυθούν, λοιπόν, άλλοτε ήθελαν εύκολες και άλλοτε δύσκολες διαδικασίες. Εύκολες διαδικασίες υπήρξαν οι χρήσεις φίλτρων για την εξομάλυνση του φωτισμού και τη μείωση του θορύβου. Δύσκολες διαδικασίες θεωρούνται οι μέθοδοι που υλοποιήθηκαν για την επίλυση των προβλημάτων υψηλού επιπέδου. Για την επίλυση αυτών των προβλημάτων χρησιμοποιήθηκαν πολυσύνθετες διαδικασίες. Μία από αυτές ήταν ο τρόπος επίλυσης της ανομοιομορφίας των χρωμάτων των εικονοστοιχείων κατά το στάδιο της κατάτμησης κίνησης όπου για κάθε segmented region έπρεπε να μετρηθεί το επικρατόν χρώμα και να οριστεί ως το χρώμα κάθε περιοχής. Παρόλα αυτά σύνθετες ή με μέθοδοι, και μόνο η χρήση τέτοιων διαδικασιών αύξησε την υπολογιστική πολυπλοκότητα της συνολικής διαδικασίας που έχει αναλυθεί έως τώρα.

1.5 Στατική κάμερα και Δυναμική κάμερα

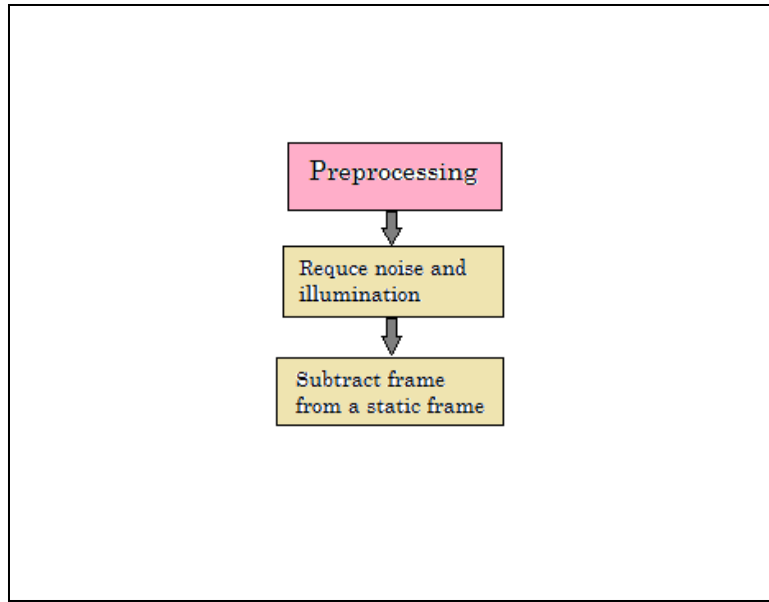
Με τον όρο στατική κάμερα εννοούμε μια σταθερή κάμερα όπου η καταγραφή κάθε σκηνής γίνεται από ένα σημείο παρατήρησης και έτσι όλη η δράση οφείλεται στην κίνηση των αντικειμένων. Αυτό σημαίνει ότι κάθε αλλαγή στο σκηνικό θα πρέπει

να δείχνει την ύπαρξη ενός κινούμενου αντικειμένου που θα μπορεί να εντοπιστεί καθόλη τη βίντεο ακολουθία. Δυστυχώς υπάρχουν όμως, περιπτώσεις που οι αλλαγές στο σκηνικό δεν οφείλονται μόνο σε αντικείμενα του ενδιαφέροντος αλλά και σε ανεπιθύμητο θόρυβο. Για παράδειγμα αυτό συμβαίνει όταν παρουσιάζονται σταδιακές ή απότομες αλλαγές στο φωτισμό, όταν συμβαίνουν ανομοιόμορφες αλλαγές στο φόντο λόγω θορύβου ή όταν η κάμερα υπόκεινται σε μικρές δονήσεις. Με βάση λοιπόν τα παραπάνω, το αρχικό πρόβλημα εντοπισμού των κινούμενων αντικειμένων, μετατρέπεται ουσιαστικά σε πρόβλημα εντοπισμού των κινούμενων αντικειμένων ενδιαφέροντος. Στην περίπτωση της συγκεκριμένης εργασίας ως αντικείμενα ενδιαφέροντος θεωρούνται οι μάζες των αντικειμένων, δηλαδή το περίγραμμα κάθε αντικειμένου μαζί με τον αριθμό των εικονοστοιχείων που τις αποτελούν [9].

Στη περίπτωση της κινούμενης κάμερας. Πέρα από τα προβλήματα θορύβου θα πρέπει να αντιμετωπιστεί και η συνεχής κίνηση-αλλαγή του φόντου λόγω κίνησης του φακού. Σε τέτοιου είδους εφαρμογές είναι λογικό πως η πολυπλοκότητα του προβλήματος είναι ανάλογη της πολυπλοκότητας κίνησης της κάμερας, καθώς απότομη ή περιστροφική κίνηση αλλάζει σε μεγάλο βαθμό το σκηνικό. Στην περίπτωση που τα κινούμενα αντικείμενα που βρίσκονται στη σκηνή είναι παραπάνω από ένα το σύστημα θα πρέπει να υπολογίσει τη σχετική ταχύτητα του κάθε αντικειμένου με το φόντο ώστε να είναι σε θέση να τα διακρίνει μεταξύ τους. Μία τέτοια διαδικασία γίνεται σαφώς πιο δύσκολη όταν τα αντικείμενα του ενδιαφέροντος κινούνται σε μικρές αποστάσεις και τα χρωματικά τους γνωρίσματα δεν αλλάζουν κατά πολύ. Μία από τις πιο γνωστές μεθόδους για την κατάτμηση της κίνησης με κινούμενη κάμερα, όταν η κάμερα κινείται με ομαλό τρόπο και η δειγματοληψία των εικόνων είναι αρκετά μεγάλη, είναι ο υπολογισμός της οπτικής ροής [9].

Κεφάλαιο 2: Αφαίρεση Φόντου

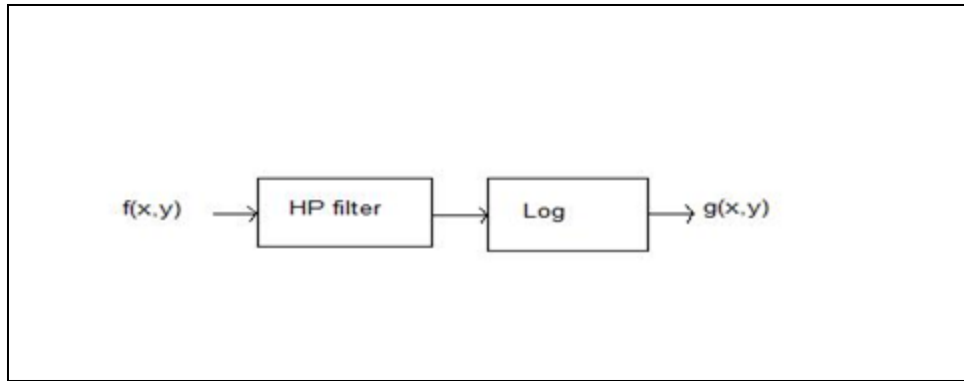
Ένα πολύ σημαντικό βήμα για τον εντοπισμό των αντικειμένων είναι η αφαίρεση φόντου, όπως αναφέρθηκε στο κεφάλαιο 1 . Με τον όρο αφαίρεση φόντου εννοούνται οι διαδικασίες χαμηλού επιπέδου που πρέπει να πραγματοποιηθούν έτσι ώστε να εξαχθεί από το κάθε καρέ μόνο η πληροφορία που ανήκει στα αντικείμενα. Αυτό μπορεί να ακούγεται απλό αλλά δεν είναι τόσο καθώς υπάρχουν παράγοντες θορύβου που επηρεάζουν τα αποτελέσματα. Σύμφωνα με αυτούς τους παράγοντες, ο θόρυβος ή ο φωτισμός δημιουργούν διακυμάνσεις στην ένταση των εικονοστοιχείων, διακυμάνσεις που για ένα υπολογιστικό σύστημα μπορεί να θεωρηθούν και αντικείμενα. Έτσι ρυθμίζοντας το θόρυβο και εξαλείφοντας εικονοστοιχεία θορύβου, δηλαδή εικονοστοιχεία που ανήκουν στο φόντο, κάποιες φορές «χάνεται» και χρήσιμη πληροφορία, κάτι το οποίο όμως είναι αρκετά δύσκολο να ρυθμιστεί. Η δυσκολία λοιπόν, σε αυτό το στάδιο της προεπεξεργασίας των καρέ είναι ότι το αποτέλεσμα δε μπορεί ποτέ να είναι άρτιο, δηλαδή ένα υπολογιστικό σύστημα δε μπορεί ποτέ να εξαγάγει κάθε αντικείμενο όπως φαίνεται «με το μάτι». Παρακάτω, σχήμα 2, φαίνεται το διάγραμμα που δείχνει το μοτίβο της προεπεξεργασίας που ακολουθήθηκε [7],[9],[12].



Σχήμα 2: Διαδικασία Preprocessing

2.1 Μείωση θορύβου και φωτισμού (Reduce noise and illumination)

Μία καλή μέθοδος μείωσης θορύβου και φωτισμού θεωρείται μία διαδικασία μέσω της οποίας μπορούν να μειωθούν παράλληλα ο θόρυβος και ο φωτισμός. Οι μέθοδοι αντιμετώπισης του θορύβου και του φωτισμού είναι αρκετές. Κάποιες από αυτές είναι οι Histogram Equalization, Homomorphic Filtering, isotropic and anisotropic smoothing. Στα συγκεκριμένα βίντεο που επεξεργαστήκαμε κρίθηκε απαραίτητη η χρήση της μεθόδου LogAbout η οποία βασίζεται στη μέθοδο Log [6]. Στο σχήμα 3 παρατηρείται η διαδικασία η οποία ακολουθήθηκε.



Σχήμα 3: LogAbout Method

Η LogAbout μέθοδος είναι μια εύκολη στην υλοποίησης μέθοδος και παράλληλα δεν αυξάνει κατά πολύ την υπολογιστική πολυπλοκότητα. Οι λόγοι που χρησιμοποιήθηκε όμως αυτή η μέθοδος είναι γιατί ήταν αρκετά αποδοτική για τη χρήσιμη πληροφορία που κρατούσε σε σχέση με το αντικείμενό μας δηλαδή, τον εντοπισμό των μαζών των κινούμενων αντικειμένων. Πιο συγκεκριμένα, αποτελείται από το πέρασμα κάθε εικόνας από ένα Gaussian φίλτρο και από τη μη-γραμμική μέθοδο Log. Αν και η Log μέθοδος μπορεί να μειώσει σημαντικά το φωτισμό, συμπιέζει τα high gray εικονοστοιχεία ταυτόχρονα. Έτσι μερικές ακμές, δηλαδή κομμάτι χρήσιμης πληροφορίας χάνεται. Με τη χρήση ενός Gaussian φίλτρου ή μίας high pass μάσκας 3x3 μπορούμε να αυξήσουμε την ένταση των υψηλών συχνοτήτων έντασης εικονοστοιχείων ενώ παράλληλα να μην χάνουμε πληροφορία από τα χαμηλής συχνότητας εικονοστοιχεία [].

Όπως παρατηρείται στο παραπάνω σχήμα 3, $f(x,y)$ είναι η είσοδος δηλαδή κάθε καρέ το οποίο εισέρχεται στη διαδικασία. Κάθε εικονοστοιχείο λοιπόν του καρέ περνάει από ένα high-pass φίλτρο, φίλτρο το οποίο μειώνει αρκετά το θόρυβο και εξομαλύνει τις διακυμάνσεις, ανομοιομορφίες του φωτισμού ενώ παράλληλα αυξάνει την ένταση των high gray εικονοστοιχείων, δηλαδή των εικονοστοιχείων που έχουν υψηλές τιμές έντασης. Αυτό συμβαίνει για να μην χάνονται κάποιες ακμές.

Στη συνέχεια της διαδικασίας λογαριθμίζουμε κάθε εικόνα κάτι το οποίο βοηθάει στο να εξομαλυνθούν οι τιμές του φωτισμού κάθε εικόνας ενώ παράλληλα συμπιέζεται η ένταση της χρήσιμης πληροφορίας, γι' αυτό χρησιμοποιούμε το Gaussian φίλτρο. Η μέθοδος υλοποιείται ως εξής:

$$g(x, y) = a + \frac{\log(f(x, y) + 1)}{b \log(c)}$$

Όπου:

$g(x,y)$: the output , the original image log

$f(x,y)$: the input, the original image

a,b,c : parameters which can control the location and the shape of the curve $g-f$

(πειραματικές στην περίπτωση μας)

Έτσι βάσει της σύνθετης αυτής μεθόδου μπορεί να μειώνεται παράλληλα και ο θόρυβος, μέσω του φίλτρου αλλά και ο φωτισμός μέσω της μεθόδου Log. Το τελικό αποτέλεσμα μας δίνει την ίδια εικόνα με την αρχική, έξοδος $g(x,y)$, στην οποία τα εικονοστοιχεία της χρήσιμης πληροφορίας, πια, έχουν μεγαλύτερες τιμές έντασης από ότι οι τιμές των εικονοστοιχείων του φόντου.

2.2 Παρουσίαση μεθόδων αφαίρεσης φόντου

Η επιλογή της μεθόδου διαχωρισμού των αντικειμένων από το φόντο είναι πολύ σημαντική για τις μετέπειτα διαδικασίες αναγνώρισης και ταυτοποίησης αυτών. Οι μέθοδοι που επιλέχθηκαν για την αφαίρεση του φόντου ήταν δύο:

- Αφαίρεση διαδοχικών καρτέ
- Αφαίρεση τρέχοντος καρτέ από ένα σταθερού φόντου

Η πρώτη ήταν η διαδοχική αφαίρεση των καρτέ δηλαδή η αφαίρεση κάθε τρέχοντος καρτέ από το προηγούμενό του. Η δεύτερη ήταν η αφαίρεση κάθε καρτέ από ένα σταθερού φόντου, δηλαδή από ένα καρτέ που απεικονίζει εξολοκλήρου το δισδιάστατο χώρο στον οποίο θα κινούνταν τα αντικείμενα του βίντεο. Σύμφωνα με την πρώτη μέθοδο αρχικά ορίζουμε έναν πίνακα με διαστάσεις ίδιες με αυτές των καρτέ τον οποίο αρχικοποιούμε. Στη συνέχεια για κάθε δύο διαδοχικά καρτέ παίρνουμε κάθε εικονοστοιχείο και βρίσκουμε τη διαφορά της τιμής του μεταξύ των δύο καρτέ. Ορίζουμε κατώφλι (threshold $0 < \tau < 1$). Τέλος, αν η διαφορά των τιμών του εικονοστοιχείου είναι μικρότερη της τιμής κατωφλίου, τότε βάζουμε στην αντίστοιχη θέση του πίνακα την τιμή 1, ενώ αν η διαφορά είναι μεγαλύτερη του κατωφλίου, τότε στην αντίστοιχη θέση του πίνακα βάζουμε την τιμή 0 [9]. Η πρώτη μέθοδος της διαδοχικής αφαίρεσης δε λειτουργούσε καλά καθώς με αυτόν τον τρόπο, αντικείμενα που βρίσκονταν στην ίδια θέση, άρα ήταν σταθερά, αναγνωρίζονταν ως φόντο και εξαλείφονταν. Έτσι, κρίθηκε άκρως απαραίτητη η χρήση της δεύτερης μεθόδου. Σύμφωνα με τη δεύτερη μέθοδο, τώρα, για κάθε βίντεο προς επεξεργασία, ορίζεται ένα κενό καρτέ ως καρτέ σταθερού φόντου και κάθε τρέχον καρτέ αφαιρείται από αυτό. Η δεύτερη μέθοδος γενικά θεωρείται πιο απλή και πρακτικά πιο κατάλληλη για τις συγκεκριμένες περιπτώσεις βίντεο.

2.3 Σύγκριση αποτελεσμάτων

Παρακάτω γίνεται σύγκριση των αποτελεσμάτων των δύο μεθόδων. Για τα ίδια καρέ θα εμφανιστούν τα αντίστοιχα αποτελέσματα. Θα δοθεί έμφαση στο καρέ 1078 στο οποίο φαίνεται ένα σταθερό αντικείμενο. Όπως μπορεί κανείς να εξετάσει συγκρίνοντας τις εικόνες 2,4 με 6,7 μπορεί να δει ότι με την πρώτη μέθοδο η βαλίτσα που φαίνεται στο καρέ 1021 που πάει να αφήσει ο άνθρωπος στο καρέ 1078 (εικόνα 4) έχει εξαλειφθεί σε αντίθεση με τη δεύτερη μέθοδο όπου η βαλίτσα εντοπίζεται (εικόνα 7).

➤ Αποτελέσματα της πρώτης μεθόδου:



Εικόνα 1

Αυθεντική εικόνα του καρέ 1021

frame1021



Εικόνα 2

Διαδοχική αφαίρεση του καρέ 1021 από το καρέ 1020



Εικόνα 3

Αυθεντική εικόνα του καρέ 1078

frame1078



Εικόνα 4

Διαδοχική αφαίρεση του καρέ 1078 από το καρέ 1077

Από τα αποτελέσματα της πρώτης μεθόδου μπορεί κανείς να δει στην εικόνα 4 ότι αρχικά απουσιάζει η βαλίτσα που έχει αφήσει ο άνθρωπος. Αυτό είναι λογικό καθώς το υπολογιστικό σύστημα καταγράφει τη βαλίτσα ως φόντο οπότε με τη διαδοχική αφαίρεση των καρέ διαγράφεται και η βαλίτσα πέρα από το φόντο. Επίσης μπορεί κανείς να παρατηρήσει τα αποτελέσματα της μεθόδου LogAbout. Όπως παρατηρείται ειδικά στην εικόνα 4, έχει χαθεί αρκετή χρήσιμη πληροφορία κυρίως από το δεύτερο άνθρωπο. Άρα η μέθοδος LogAbout και η μέθοδος της διαδοχικής αφαίρεσης των καρέ δε λειτουργούσαν αποδοτικά. Ας δούμε όμως και τα αποτελέσματα της δεύτερης μεθόδου.

➤ Αποτελέσματα της δεύτερης μεθόδου:



Εικόνα 5

Αυθεντική εικόνα σταθερού φόντου

frame1021



Εικόνα 6

Αφαίρεση καρέ 1021 από του καρέ σταθερού φόντου

frame1078



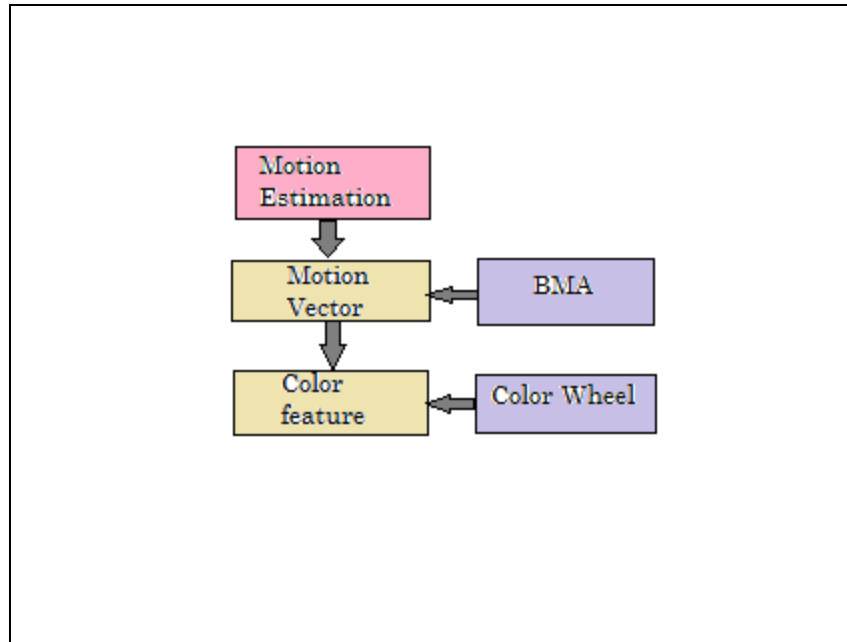
Εικόνα 7

Αφαίρεση καρέ 1078 από του καρέ σταθερού φόντου

Γενικά, η δεύτερη μέθοδος λοιπόν, θεωρήθηκε καλύτερη καθώς εντοπίζεται η βαλίτσα. Εκτός αυτού μπορεί κανείς να παρατηρήσει ότι στο καρέ 1078 (εικόνα 7) η πληροφορία η οποία έχει διατηρηθεί είναι σαφώς μικρότερη από την πραγματική, πρόβλημα που αναφέρθηκε και οφείλεται στο στάδιο μείωσης του θορύβου, που εκτός που το φόντο κάποιες φορές χάνουμε και χρήσιμη πληροφορία, πρόβλημα όμως που στην ουσία δε θα λύσουμε καθώς η μάζα που έχουμε εντοπίσει είναι ικανοποιητική για τον τελικό μας στόχο. Επίσης παρατηρείται ελλιπής πληροφορία από το παντελόνι του ανθρώπου στην εικόνα 6, κάτι που οφείλεται στο γεγονός ότι το γενικά τα ρούχα του ανθρώπου αυτού είχαν ίδιο χρώμα με το φόντο. Άρα η δεύτερη μέθοδος λειτουργεί γενικά καλά και σε συνδυασμό με τη μέθοδο LogAbout δίνουν ένα αρκετά ικανοποιητικό αποτέλεσμα για τον αυτοσκοπό μας, δηλαδή τον εντοπισμό των μαζών των κινούμενων αντικειμένων.

Κεφάλαιο 3: Εκτίμηση κίνησης

Στο επόμενο στάδιο του εντοπισμού και της ταυτοποίησης των αντικειμένων σκοπός είναι η εκτίμηση της κίνησης. Με τον όρο εκτίμηση κίνησης εννοούμε τη δημιουργία ενός διανύσματος κίνησης (motion vector) το οποίο να δείχνει τη φορά και τη διεύθυνση της κίνησης, διάνυσμα που στην πορεία μέσω ενός χρωματικού κύκλου (color wheel), όπως φαίνεται στο παρακάτω σχήμα 4, παίρνει ένα χαρακτηριστικό (feature), το χρώμα [2].



Σχήμα 4: Διάγραμμα εκτίμησης της κίνησης

3.1 Εκτίμηση κίνησης (motion estimation)

Το αρχικό στάδιο για τον εντοπισμό των αντικειμένων είναι να εξετάσουμε την κίνηση σε επίπεδο εικονοστοιχείων (pixels). Αυτός ο υπολογισμός μπορεί να γίνει μέσω δύο μεθόδων της εκτίμησης κίνησης (motion estimation) και της οπτικής ροής (optical flow). Εκτίμηση κίνησης είναι η διαδικασία καθορισμού των motion vectors που περιγράφουν τη μεταφορά της πληροφορίας ανάμεσα σε δύο 2D εικόνες, δηλαδή σε δύο γειτονικά καρέ μίας βίντεο ακολουθίας. Η διαφορά της από την οπτική ροή είναι ότι στη motion estimation διαδικασία δεν είναι απαραίτητη η 1:1 αντιστοιχία των θέσεων των εικονοστοιχείων που σημαίνει ότι μπορεί να γίνει επεξεργασία και ομάδας εικονοστοιχείων (macro blocks). Επίσης, βέβαια, σημαντικό είναι το γεγονός ότι με την οπτική ροή μελετάται η αντικειμενική κίνηση των εικονοστοιχείων στο χώρο 3D. Γενικότερα σε επίπεδο επεξεργασίας, η motion estimation διαδικασία κρίθηκε

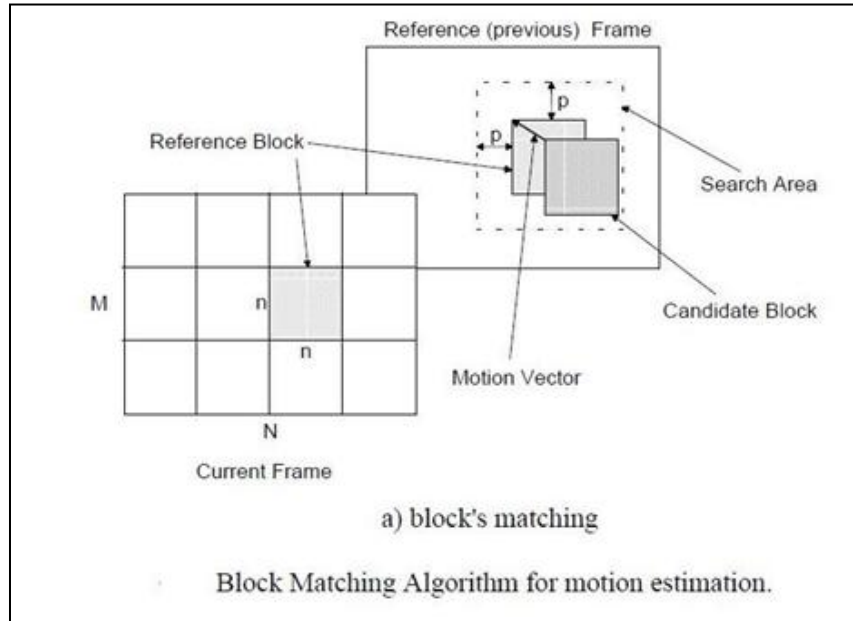
καταλληλότερη για το ύψος της κίνησης την οποία θέλαμε να εξετάσουμε καθότι, όπως έχει αναφερθεί μελετάται η μάζα των αντικειμένων. Η optical flow διαδικασία δεν είναι τόσο εύρωστη σε έντονες διακυμάνσεις της έντασης των εικονοστοιχείων καθώς βασίζεται στην υπόθεση ότι η ένταση των εικονοστοιχείων πρέπει να είναι σταθερή σε μία ακολουθία εικόνων.

3.2 Υπολογισμός του διανύσματος κίνησης μέσω Block Matching Algorithm

Ο πρώτος βασικός αλγόριθμος, υψηλού επιπέδου, που υλοποιήθηκε ήταν ο Block Matching Algorithm (BMA). Ο BMA είναι μια τεχνική κωδικοποίησης κίνησης σε βίντεο ακολουθία. Σκοπός του είναι η ανίχνευση της κίνησης μεταξύ δύο διαδοχικών εικόνων, αρχικά σε επίπεδο blocks και έπειτα σε επίπεδο εικονοστοιχείων (pixels), για την εξαγωγή του motion vector. Αρχικά ο αλγόριθμος υλοποιείται σε επίπεδο block. Διαιρώντας κάθε καρέ σε μη επικαλυπτόμενα τετράγωνα κομμάτια καθορίζονται τα blocks. Σκοπός του BMA είναι να ταιριάζει κάθε block του τρέχοντος καρέ (current frame) με ένα του καρέ προορισμού (destination frame) κάνοντας shift το current frame σε μια προκαθορισμένη γειτονιά από pixels (search area) στο destination frame [4],[5]. Πιο συγκεκριμένα:

- Για κάθε shift (σύγκριση) υπολογίζεται το άθροισμα των αποστάσεων των gray values των δύο blocks.
- Το shift του block με το minimum total distance θεωρείται το best matching (βάσει του κριτηρίου SAD-sum of absolute differences). Σε ιδανικές συνθήκες τα δύο matching blocks ταυτίζονται.
- Τέλος, ενώνοντας την απόσταση μεταξύ των δύο εξεταζόμενων blocks, παίρνουμε το motion vector (direction και magnitude).

Σχηματικά η διαδικασία φαίνεται στο σχήμα 5.



Σχήμα 5

Όπως φαίνεται στο παραπάνω σχήμα, χωρίζουμε κάθε frame_t σε $M \times N$ blocks. Για κάθε block ορίζουμε μία search area μεγέθους $(n+p) \times (n+p)$, όπου p είναι ο searching factor. Ψάχνοντας σε κάθε πιθανή θέση της search area του frame_{t+1} όπου μπορεί να βρίσκεται το current block, υπολογίζουμε το SAD, σύμφωνα με τον τύπο. Το block με το minimum SAD θεωρείται το best-matching.

$$\text{SAD}(\hat{u}, \hat{v}) = \sum_{j=0}^{N-1} \sum_{i=0}^{N-1} |g_t(x+i, y+j) - g_{t+1}(x+\hat{u}+i, y+\hat{v}+j)|$$

Όπου:

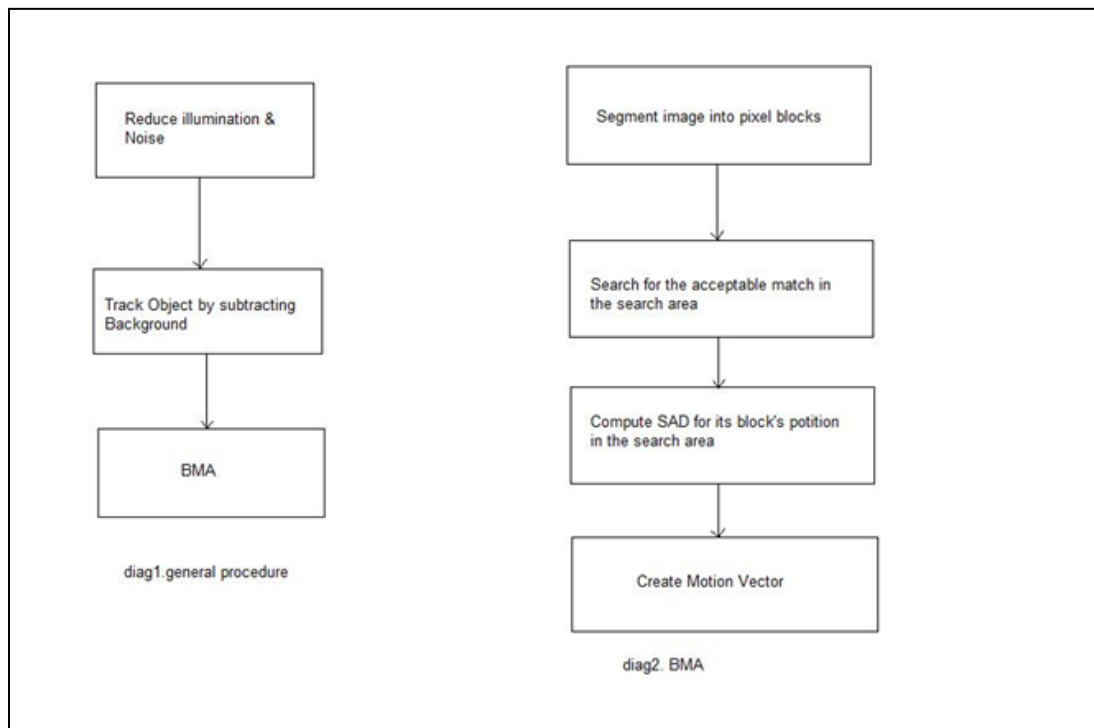
g_t : is the gray value of a pixel in the current frame

g_{t+1} : is the gray level of a pixel in the next frame

u : μετατόπιση κατά άξονα x

y : μετατόπιση κατά άξονα y

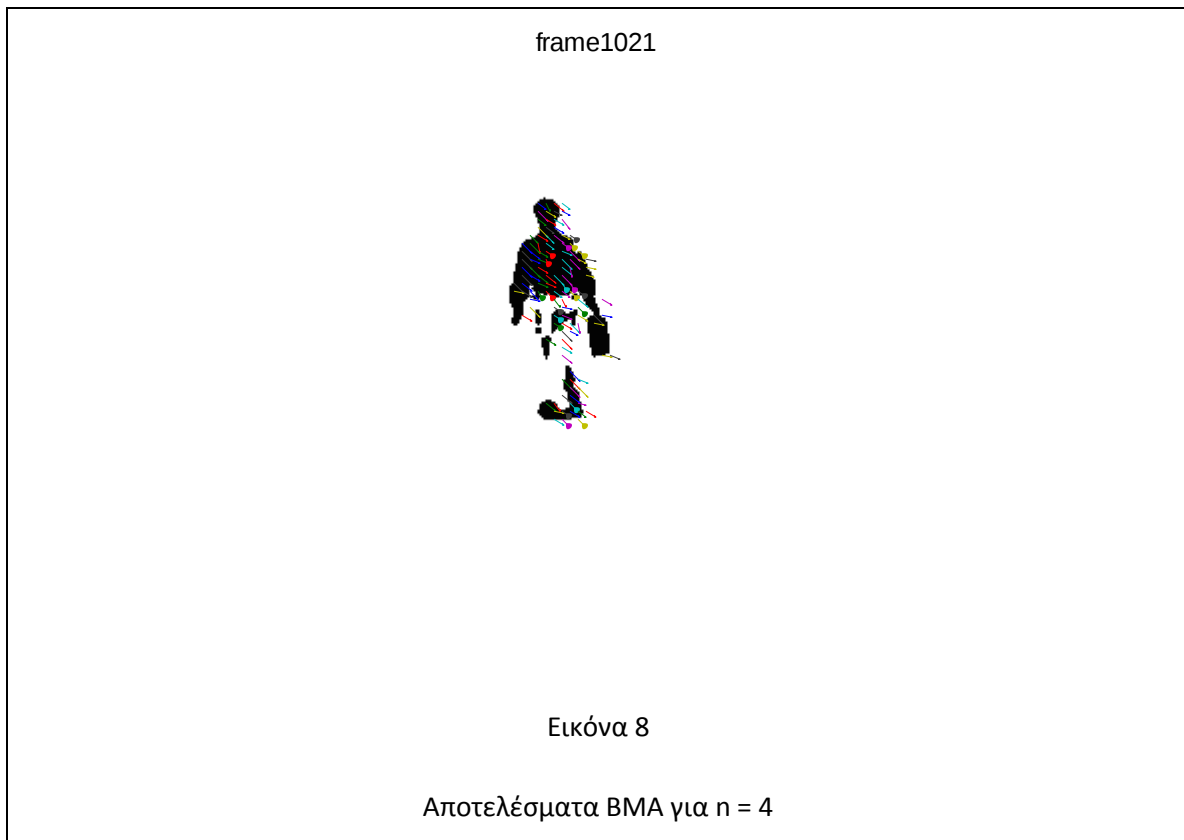
Έχοντας λοιπόν, την τρέχουσα θέση του block στο current frame και την αντίστοιχη θέση του στο next frame , ενώνοντας τα σημεία δημιουργούμε τον motion vector. Παρακάτω φαίνονται τα στάδια του αλγορίθμου από το preprocessing στάδιο έως και το στάδιο του υπολογισμού του motion vector.



Σχήμα 6: Σύνθετη διαδικασία για τη δημιουργία του διανύσματος κίνησης

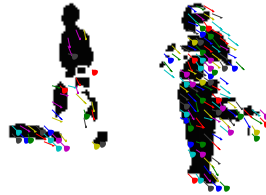
Στη συνέχεια, μετά την υλοποίηση του παραπάνω αλγορίθμου, ο αλγόριθμος εφαρμόζεται για κάθε ρίχει της χρήσιμης πληροφορίας όπως θα δούμε στα αποτελέσματα παρακάτω. Θα δούμε δείγματα για $n = 4$, $n = 8$ και $n = 1$.

➤ $n = 4$:



Αρχικά παρουσιάζονται τα αποτελέσματα για $n = 4$ δηλαδή, για block αριθμού εικονοστοιχείων ίσο με 4. Πρακτικά αυτό σημαίνει, όπως φαίνεται στην παραπάνω εικόνα πως κάθε 4 εικονοστοιχεία έχω έναν motion vector.

frame1078



Εικόνα 9

Αποτελέσματα BMA για $n = 4$

Όπως παρατηρήθηκε και παραπάνω, για $n = 4$ η εκτίμηση της κίνησης δεν είναι τόσο ακριβής καθώς τα διανύσματα κίνησης δεν μπορούν να καθορίσουν πλήρως την κίνηση κάθε αντικειμένου.

➤ $n = 8$:

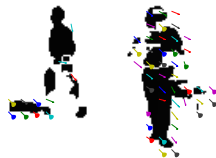
frame1021



Εικόνα 10

Αποτελέσματα BMA για $n = 8$

frame1078



Εικόνα 11

Αποτελέσματα BMA για $n = 8$

Όπως παρατηρείται από τα παραπάνω αποτελέσματα, μεγαλώνοντας την παράμετρο n από $n = 4$ σε $n = 8$ η εκτίμηση της κίνησης και πάλι είναι λιγότερο σαφής. Μπορεί λοιπόν κανείς να παρατηρήσει ότι όσο μεγαλώνει το n τόσο λιγότερο ακριβή αποτελέσματα έχουμε. Η ιδανική περίπτωση για την ακριβέστερη εκτίμηση κίνησης είναι για $n = 1$.

➤ $n = 1$:

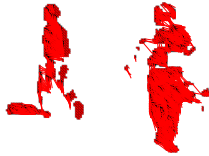
frame1021



Εικόνα 12

Αποτελέσματα BMA για $n = 1$

frame1078



Εικόνα 13

Αποτελέσματα BMA για $n = 1$

Η τελευταία περίπτωση που έχουμε να παρουσιάσουμε είναι για $n = 1$. Η επιλογή του $n = 1$ έγινε τελευταία γιατί ήταν κι εκείνη που χρησιμοποιήσαμε στην εργασία αυτή. Όπως παρατηρούμε από τα παραπάνω αποτελέσματα, όσο μικρότερο είναι το n , δηλαδή το μέγεθος του block, τόσο ακριβέστερη γίνεται η εκτίμηση της κίνησης. Για μεγάλο n , δε μπορούμε να εκτιμήσουμε με ακρίβεια την κίνηση, καθώς με το να υπάρχει και ο θόρυβος τα αποτελέσματα δεν είναι τόσο ομοιόμορφα όσο αν οι εντάσεις της χρήσιμης πληροφορίας ήταν πιο ομοιόμορφες. Έτσι, για $n = 1$, έχουμε τη ιδανικότερη περίπτωση εκτίμησης της κίνησης, γιατί ακόμη και τα προβλήματα ανομοιομορφίας των εντάσεων δε μας αγγίζουν, όπως φαίνεται στις εικόνες 12,13. Γενικά λοιπόν, ο αλγόριθμος λειτουργεί με αρκετή ακρίβεια για μικρό block size αλλά παράλληλα είναι «ακριβός» για όγκο δεδομένων. Έτσι επιλέχτηκε να βρεθούν τα motion vectors μόνο της χρήσιμης πληροφορίας. Από τη άλλη, για μεγαλύτερο block size, τα αποτελέσματα γίνονται λιγότερο ακριβή αλλά μειώνεται η πολυπλοκότητα του αλγορίθμου.

3.3 Χρωματισμός των διανυσμάτων κίνησης βάσει του χρωματικού κύκλου

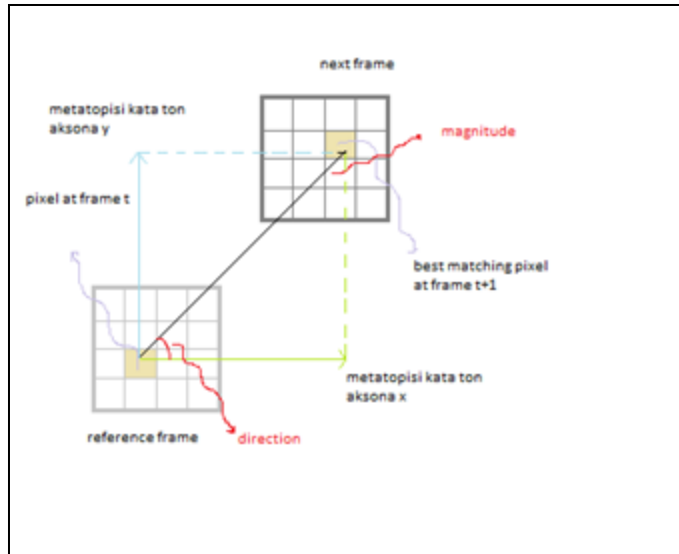
Στη συνέχεια, για το στόχο του Object Tracking, δηλαδή του εντοπισμού των αντικειμένων πρέπει να γνωρίζω τη θέση που βρίσκεται κάθε αντικείμενο κάθε χρονική στιγμή, δηλαδή σε κάθε καρέ. Έως τώρα λοιπόν με τη BMA διαδικασία έχει επιτευχθεί η εξαγωγή της μετατόπισης κάθε pixel χρήσιμης πληροφορίας, ουσιαστικά δηλαδή, της θέσης που βρίσκεται κάθε εξεταζόμενο pixel. Εν συνεχεία, με τον όρο motion estimation σημαίνει υπολογισμός του μήκους και της κατεύθυνσης (magnitude και direction) κάθε motion vector ως εξής:

Υποθέτοντας ότι a , b είναι οι μετατοπίσεις κατά άξονες x , y αντίστοιχα, έχουμε:

$$m = \sqrt{a^2 + b^2}, m \in [0,1]$$

$$d = \frac{\arctan(a,b)}{\pi}$$

Όπως παρατηρείται από το παρακάτω σχήμα, μέσω του BMA γενικά έχει δημιουργηθεί ένας πίνακας ο οποίος αποτελείται από τις συντεταγμένες κάθε pixel στο τρέχον και του προορισμού καρέ. Κατά το στάδιο του motion estimation γίνεται υπολογισμός των magnitude (μήκος) και direction (κατεύθυνση) κάθε διανύσματος κίνησης αφού γίνεται αναφορά σε ένα διανυσματικό μέγεθος.

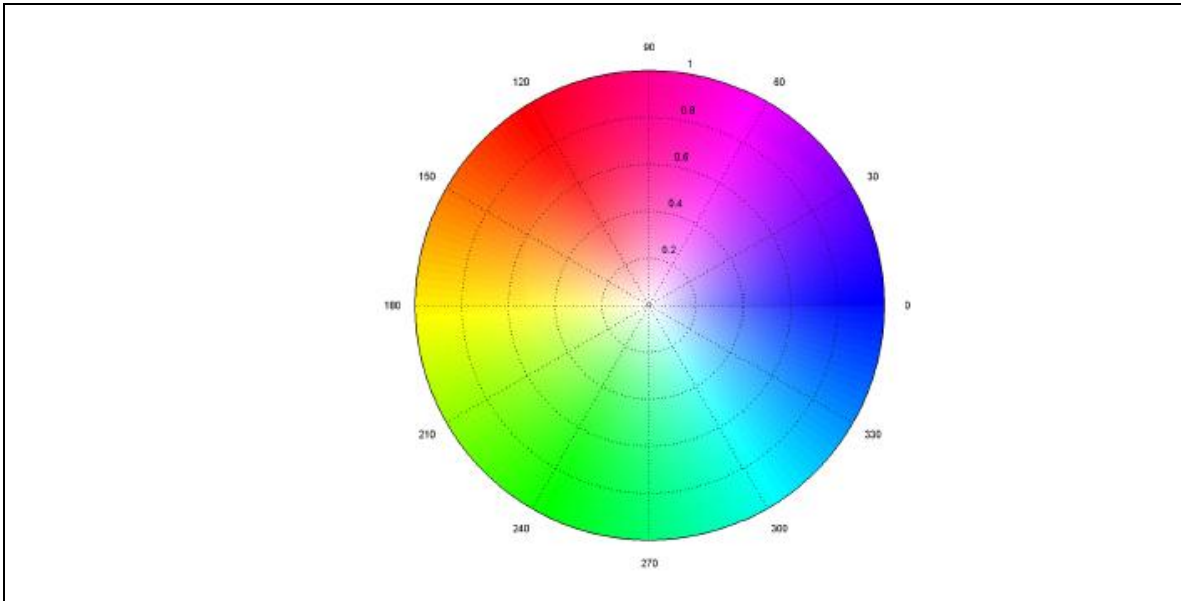


Σχήμα 7: motion vector

Αφού ληφθεί αυτή η χρήσιμη πληροφορία στόχος είναι να γίνει ομαδοποίηση των pixels βάσει κάποιων κριτηρίων. Ένα πολύ σημαντικό κριτήριο ομαδοποίησης είναι το χρώμα. Το χρώμα είναι απαραίτητος παράγοντας που ωθεί την εργασία στο να δουλεύουμε από επίπεδο εικονοστοιχείου σε επίπεδο εικονοστοιχείων και μετέπειτα σε επίπεδο περιοχών. Είναι απαραίτητο καθώς δίνει τη δυνατότητα κάθε διάνυσμα κίνησης να παίρνει ένα χρώμα από έναν χρωματικό κύκλο, χρώμα που αντιστοιχεί στην κατεύθυνσή του (διεύθυνση και φορά). Η απόχρωση τώρα κάθε διανύσματος αντιστοιχεί στο μήκος του. Έτσι το χρώμα βοηθάει στο να ομαδοποιηθούν εικονοστοιχεία κοντινής ή ίδιας απόχρωσης και με την προαιρετική χρήση ενός bounding box να έχουμε μια ομοιομορφία στην φορά της κίνησης κάθε κομματιού κάθε αντικειμένου. Έτσι κρίθηκε απαραίτητη η διαδικασία δημιουργίας ενός Color Coding, δηλαδή ενός χρωματικού κώδικα κατά τον οποίο κάθε εικονοστοιχείο πλέον θα αντιστοιχεί σε ένα χρώμα.

3.4 Χρωματικός κύκλος για ομαδοποίηση των διανυσμάτων κίνησης

Γενικά ο color wheel βασίζεται στην HSV color αναπαράσταση. Η κατασκευή του είναι αποτέλεσμα του συνδυασμού των basic colors (red, green, blue) και των primary colors (cyan, magenta, yellow), δηλαδή του συνδυασμού των basic colors. Σκοπός του color wheel είναι να αντιστοιχίσουμε κάθε motion vector σε μία απόχρωση βάσει του magnitude και του direction του. Το αυθεντικό μοντέλο του color wheel είναι το παρακάτω:



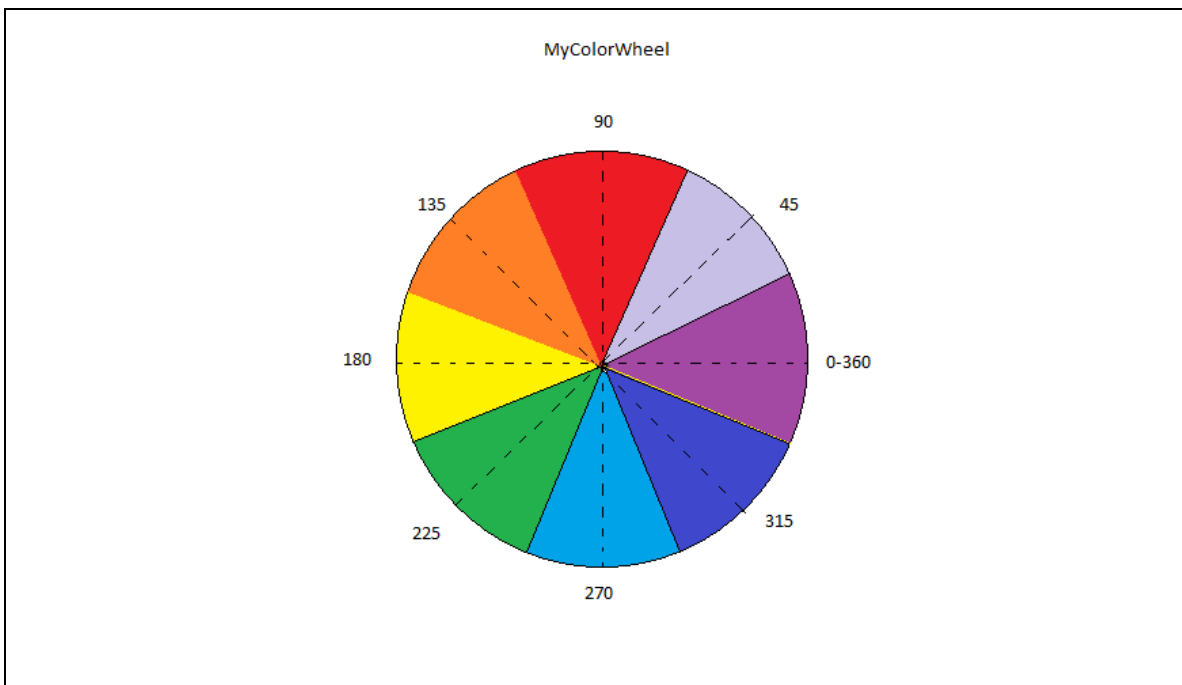
Σχήμα 8: Authentic Color Wheel

Βασισμένοι στο παραπάνω μοντέλο λοιπόν, δημιουργήσαμε έναν χρωματικό κώδικα, ο οποίος εξυπηρετούσε τις ανάγκες των δειγμάτων των videos. Πιο συγκεκριμένα, χρησιμοποιήσαμε τα τρία basic colors (red, green, blue) καθώς και τους τρεις συνδυασμούς τους, δηλαδή yellow, cyan, magenta ενώ προσθέσαμε και κάποιες αποχρώσεις αυτών το ανοιχτό magenta και το orange. Κάθε χρώμα στο χρωματικό κώδικα καθορίζει και μία κατεύθυνση στην οποία μπορεί να κινηθεί ένα αντικείμενο στο δισδιάστατο χώρο που κινούνται τα αντικείμενα των videos.

Έτσι δημιουργήθηκαν οχτώ πιθανές κατευθύνσεις, χωρίζοντας τον κύκλο ανά 45°. Αναλυτικότερα, έχουμε καθορίσει τις εξής κατευθύνσεις:

1. 0° - 22.5° & 337.5° - 360°: Magenta

2. $22.5^{\circ} - 67.5^{\circ}$: Light Magenta
3. $67.5^{\circ} - 112.5^{\circ}$: Red
4. $112.5^{\circ} - 157.5^{\circ}$: Orange
5. $157.5^{\circ} - 202.5^{\circ}$: Yellow
6. $202.5^{\circ} - 247.5^{\circ}$: Green
7. $247.5^{\circ} - 292.5^{\circ}$: Cyan
8. $292.5^{\circ} - 337.5^{\circ}$: Blue



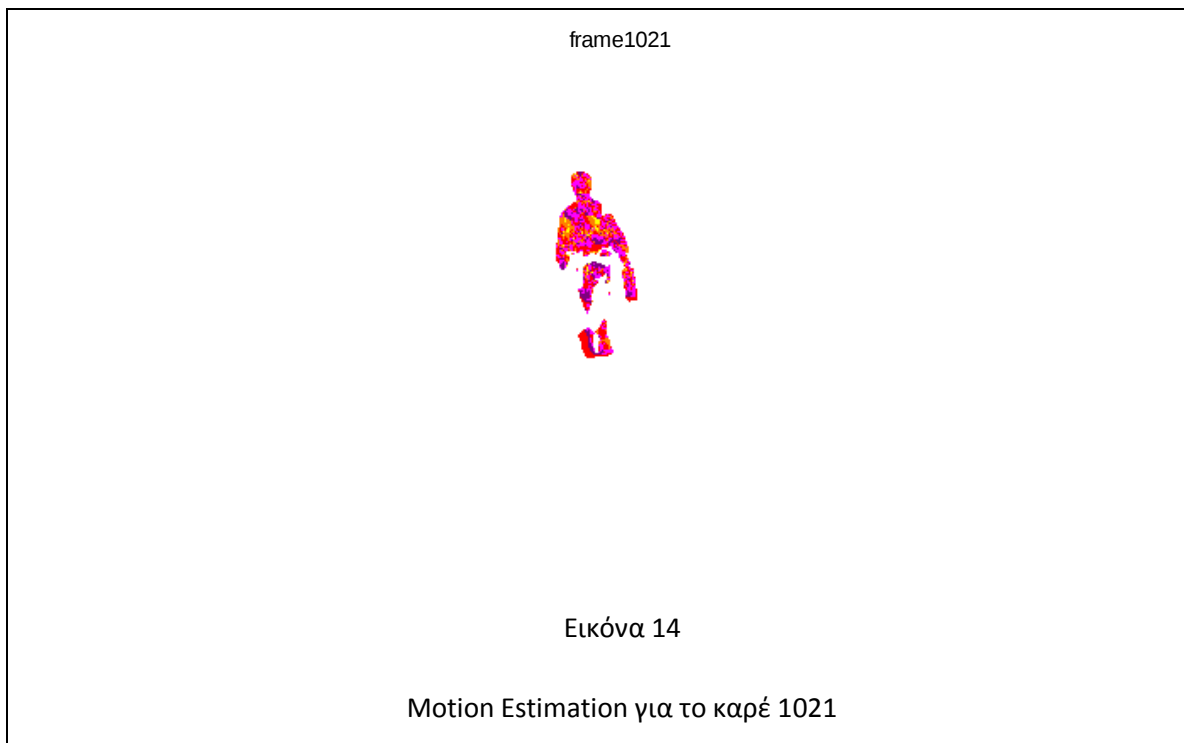
Σχήμα 9: Color Wheel

Να αναφερθεί ότι εργαστήκαμε σε κλίμακα $[360,3]$, όπου το «3» καθορίζεται από το RGB μοντέλο, δηλ. καθορίζει την ένταση κάθε pixel. Για κάθε pixel's direction υπάρχει πια κι ένα χρώμα που ουσιαστικά ,λοιπόν, δείχνει προς τα πού κινείται. Έτσι

έχουμε καταφέρει να δημιουργήσουμε ένα χρωματισμένο βέλος το οποίο δε δείχνει μόνο το που βρίσκεται κάθε εικονοστοιχείο αλλά δείχνει και το προς τα πού κινείται.

Σημείωση: Στο δικό μας color wheel δε χρησιμοποιήσαμε την ακτίνα, γιατί αν και θα είχαμε μεγαλύτερη ακρίβεια, αυτό το μοντέλο πληρούσε της προϋποθέσεις των συγκεκριμένων videos.

Παρακάτω φαίνονται τα αποτελέσματα του motion estimation από τα δείγματα των βίντεο που επεξεργάστηκαν.



Στην παραπάνω εικόνα παρατηρείται ότι η επικρατούσα τιμή είναι το ανοιχτό μοβ ενώ θα έπρεπε να υπερισχύουν τα χρώματα κόκκινο-κίτρινο του δικού μας χρωματικού κύκλου.

frame1046



Εικόνα 15

Motion Estimation για το καρέ 1046

Στο καρέ 1046 ο πρώτος άνθρωπος του βίντεο αφήνει τη βαλίτσα. Παρατηρείται ότι η κατεύθυνσή της δεν είναι μηδενική.

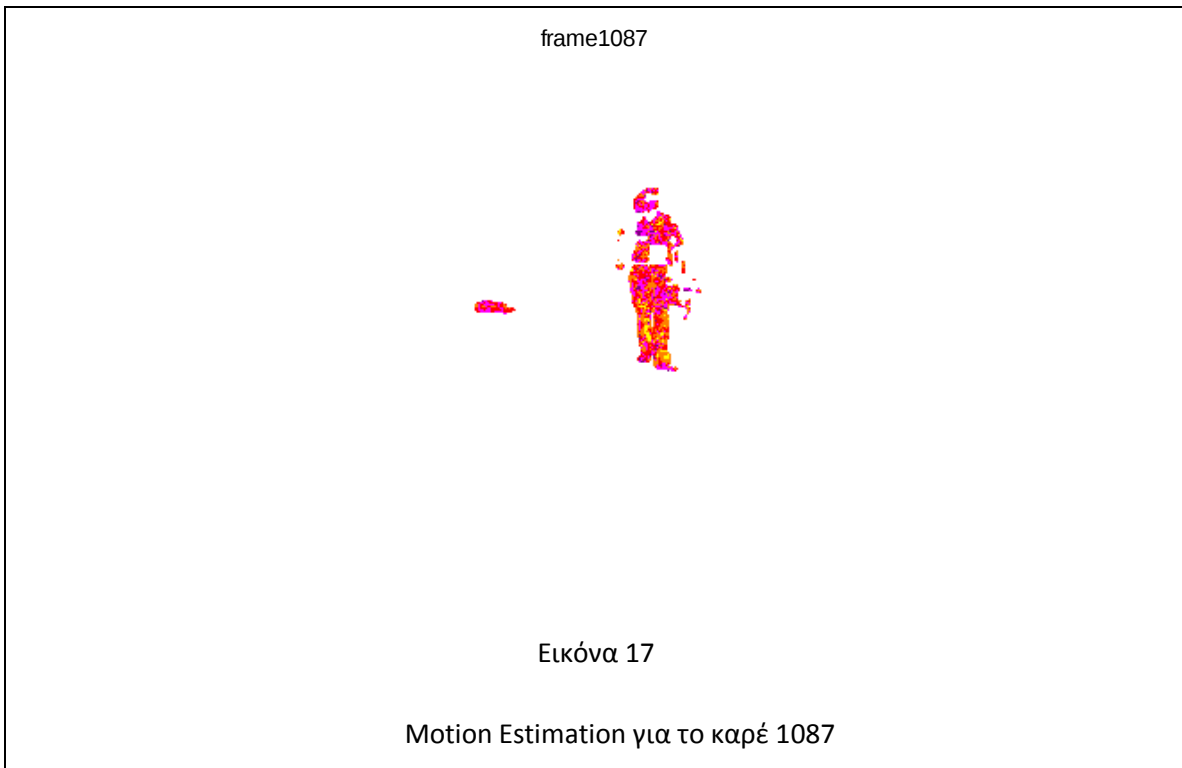
frame1078



Εικόνα 16

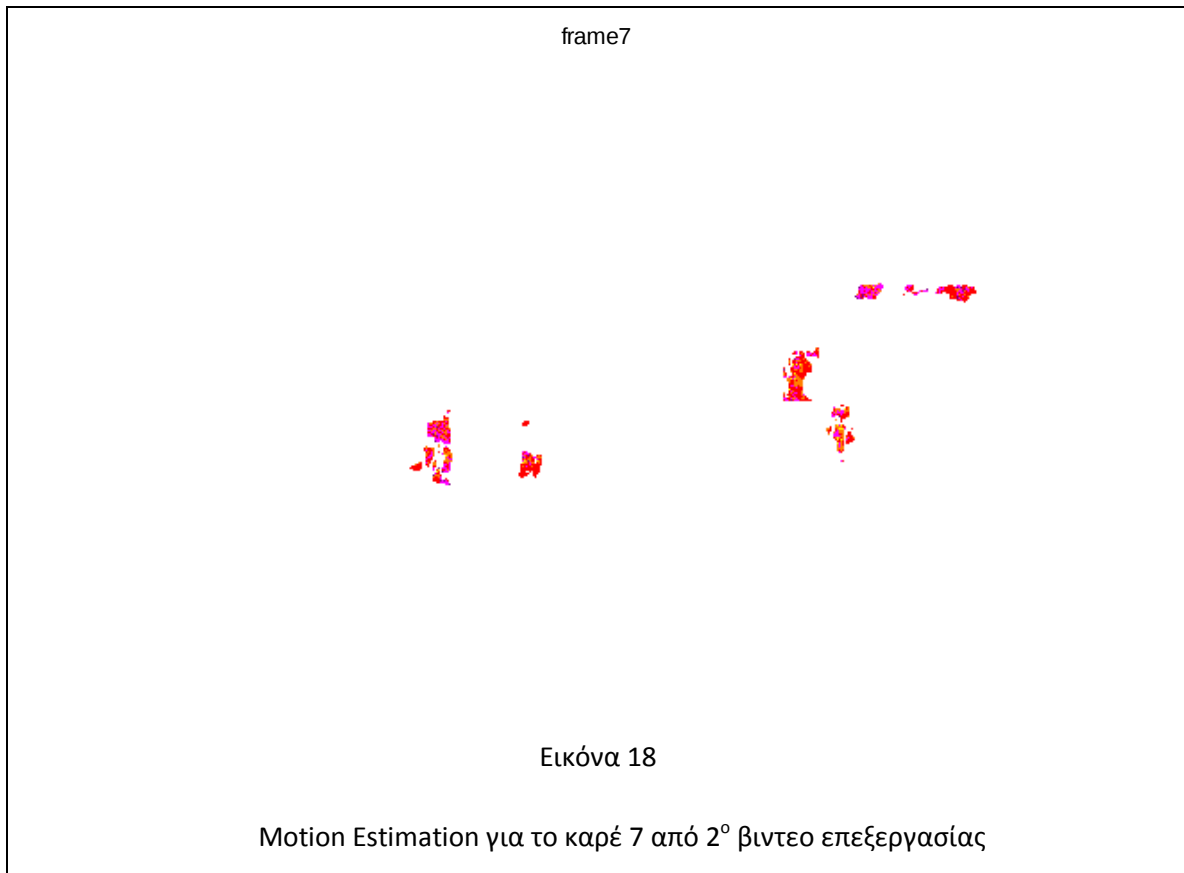
Motion Estimation για το καρέ 1078

Σε αυτό το καρέ παρατηρείται ότι ο πρώτος από αριστερά άνθρωπος έχει επικρατέστερη κατεύθυνση την κόκκινη ενώ ο δεύτερος την ανοιχτό μοβ προς κόκκινο. Όσον αφορά τη βαλίτσα παρατηρούμε ότι , όπως φαίνεται και στο παρακάτω καρέ, ο χρωματισμός της θα έπρεπε να είναι σκούρο μοβ αλλά επικρατεί το ανοιχτό μοβ. Αυτό συμβαίνει λόγω θορύβου, όπως αναλύεται και παρακάτω.



Ομοίως το ίδιο θα έπρεπε να ισχύει και στις παραπάνω εικόνες καθώς ο πρώτος άνθρωπος θα έπρεπε να είχε επικρατόν χρώμα το κόκκινο με κίτρινο, χρώματα που υπάρχουν αλλά δεν υπερισχύουν, η βαλίτσα θα έπρεπε να ήταν χρώματος φούξια ενώ ο δεύτερος κατά σειρά άνθρωπος θα έπρεπε να ήταν χρώματος σκούρο μοβ. Οι έντονες αυτές αποκλίσεις που εμφανίζονται και στο επόμενο βίντεο αλλά όχι τόσο έντονα όσο σε αυτό το βίντεο οφείλονται σε αρκετούς παράγοντες. Αρχικά η έντονη ανομοιομορφία των εντάσεων των εικονοστοιχείων δημιουργεί και ανομοιομορφία στις τιμές των διανυσμάτων κίνησης. Επίσης από το χρωματικό κύκλο παρατηρείται ότι

τα χρώματα θα έπρεπε να ήταν κάπως διαφορετικά, εννοώντας ότι από το κόκκινο θα έπρεπε να πηγαίνουμε στο φούξια και έπειτα στο ανοιχτό μοβ. Τέλος επειδή εργαζόμαστε για πολύ μικρές τιμές μήκους του διανύσματος κίνησης από καρέ σε καρέ η διεύθυνση του καθενός δείχνει σχεδόν κάθετη.

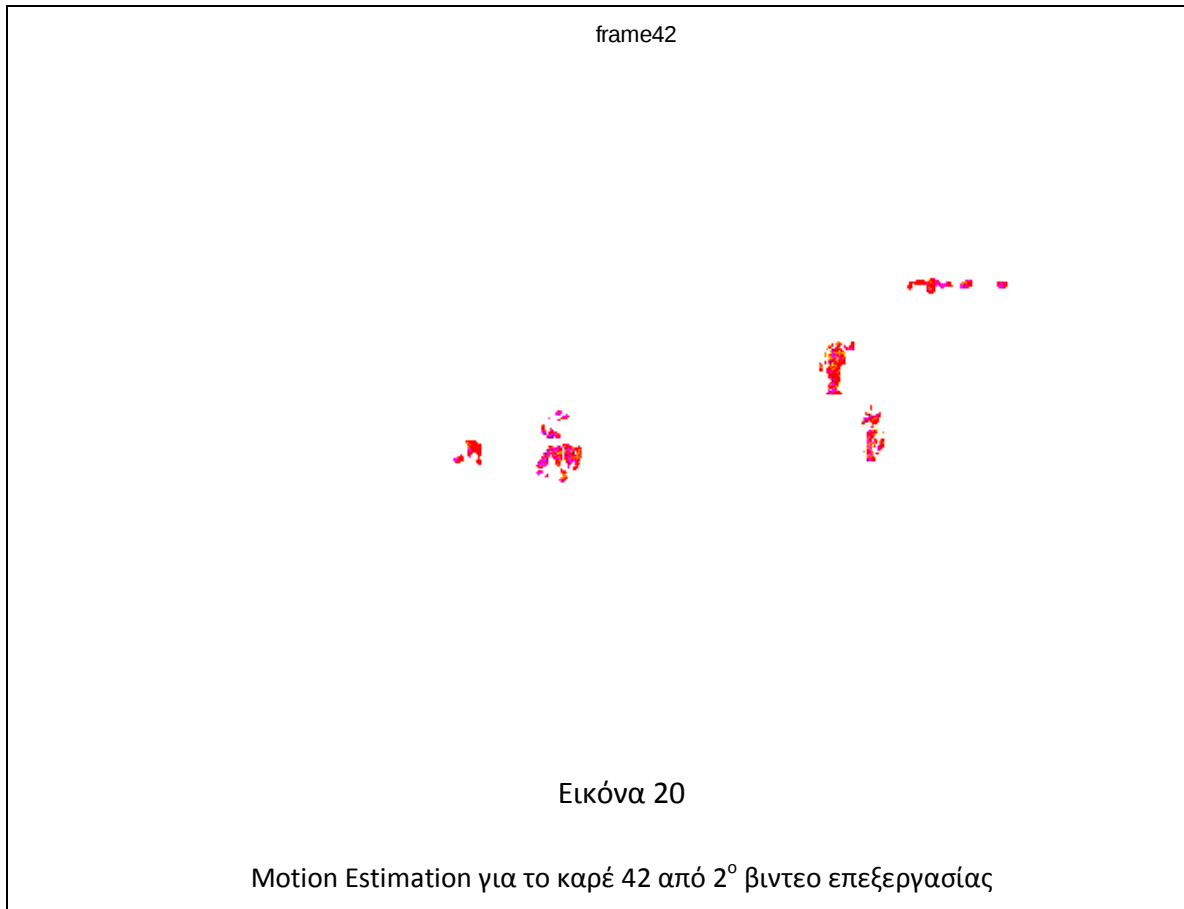


frame27



Εικόνα 19

Motion Estimation για το καρέ 27 από 2^ο βίντεο επεξεργασίας



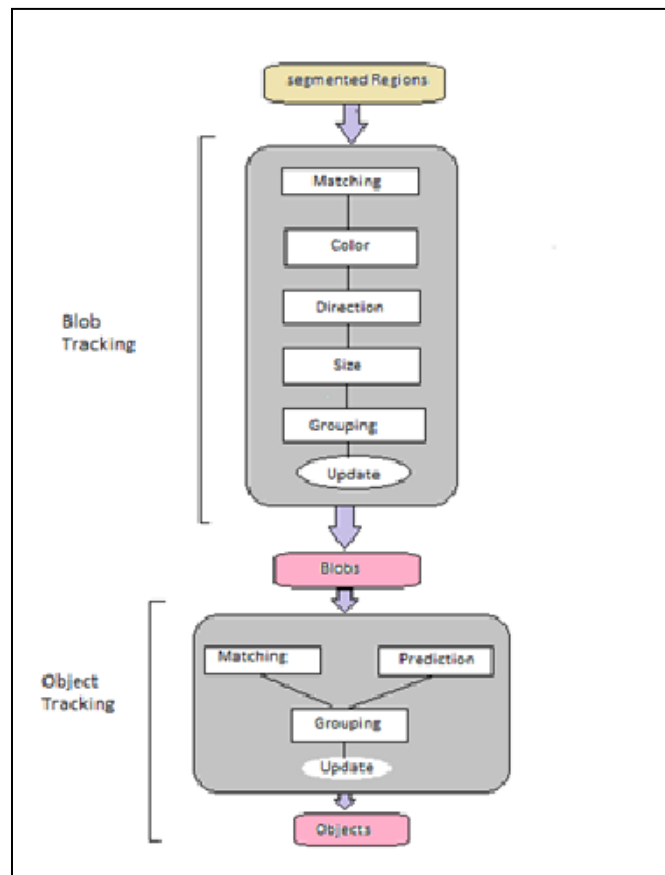
Γενικά από τα παραπάνω αποτελέσματα, αρχικά, παρατηρείται ότι κάθε pixel χρήσιμης πληροφορίας έχει πάρει κι ένα χρώμα, από αυτά του Color Wheel. Οι κατευθύνσεις οι οποίες επικρατούν για το δισδιάστατο χώρο στον οποίο κινούνται τα αντικείμενα, είναι οι 0° - 180° . Πρακτικά αυτό σημαίνει ότι τα αντικείμενα κινούνται σε κατευθύνσεις οριζόντια, κάθετα, δεξιά ή αριστερά. Βέβαια εξίσου εμφανής είναι και η ανομοιομορφία των χρωμάτων που οφείλεται σε δύο κατά βάση, παράγοντες. Ο ένας σαφώς και είναι ο θόρυβος, που έχει αναλυθεί σε προηγούμενο κεφάλαιο, ενώ ο δεύτερος οφείλεται στο γεγονός ότι τα άκρα των αντικειμένων, αφού τα περισσότερα αντικείμενα είναι άνθρωποι, δε συμφωνούν πάντα με το σώμα. Η ανομοιομορφία αυτή, όπως και τα υπόλοιπα προβλήματα θορύβου, θα αλλάξουν στο επόμενο κεφάλαιο, στο κεφάλαιο του Blob Tracking.

Βελτιώσεις σε αυτό το κομμάτι της εκτίμησης κίνησης μπορούν να γίνουν. Αρχικά πρέπει να δημιουργηθεί ένας νέος χρωματικός κώδικας σύμφωνα με τον οποίο να είναι πιο έντονες οι διαφοροποιήσεις κοντινών φορών διανυσμάτων κίνησης. Πρέπει να προστεθούν παραπάνω χρώματα έτσι ώστε να έχουμε μεγαλύτερη ακρίβεια. Μία εξίσου καλή βελτιστοποίηση του χρωματικού κύκλου είναι η χρήση της ακτίνας δηλαδή το να χρησιμοποιηθεί και το μέτρο του διανύσματος κίνησης κατά που σε αυτή την εργασία κρατήθηκε στο περιθώριο. Τέλος, ίσως θα μπορούσε να χρησιμοποιηθεί άλλη μέθοδος μείωσης του φωτισμού άρα και της ανομοιομορφίας των χρωμάτων, μέθοδος που θα ταιριάζει περισσότερο στο συνδυασμό της χρήσης του χρωματικού κύκλου με τη γενική μείωση του θορύβου.

Κεφάλαιο 4: Πολυεπίπεδος αλγόριθμος εντοπισμού κίνησης

Έως τώρα έχει γίνει επεξεργασία των καρέ σε επίπεδο pixels. Στη συνέχεια της εργασίας σκοπός είναι η ανάλυση του πολυεπίπεδου αλγορίθμου εντοπισμού (multi-level tracking algorithm), όπως φαίνεται στο παρακάτω σχήμα. Ο multi-level tracking αλγόριθμος μας μεταφέρει σε επίπεδο ομάδας εικονοστοιχείων πια και συνδέει τις δύο διαφορετικές μεθόδους με τις οποίες ασχοληθήκαμε, ουσιαστικά σε αυτήν την εργασία, την κατάτμηση κίνησης και τον εντοπισμό και την ταυτοποίηση των μαζών των αντικειμένων. Με τη βοήθεια αυτού του αλγορίθμου καταφέρνουμε να ομαδοποιήσουμε τα χρωματισμένα εικονοστοιχεία βάσει κριτηρίων και αρχικά να εντάξουμε σε ομάδες τις άμορφες μάζες που ανήκουν σε ένα αντικείμενο σε επίπεδο frame και έπειτα να συντάξουμε και να αναγνωρίσουμε τα αντικείμενα σε επίπεδο frame-by-frame. Ο multi-level tracking αλγόριθμος αποτελείται από τρία στάδια, το segmented regions, το blob tracking και εν συνεχεία το object tracking. Η βασική ιδέα του είναι να εξάγουμε σε κάθε frame τις επιμέρους περιοχές, στη συνέχεια να τις ομαδοποιήσουμε, σε επίπεδο frame ακόμη, και να δημιουργήσουμε τα blobs και τέλος, να χρησιμοποιήσουμε αυτήν την πληροφορία και δουλεύοντας σε επίπεδο frame-by-frame να ομαδοποιηθούν εκ νέου παίρνοντας το επιθυμητό αποτέλεσμα, δηλαδή την εξαγωγή των objects. Αναλυτικότερα, ο multi-level tracking αλγόριθμος έχει ως τελικό στόχο τον εντοπισμό των αντικειμένων. Αποτελείται από τρία τμήματα το segmented regions, δηλαδή εξαγωγή των τμημάτων των περιοχών που στη συνέχεια, στο δεύτερο επίπεδο του αλγορίθμου, το blob tracking, ομαδοποιούνται βάσει

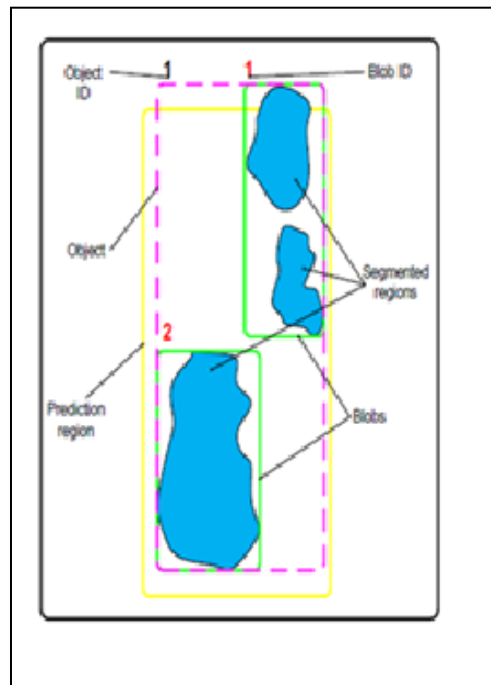
χαρακτηριστικών (features), έτσι ώστε στο τρίτο επίπεδο πια, του object tracking, να συνταχθούν βάσει κριτηρίων, δηλαδή να εντοπιστούν και να ταυτοποιηθούν τα αντικείμενα. Τα δύο πιο βασικά κομμάτια του multi-level tracking algorithm είναι οι αλγόριθμοι υψηλού επιπέδου Blob tracking algorithm και Object tracking algorithm, που θα αναλυθούν παρακάτω [2].



Σχήμα 10: Multi-Level Tracking Diagram

4.1 Διαχωρισμένες περιοχές

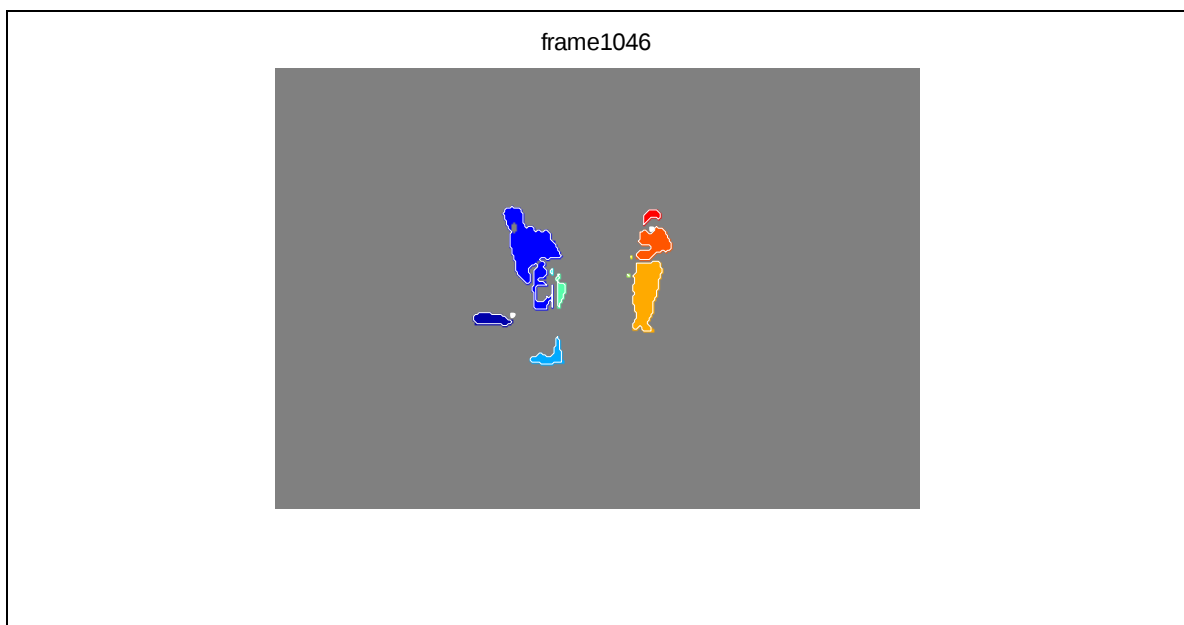
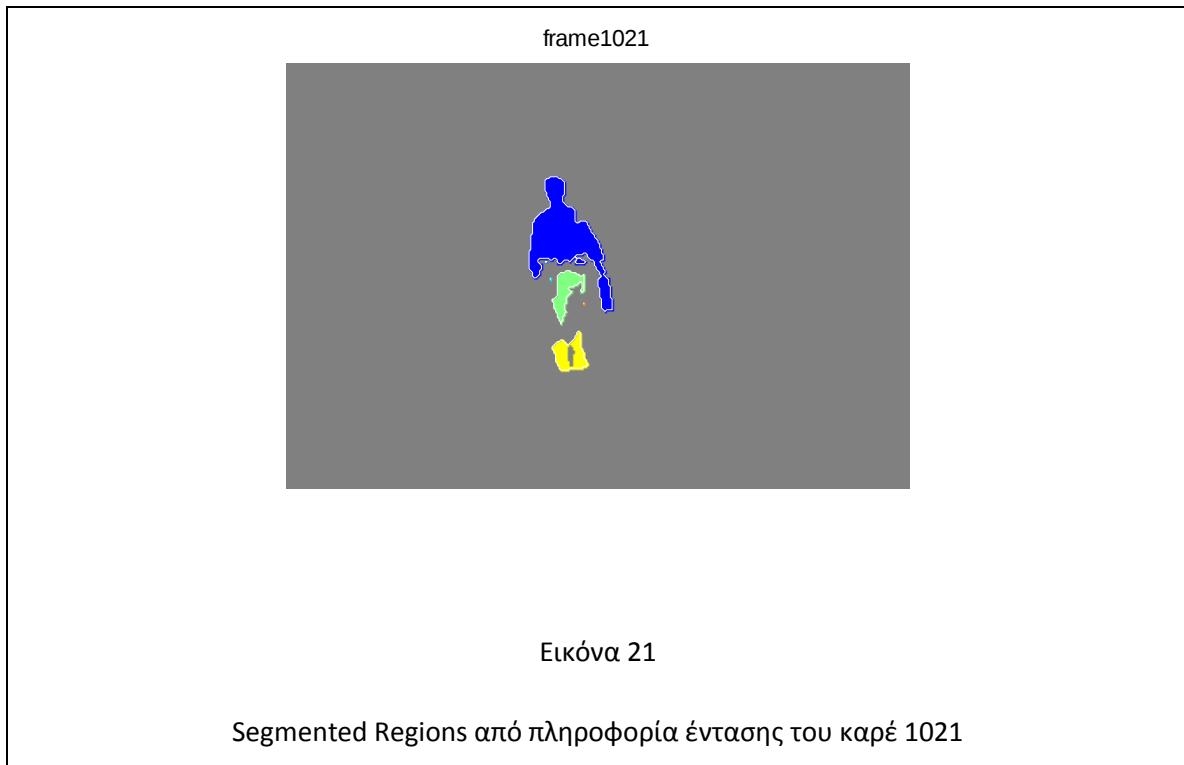
Ως αρχική διαδικασία ορίζεται η εξαγωγή των τμημάτων των περιοχών που ορίζουν τα αντικείμενα (segmented regions). Στο πρώτο επίπεδο του multi-level tracking αλγορίθμου εργαζόμαστε σε επίπεδο καρέ. Από κάθε καρέ σκοπός είναι να ορίσουμε τα «κομμάτια» χρήσιμης πληροφορίας, κομμάτια που στην πορεία θα ομαδοποιηθούν και θα οριστούν τα αντικείμενα. Η αρχική διαδικασία εξαγωγής έγινε με τη χρήση της *bwboundaries*, συνάρτησης της matlab. Η είσοδος λοιπόν, αποτελείται από το περίγραμμα των segmented regions, τα κέντρα τους καθώς και τον αριθμό των pixels από τα οποία αποτελείται καθεμία, όπως φαίνεται στο σχήμα 11 [2].



Σχήμα 11: Segmented Regions

Στο παραπάνω σχήμα αρχικά με «μπλε» χρώμα φαίνονται τα κομμάτια που συνθέτουν αργότερα, αφού ομαδοποιηθούν ένα blob, στο δεύτερο επίπεδο του αλγορίθμου. Τέλος, αφού ομαδοποιηθούν εκ νέου τα blobs, δηλαδή οι άμορφες μάζες

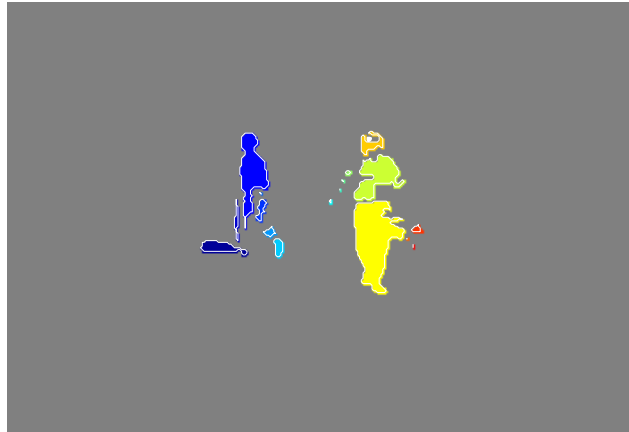
γίνεται εξαγωγή και ταυτοποίηση των αντικειμένων στο τρίτο επίπεδο του αλγορίθμου.
Παρακάτω φαίνονται τα αποτελέσματα της εξαγωγής των segmented regions.



Εικόνα 22

Segmented Regions από πληροφορία έντασης του καρέ 1046

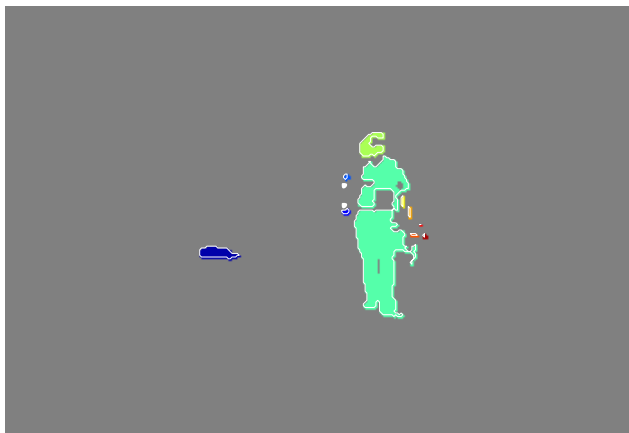
frame1078



Εικόνα 23

Segmented Regions από πληροφορία έντασης του καρέ 1078

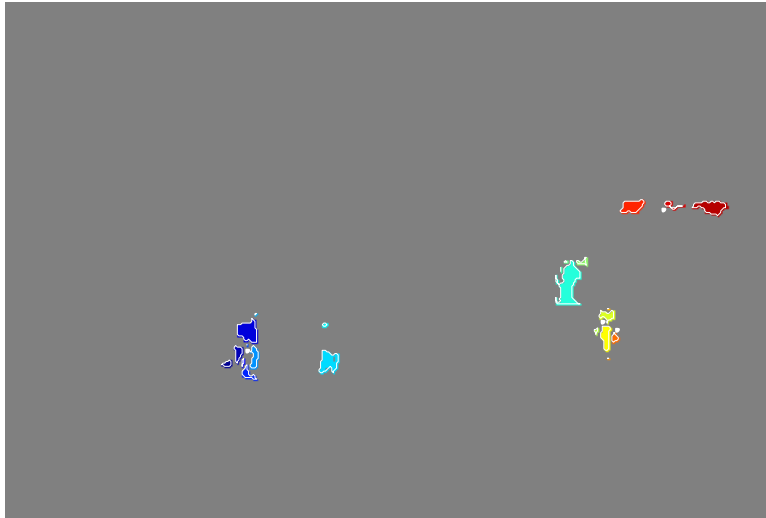
frame1087



Εικόνα 24

Segmented Regions από πληροφορία έντασης του καρέ 1087

frame7



Εικόνα 25

Segmented Regions από πληροφορία έντασης του καρέ 7 του 2^{ου} βίντεο

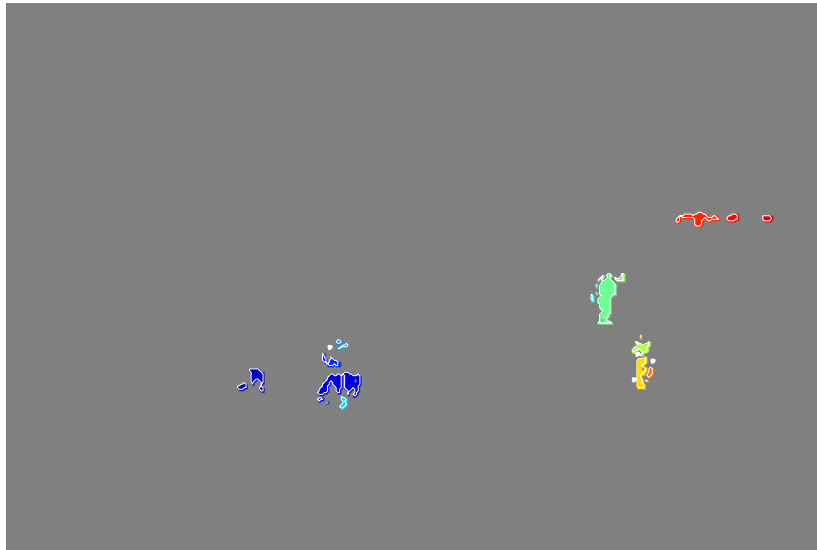
frame27



Εικόνα 26

Segmented Regions από πληροφορία έντασης του καρέ 27 του 2^{ου} βίντεο

frame42



Εικόνα 27

Segmented Regions από πληροφορία έντασης του καρέ 42 του 2^{ου} βίντεο

Από τα παραπάνω αποτελέσματα μπορεί κανείς να διακρίνει ότι αρχικά ένα αντικείμενο είναι χωρισμένο σε πολλές μικρές αλλά και μεγάλες περιοχές. Κάθε περιοχή βάσει της έτοιμης συνάρτησης `bwboundaries` έχει το δικό της χρώμα. Επίσης είναι εμφανής ο διαχωρισμός με ένα άσπρο περίγραμμα κάθε περιοχής.

4.1.1 Ομαδοποίηση κατευθύνσεων

Πριν προχωρήσουμε όμως στην ανάλυση του δεύτερου αλγορίθμου υψηλού επιπέδου, του blob tracking algorithm, όπως φαίνεται από το σχήμα 10, για την ομαδοποίηση των παραπάνω περιοχών χρειάζεται το χαρακτηριστικό χρώμα, που αναλύθηκε στο κεφάλαιο 3. Όμως, όπως είχε αναφερθεί λόγω του θορύβου, το χρώμα από τη motion estimation διαδικασία δεν μπορεί να οριστεί ως ένα έγκυρο χαρακτηριστικό ομαδοποίησης των segmented regions. Έτσι καθίσταται υποχρεωτικό να ομαδοποιήσουμε με κάποιο τρόπο τις κατευθύνσεις των motion vectors, δηλαδή να μπορεί να γίνει το χρώμα κάθε segmented region ένα επαρκές και έγκυρο χαρακτηριστικό ομαδοποίησης, δηλαδή εν ολίγοις να μπορέσει να ξεπεραστεί το πρόβλημα του θορύβου που υπήρχε. Η ανομοιομορφία λοιπόν του motion estimation μπόρεσε να λυθεί μόνο σε αυτό το στάδιο εξαγωγής των segmented regions. Το σκεπτικό επίλυσης αυτού του προβλήματος βασίστηκε στο γεγονός ότι κάθε segmented region πρέπει να έχει μία και μοναδική κατεύθυνση, δηλαδή ένα χρώμα που θα καθορίζει αυτό το κομμάτι χρήσιμης πληροφορίας προς τα πού κινείται. Έτσι από κάθε κομμάτι μετρήθηκε η επικρατέστερη κατεύθυνση η οποία ορίστηκε και ως η κατεύθυνση κάθε segmented region. Βάσει αυτού καταφέραμε να ορίσουμε το χρώμα ως ένα έγκυρο και σαφές χαρακτηριστικό ομαδοποίησης των segmented regions. Παρακάτω φαίνονται ενδεικτικά αποτελέσματα σε grayscale τιμές έντασης αυτής της επίλυσης.

frame1021



Εικόνα 28

Ομαδοποιημένες διευθύνσεις του καρέ 1021

Στο παραπάνω καρέ παρατηρείται ότι τα πόδια έχουν διαφορετική κατεύθυνση από ότι το υπόλοιπο σώμα. Αυτό συμβαίνει λόγω του ότι και η ανθρώπινη κίνηση στην πραγματικότητα έτσι είναι. Τα άκρα μας δε συμβαδίζουν πάντα με την κτεύθυνση του σώματός μας.

frame1046



Εικόνα 29

Ομαδοποιημένες διευθύνσεις του καρέ 1046

frame1078



Εικόνα 30

Ομαδοποιημένες διευθύνσεις του καρέ 1078

frame1087



Εικόνα 31

Ομαδοποιημένες διευθύνσεις του καρέ 1087

Από τα παραπάνω αποτελέσματα παρατηρείται πλέον η ομοιομορφία των κατευθύνσεων κάθε segmented region. Εκτός αυτού βέβαια παρατηρείται και η εξάλειψη μικρών περιοχών. Επειδή όπως έχει αναφερθεί σκοπός αυτής της εργασίας είναι ο εντοπισμός της μάζας των αντικειμένων , καταφέραμε να «διώξουμε» περιοχές μικρού μεγέθους (κάτω από 20 εικονοστοιχεία περίγραμμα, τιμή που υπολογίστηκε πρακτικά, υπολογίζοντας το μέσο όρο των μεγεθών των περιγραμμάτων των segmented regions), περιοχές που στην ουσία δε δημιουργούν κάποιο πρόβλημα στον μετέπειτα εντοπισμό των αντικειμένων. Ο λόγος που εξαλείφθηκαν αυτές οι μικρές περιοχές είναι ότι όχι πως ήταν μη χρήσιμη πληροφορία αλλά δημιουργούσαν πρόβλημα στην μετέπειτα επεξεργασία της μάζας.

4.2 Αλγόριθμος εντοπισμού άμορφης μάζας

Ο δεύτερος αλγόριθμος, υψηλού επιπέδου, που κρίθηκε απαραίτητο να υλοποιηθεί για τον εντοπισμό των αντικειμένων είναι ο blob tracking αλγόριθμος, δηλαδή ο αλγόριθμος εντοπισμού της άμορφης μάζας. Όπως παρατηρείται στο σχήμα 11, στόχος του αλγορίθμου είναι η ομαδοποίηση των segmented regions που έχουν εντοπιστεί για τη δημιουργία των blobs. Η διαδικασία του blob tracking βασίζεται στο γεγονός ότι, segmented regions που έχουν εντοπιστεί σε κάθε καρέ θα έχουν κάποια κοινά στοιχεία, έτσι ώστε να θεωρηθεί ότι ανήκουν στο ίδιο blob. Τα χαρακτηριστικά ενός blob είναι αρκετά. Μερικά από αυτά είναι η περίμετρος, το μέγεθος, το χρώμα κ.α. Ξεκινώντας λοιπόν, εξετάζοντας μία περιοχή, έπρεπε να συδυαστούν οι πιθανές περιοχές και να ενταχθούν στην ίδια ομάδα άλλες περιοχές που πληρούσαν κάποιες

προϋποθέσεις που σημαίνει ότι είχαν και τα τρία από τα παρακάτω κριτήρια ομαδοποίησης. Ο αλγόριθμος, παύει να λειτουργεί και εξάγει μία ομάδα, δηλαδή ένα blob, αφού εξεταστούν όλοι οι πιθανοί συνδυασμοί segmented regions και δεν υπάρχουν πια άλλες περιοχές που να έχουν όμοια χαρακτηριστικά με αυτήν.

Κριτήρια Ομαδοποίησης Blob Tracking:

1. Color - Direction
2. Segmented Region's Size (Perimeter)
3. Segmented Region's Distance from the current segmented region

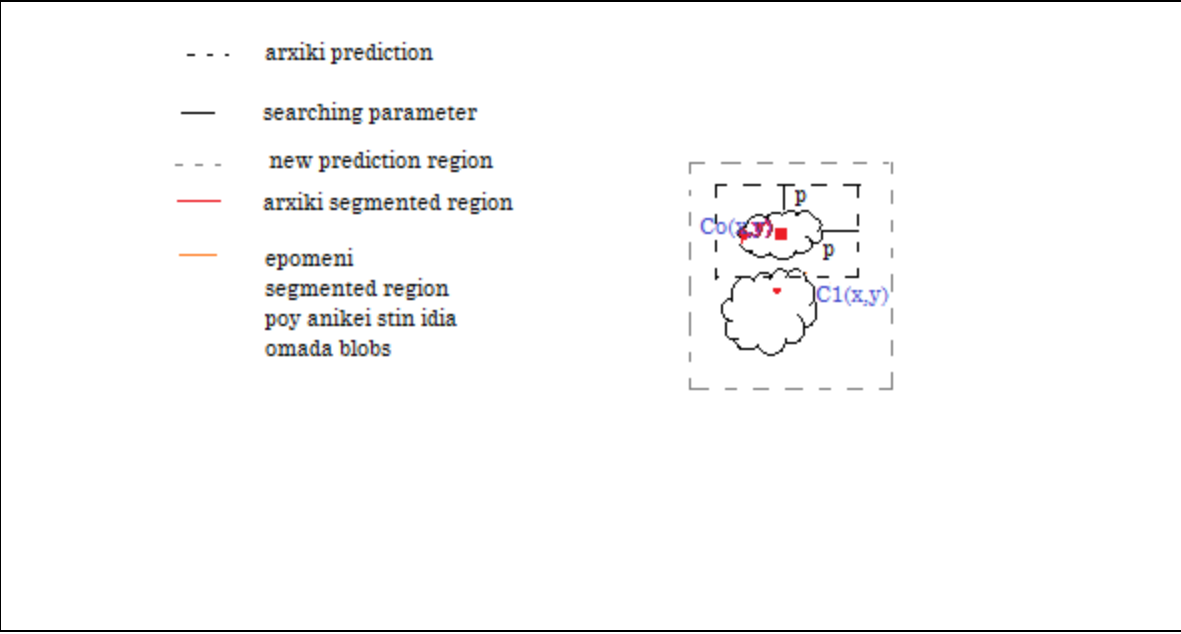
Το πρώτο κριτήριο είναι το χρώμα κάθε segmented region, δηλαδή το direction. Από το στάδιο του motion estimation έχει γίνει η όριση του direction κάθε pixel χρήσιμης πληροφορίας. Έτσι, για κάθε segmented region έχει γίνει έλεγχος για το ποια είναι η επικρατέστερη direction, ο καθορισμός της ως της κατεύθυνσης της εξεταζόμενης περιοχής και στη συνέχεια, θεωρήθηκε ότι περιοχές με ίσα directions, ή έστω πολύ κοντινά θα ανήκαν στο ίδιο blob, δηλαδή κομμάτια που ανήκαν στο ίδιο blob θα κινούνταν προς την ίδια κατεύθυνση. Το δεύτερο και τρίτο κριτήριο ομαδοποίησης ήταν ένας συνδυασμός της λογικής ότι, κομμάτια που απέιχαν μικρή απόσταση από την εξεταζόμενη περιοχή και παράλληλα είχαν μικρό μέγεθος, θα ανήκαν στην ίδια ομάδα, δηλαδή στο ίδιο blob. Αναλυτικότερα, ας δούμε πως λειτουργεί ο blob tracking αλγόριθμος.

Αρχικά λοιπόν εξετάζεται μία segmented region. Για κάθε περιοχή που εξετάζεται ορίζεται μια prediction region, όπως φαίνεται στο σχήμα 11. Η περιοχή αυτή περιλαμβάνει όλη τη segmented region. Στη συνέχεια, κάθε νέα segmented region που βρίσκεται εντός αυτής της περιοχής και έχει και τα τρία από τα παραπάνω κριτήρια ομοιότητας εντάσσεται στην ίδια ομάδα blob. Για κάθε νέα segmented region που εισάγεται σε ένα blob ορίζεται εκ νέου μία prediction region. Κάθε, τώρα, πάλι νέα

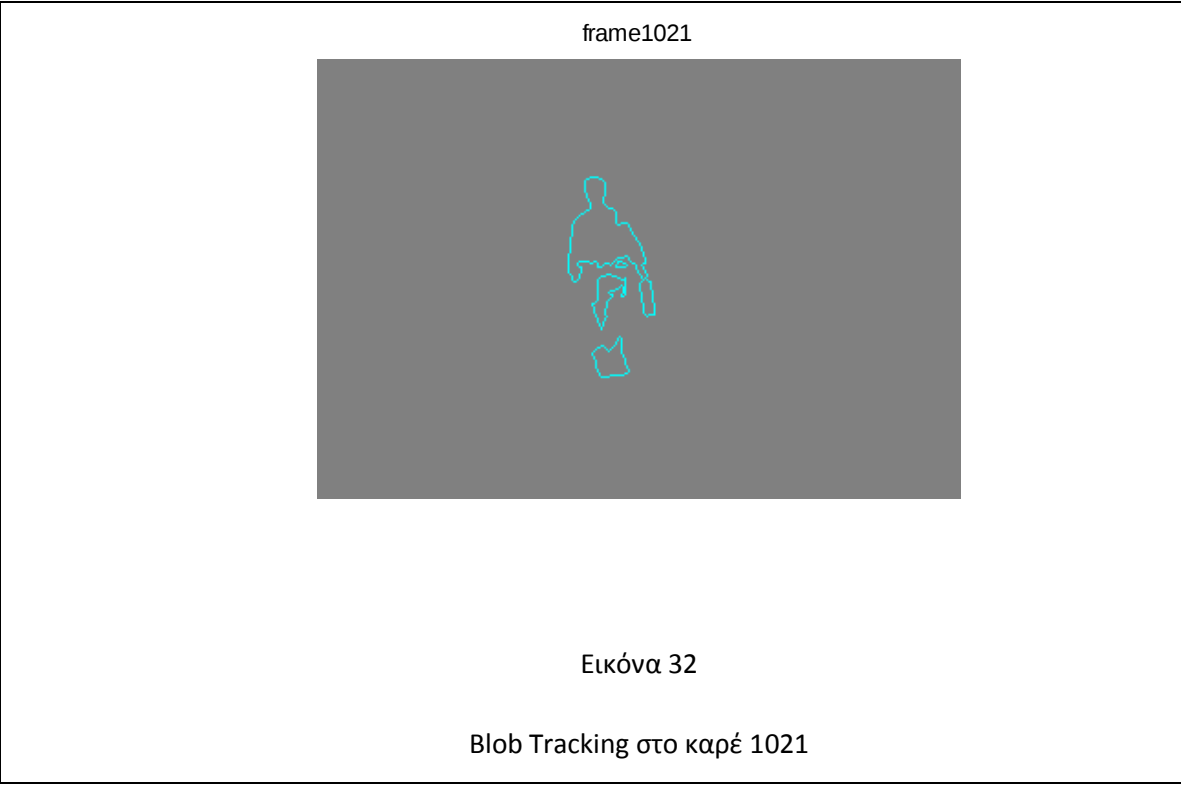
segmented region που ανήκει σε αυτή εντάσσεται στην ίδια ομάδα. Η prediction region άρα και κάθε ομάδα blob σταματάει να μεγαλώνει μόλις εξεταστούν όλες περιοχές που πιθανόν να ανήκουν μέσα στην ίδια ομάδα. Αφού ολοκληρωθεί μία ομάδα για μία νέα segmented region εξάγεται το blob και ορίζεται μία νέα prediction region κ.τ.λ. Παρακάτω εμφανίζονται ενδεικτικά αποτελέσματα των βίντεο που επεξεργάστηκαν.

4.2.1 Υπολογισμός πρόβλεψης περιοχής

Ας αναλύσουμε όμως τη μεθοδολογία δημιουργίας της prediction region, δηλαδή της πρόβλεψης της περιοχής μέσα στην οποία ορίζουμε να βρίσκεται κάθε διαχωρισμένη περιοχή. Prediction region θεωρείται η περιοχή που περικλείει κάθε segmented region. Όπως παρατηρείται στο παρακάτω σχήμα 12, αρχικά ορίζεται μία περιοχή για μία segmented region. Γι αυτήν την περιοχή γνωρίζουμε το κέντρο της έτσι από κάθε κέντρο υπολογίζουμε τη max απόσταση από τα πάνω και κάτω όρια της segmented region και βάζοντας και έναν searching parameter p , ορίζουμε μία περιοχή. Τώρα, περιοχές που βρίσκονται εντός της αρχικής prediction region και έχουν κοντινές ή ίδιες διευθύνσεις (βλ. κοντινά χρώματα στο χρωματικό κύκλο στο σχήμα 12), έχουν κοντινές αποστάσεις και κοντινά μεγέθη περιμέτρου ανήκουν στην ίδια ομάδα blob. Αφού λοιπόν κάποια νέα περιοχή εισέρχεται στην ίδια ομάδα blob, ορίζεται εκ νέου το κέντρο βάσει του τύπου $C1((x_0+x_1)/2, (y_0+y_1)/2)$, έτσι δημιουργείται μία νέα prediction region. Όπως αναφέρθηκε αναλυτικά παραπάνω, μία prediction region σταματάει να μεγαλώνει, δηλαδή ουσιαστικά ορίζεται μία και μοναδική ομάδα blob, όταν δεν υπάρχουν πια άλλες περιοχές που να πληρούν τα παραπάνω τρία κριτήρια ομαδοποίησης segmented regions.



Σχήμα 12: Εικονική δημιουργία prediction region



frame1046

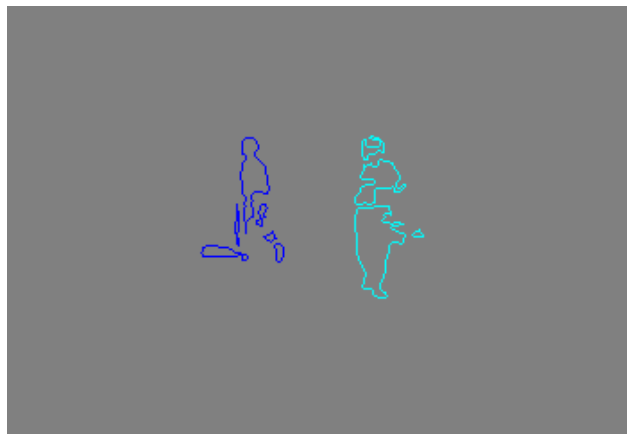


Εικόνα 33

Blob Tracking στο καρέ 1046

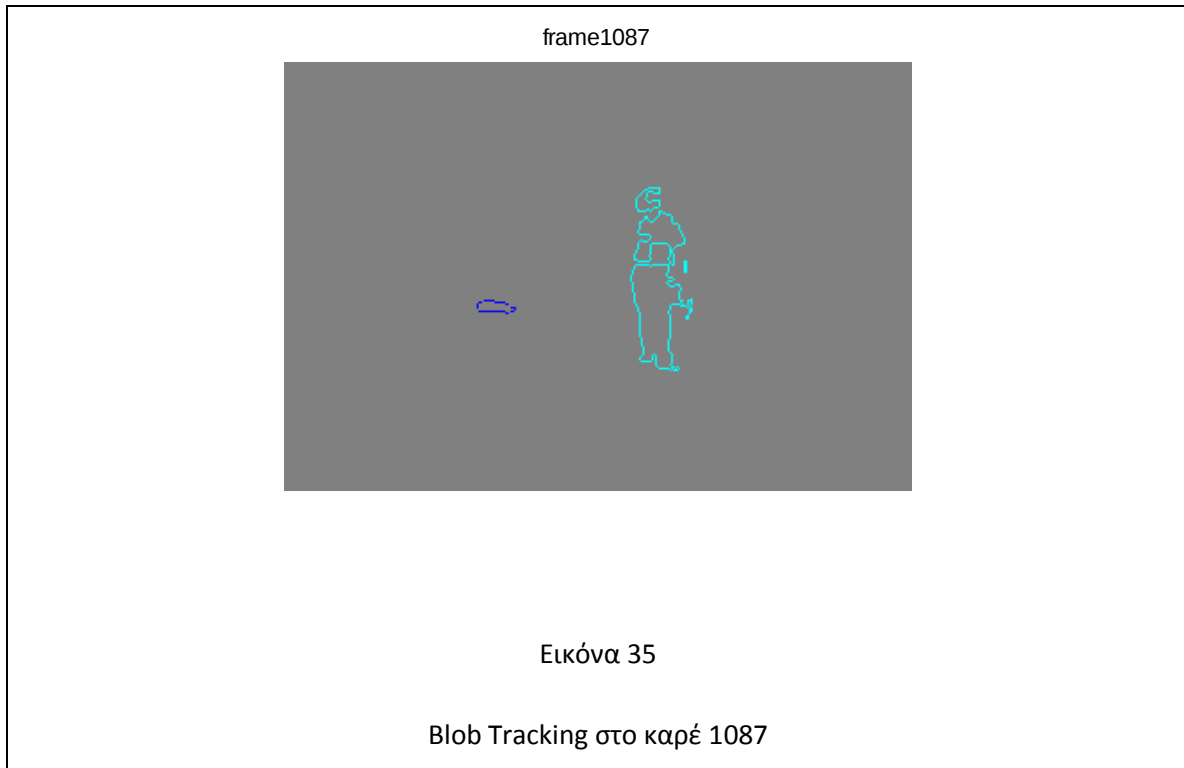
Συνδυάζει πληροφορία έντασης και κίνησης σε μία μάζα με όρια τις ακμές της.

frame1078



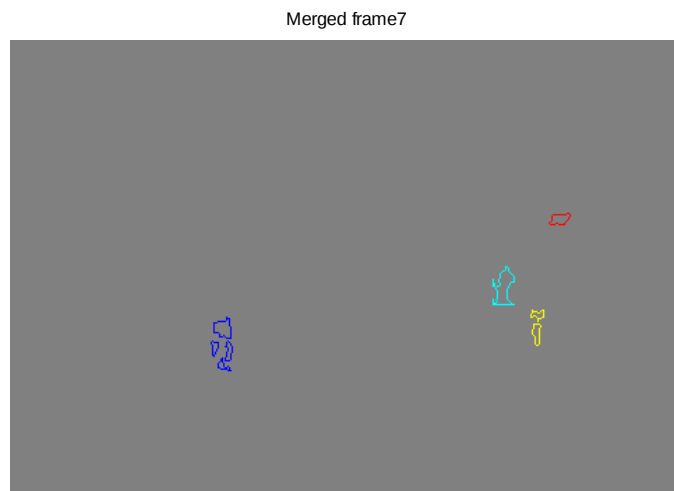
Εικόνα 34

Blob Tracking στο καρέ 1078



Όπως παρατηρείται στις εικόνες 32-35 τα αποτελέσματα είναι γενικά ικανοποιητικά. Segmented regions που ανήκουν στο ίδιο blob έχουν πάρει το ίδιο χρώμα. Παρατηρούμε ότι έχουμε περάσει από χρωματισμένες περιοχές σε χρωματισμένα περιγράμματα. Αυτή η επιλογή έγινε διότι ακόμη κι έτσι ο τελικός σκοπός που είναι ο εντοπισμός των μαζών των αντικειμένων εκπληρώνεται. Η εσωτερική περιοχή δε μας ενδιαφέρει αλλά ουσιαστικά το μόνο που κρατά το ενδιαφέρον είναι το περίγραμμα κάθε αντικειμένου. Έτσι από δω κι έπειτα λειτουργούμε βάσει περιγράμματος και μόνο. Στη συνέχεια να σημειωθεί ότι παρουσιάστηκαν κάποια προβλήματα όσον αφορά κομμάτια που ανήκαν σε άλλη ομάδα και ενώθηκαν με άλλη ομάδα. Πιο συγκεκριμένα όπως παρατηρείται στις εικόνες 33, 34 υπάρχουν τα εξής προβλήματα. Από τη στιγμή που το βίντεο είναι σε non-real time μπορούμε να γνωρίζουμε ότι η βαλίτσα για πρώτη φορά εντοπίζεται στο καρέ 1046, εικόνα 33. Παρατηρούμε όμως ότι η βαλίτσα δε διαχωρίζεται από το αντικείμενο το οποίο την αφήνει καθώς λόγω θορύβου η επικρατέστερη κατεύθυνση

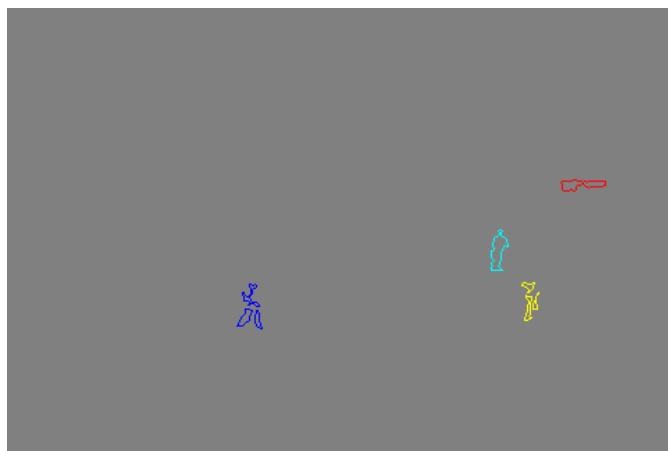
για τη βαλίτσα δεν είναι οι 0, δηλαδή σταθερό αντικείμενο, αλλά η βαλίτσα έχει την ίδια κατεύθυνση με τον άνθρωπο όπως μπορούμε να δούμε στην εικόνα 29. Αυτή η ενοποίηση των δύο αντικειμένων συμβαίνει γιατί και η prediction region της βαλίτσας αλλά και του ανθρώπου που την αφήνει επικαλύπτονται. Αυτό είναι ένα πρόβλημα επιλύθηκε παρακάτω, στο τρίτο στάδιο του multi-level tracking αλγορίθμου, δηλαδή με τον object tracking αλγόριθμο.



Εικόνα 36

Blob Tracking για το καρέ 7 του 2^{ου} βίντεο

Merged frame27



Εικόνα 37

Blob Tracking του καρέ 27 του 2^{ου} βίντεο

Merged frame42



Εικόνα 38

Στα παραπάνω αποτελέσματα παρατηρούμε ότι το blob tracking είναι άρτιο καθώς το μετέπειτα αντικείμενο είναι πλήρως εντοπισμένο και λόγω της περίπτωση του βίντεο που δεν υπάρχει κάποια επικάλυψη δεν παρουσιάζεται κάποιο πρόβλημα.

4.3 Αλγόριθμος εντοπισμού αντικειμένων

Ο τρίτος και τελευταίος αλγόριθμος υψηλού επιπέδου που υλοποιήθηκε για τον εντοπισμό και την ταυτοποίηση των αντικειμένων ήταν ο object tracking αλγόριθμος, δηλαδή ο αλγόριθμος ανίχνευσης αντικειμένων. Ο object tracking αλγόριθμος λειτουργεί σε επίπεδο frame-by-frame. Όπως παρατηρείται από το σχήμα 11, αφού έχει γίνει το «ταίριασμα» δηλαδή η ομαδοποίηση των segmented regions σε επίπεδο frame, σκοπός είναι να γίνει η «ένωση» των δεδομένων, blobs, για την εξαγωγή των αντικειμένων. Ο αλγόριθμος αυτός λειτουργεί σε δύο επίπεδα, δηλαδή προσπαθεί σε ένα καρέ να εντοπίσει τα αντικείμενα και στη συνέχεια να ταυτοποιήσει τα όμοια αντικείμενα από καρέ σε καρέ. Αναλυτικότερα, στο τελευταίο στάδιο του αλγορίθμου, στο στάδιο του object tracking, χρησιμοποιούνται τα δεδομένα που είχαμε ως έξοδο από το προηγούμενο στάδιο του multi-level tracking αλγορίθμου, δηλαδή τα blobs. Σε κάθε καρέ ενοποιούνται τα blobs που ανήκουν στην ίδια περιοχή (prediction region) διαδικασία που μοιάζει με τη διαδικασία ομαδοποίησης των segmented regions. Στη συνέχεια, βάσει των παρακάτω κριτηρίων, τα αντικείμενα πια, ταυτοποιούνται δηλαδή σε κάθε όμοιο αντικείμενο από καρέ σε καρέ δίνεται μία και μοναδική ταυτότητα από τη στιγμή που εμφανίζεται στο δισδιάστατο χώρο μέχρι που εξαφανίζεται από αυτόν. Ας υποθέσουμε ότι a, b είναι τα blobs των frame $t, t+1$ αντίστοιχα. Τότε υπολογίζοντας

τις παρακάτω παραμέτρους αντιστοιχίζουμε τα ομαδοποιημένα blobs από καρέ σε καρέ ως εξής:

Κριτήρια Object Tracking

1. $R_{\max}(a,b) = \max(\frac{SC(a,b)}{RI(a)}, \frac{SC(a,b)}{RI(b)})$
2. $R_{\min}(a,b) = \min(\frac{SC(a,b)}{RI(a)}, \frac{SC(a,b)}{RI(b)})$
3. $S(a,b) = 1 - [\frac{Abs(SRa-SRb)}{\max(SRa,SRb)}]$

Όπου:

R: είναι η επικάλυψη

SC: η τομή των ROI

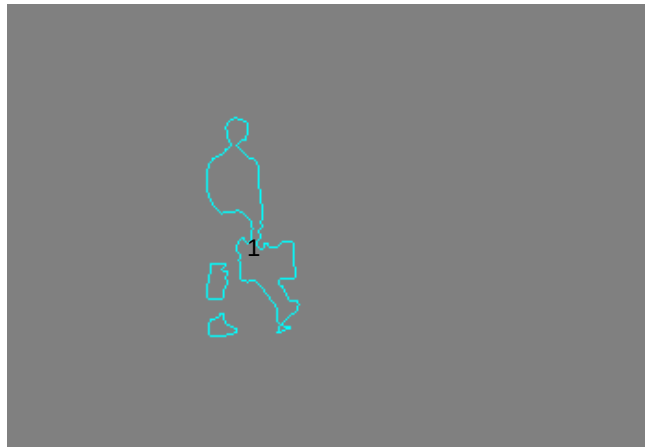
RI(i): η area του i^{th} blob's ROI

S: similarity

SR_i: the actual area of i^{th} blob

Με κύριο κριτήριο ταυτοποίησης την ομοιότητα S, και συνδυάζοντας τις άλλες παραμέτρους για να μειωθεί η περίπτωση λάθους, γίνεται αντιστοίχιση των objects των διαδοχικών frames. Όσον αφορά την ταυτοποίησή τους, ένας πιθανός τρόπος ήταν η δημιουργία μίας prediction region/window για κάθε object κάθε frame. Για αυτά τα δείγματα videos επιλέχθηκε απλά η απαρίθμηση των αντικειμένων. Έτσι, για κάθε object που εμφανίζεται στα videos, δόθηκε ένα μοναδικό ID το οποίο το ορίζει από τη στιγμή που εμφανίζεται έως και τη στιγμή που εξαφανίζεται από το video. Παρακάτω εμφανίζονται ενδεικτικά αποτελέσματα των βίντεο που επεξεργάστηκαν.

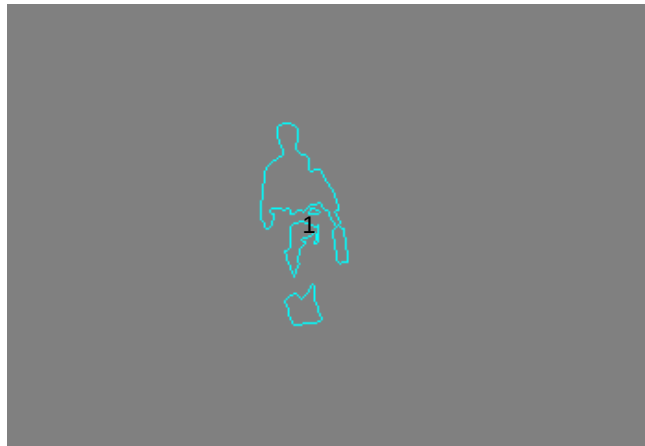
frame1014



Εικόνα 39

Object Tracking για το καρέ 1014

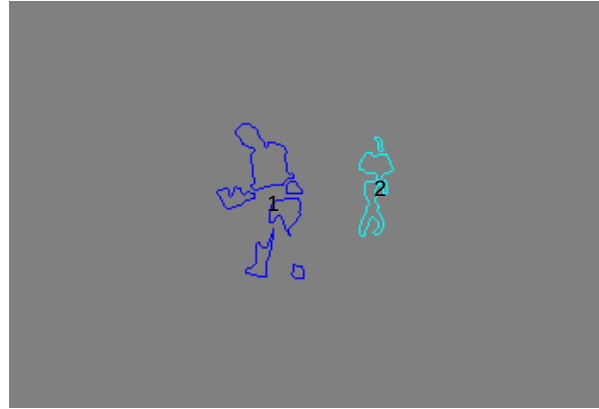
frame1021



Εικόνα 40

Object Tracking για το καρέ 1021

frame1032



Εικόνα 41

Object Tracking για το καρέ 1032

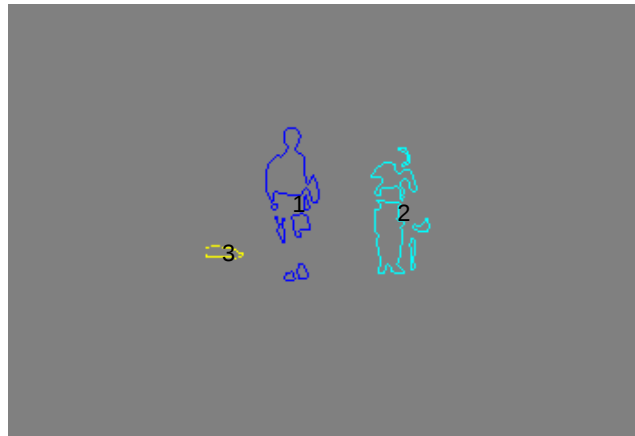
frame1046



Εικόνα 42

Object Tracking για το καρέ 1046

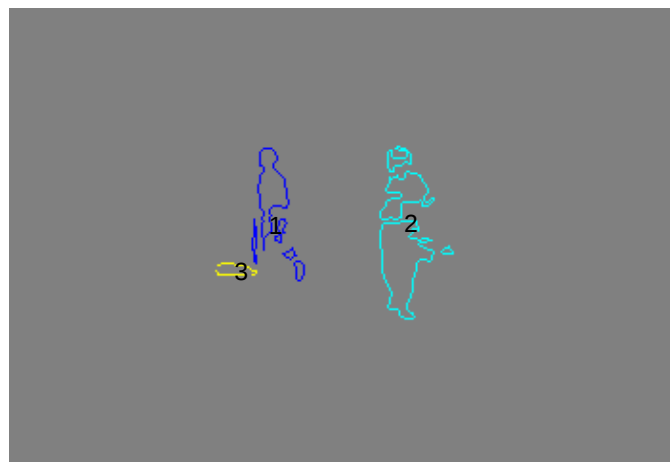
frame1058



Εικόνα 43

Object Tracking για το καρέ 1058

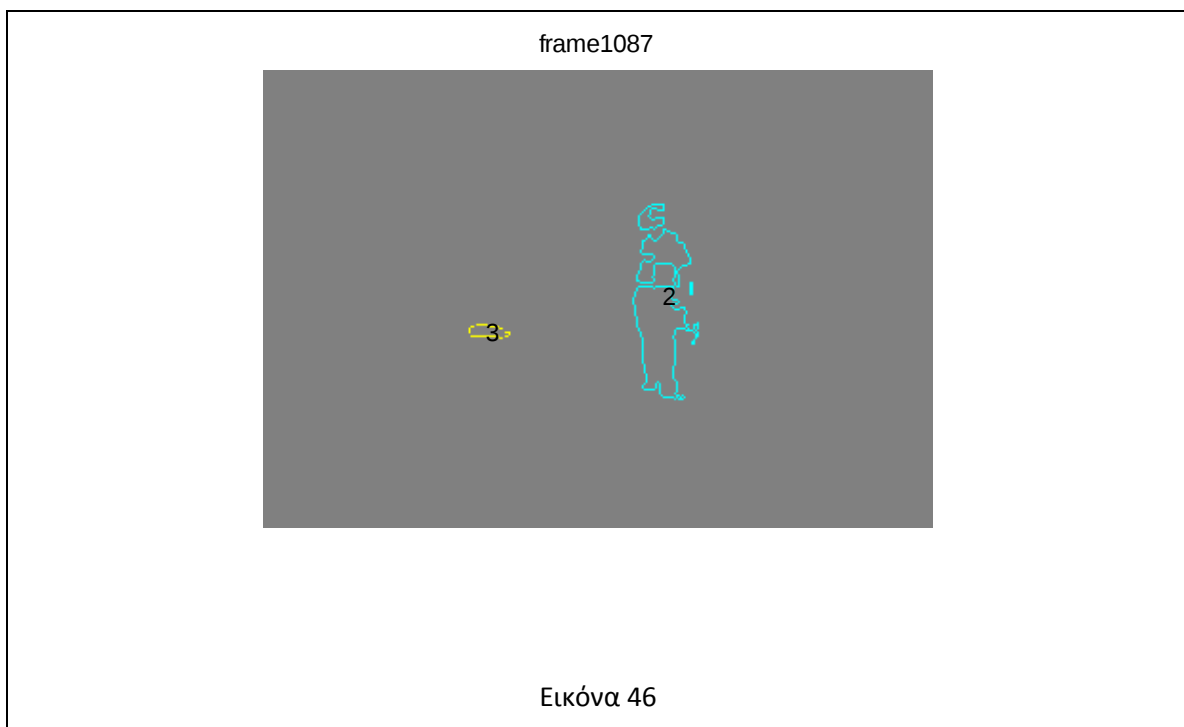
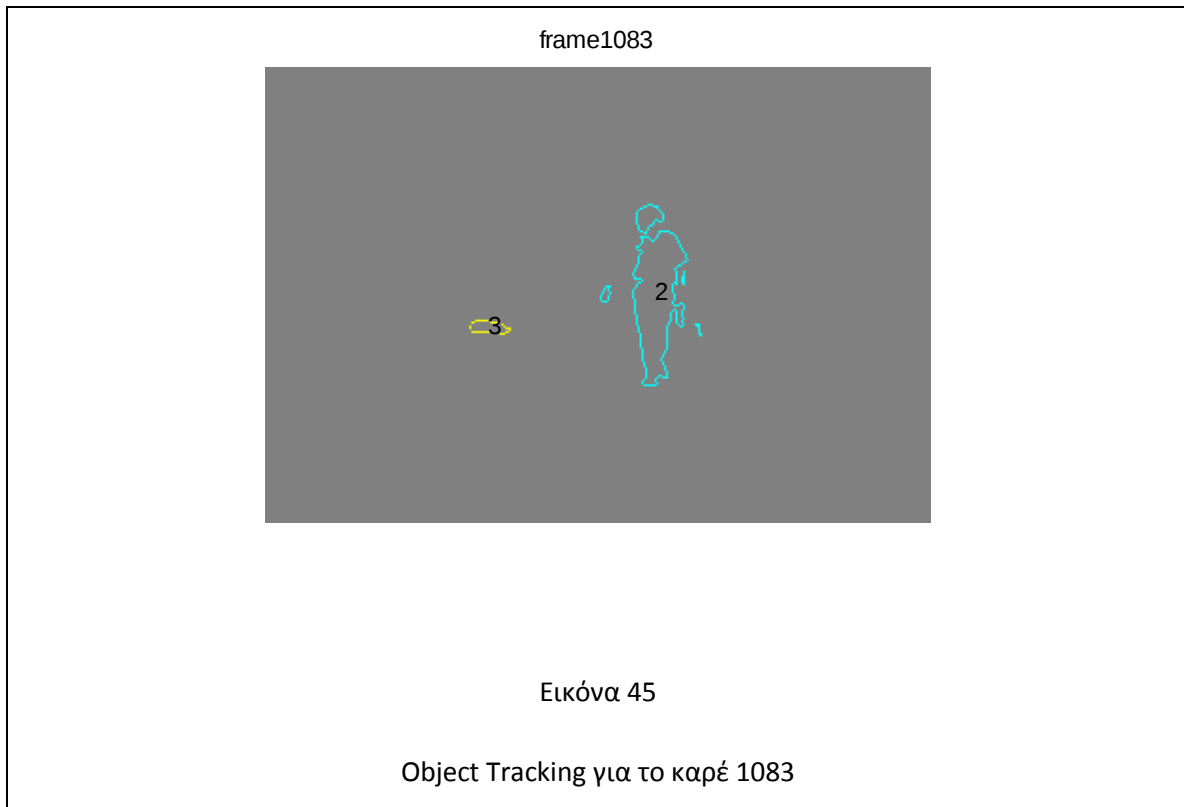
frame1078



Εικόνα 44

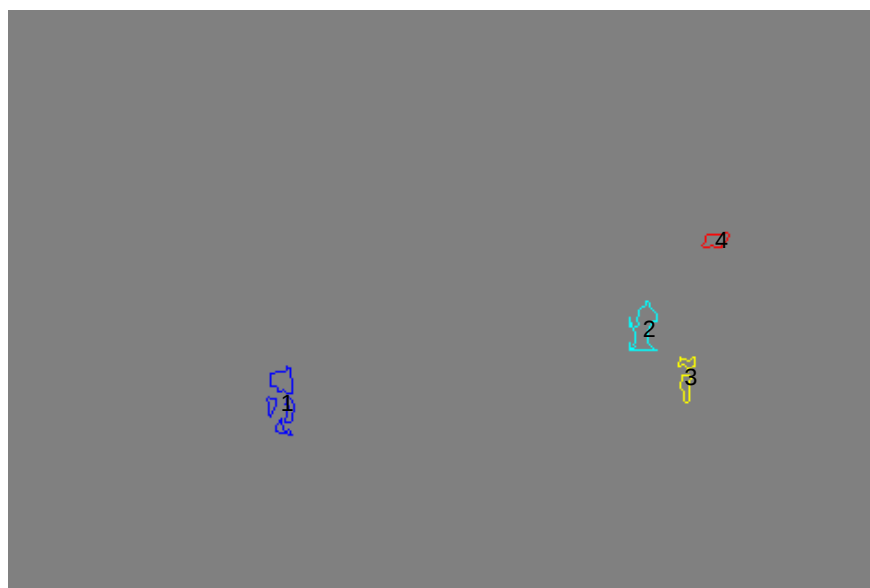
Object Tracking για το καρέ 1078

Σύμφωνα με το σχήμα 11 παρατηρούμε ότι κάθε blob έχει ένα και μοναδικό ID.



Όπως έχει αναφερθεί πιο πάνω, το βασικό πρόβλημα που συναντήσαμε κατά τον εντοπισμό του object tracking ήταν ο εντοπισμός της βαλίτσας, δηλαδή κατ'επέκταση ο διαχωρισμός του αντικειμένου από άλλα αντικείμενα και στη συνέχεια η όριση του ως σταθερού αντικειμένου. Η λύση σε αυτό το πρόβλημα δόθηκε με την παρατήρηση , αρχικά, ότι το σταθερό αυτό αντικείμενο εντοπίζεται για πρώτη φορά στο frame 1046. Χρησιμοποιώντας αυτό το δεδομένο (έπειτα από αρκετές δοκιμές που έγιναν) κρίθηκε ότι ένας εποικοδομητικός τρόπος διαχωρισμού ήταν να συνδυάσουμε τα αποτελέσματα του segmented region level με τα αποτελέσματα του blob tracking level. Για κάθε frame επομένως, μετά το frame 1046 προηγούταν μία διαδικασία διαχωρισμού της βαλίτσας από τον άνθρωπο. Με αυτόν τον τρόπο ουσιαστικά «στέλνονταν» ανανεωμένα τα blobs ως είσοδο στο object tracking level. Τέλος, να επισημάνουμε ότι τα labels που καθορίζουν το ID κάθε αντικειμένου, είναι η αντίστοιχη τιμή που έχει το αντικείμενο ως κωδικοποίηση στην εικόνα μας. Αυτό πρέπει να σημειωθεί καθώς τα χρώματα που βλέπουμε δεν είναι πάντα ενδεικτικά για τον καθορισμό όμοιων αντικειμένων κατά τη ροή του video.

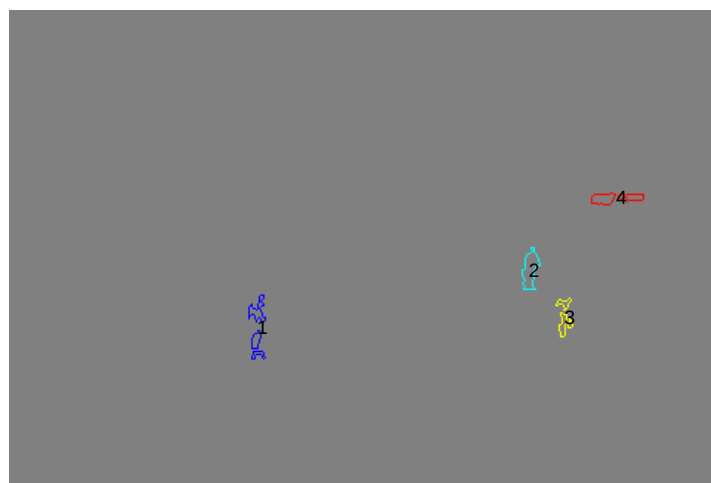
frame7



Εικόνα 47

Object Tracking για το καρέ 7

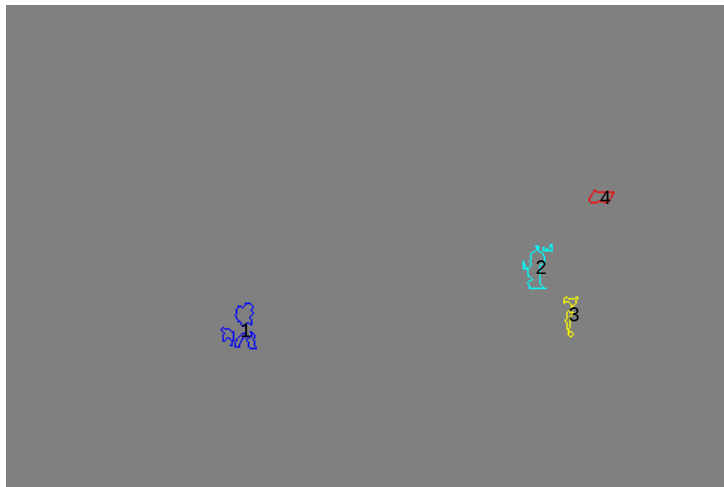
frame21



Εικόνα 48

Object Tracking για το καρέ 21

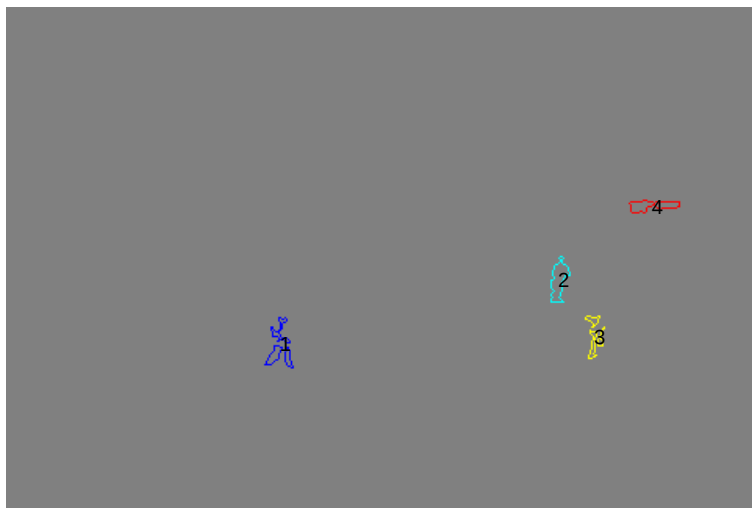
frame13



Εικόνα 49

Object Tracking για το καρέ 13

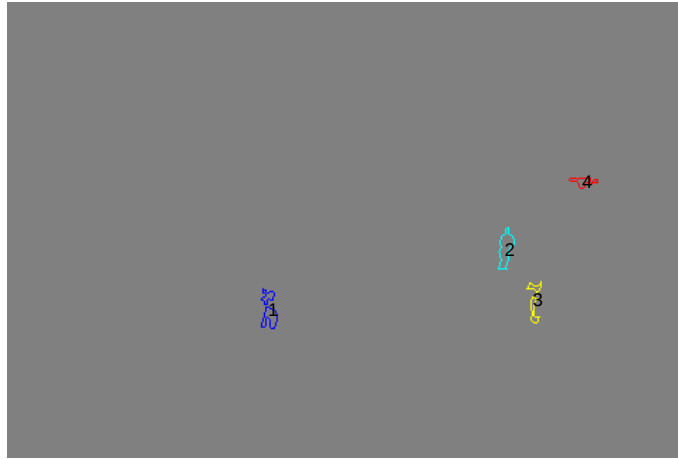
frame27



Εικόνα 50

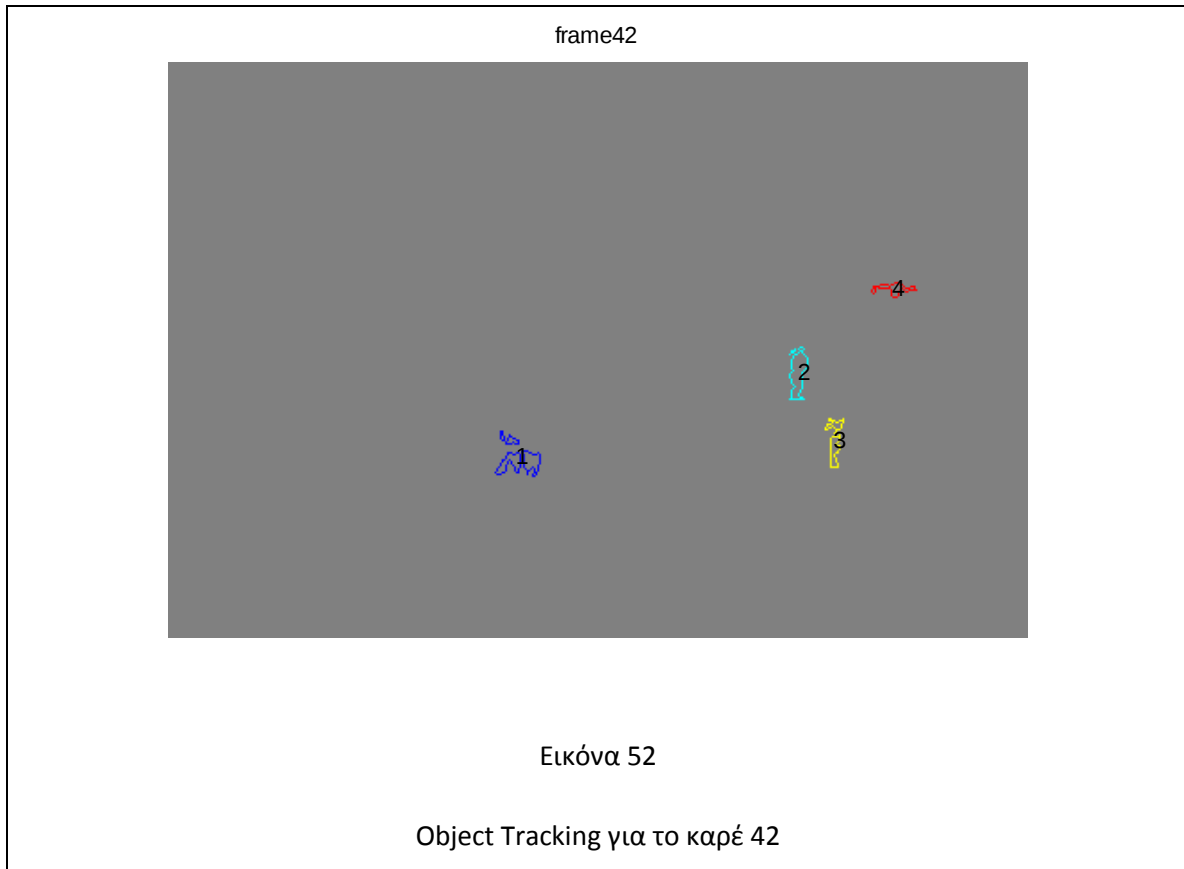
Object Tracking για το καρέ 27

frame36



Εικόνα 51

Object Tracking για το καρέ 36

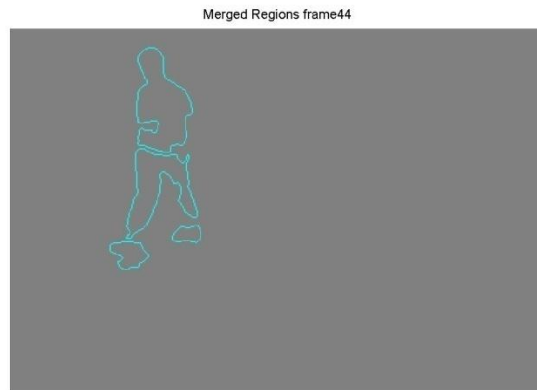


Στο δεύτερο βίντεο επεξεργασίας από μία και σταθερή κάμερα παρατηρούμε έναν άρτιο εντοπισμό των τεσσάρων αντικειμένων που εμφανίζονται στο βίντεο. Αρχικά να σημειωθεί ότι δεν υπήρχε κανένα σταθερό αντικείμενο και πως δεν είχαμε κάποιου είδους επικάλυψη μεταξύ των κινούμενων αντικειμένων. Στη συνέχεια παρατηρούμε ότι στα αντικείμενα 1,3 που εμφανίζονται στο βίντεο δεν έχουμε ίσως τη μέγιστη δυνατή πληροφορία που μπορούμε. Αυτό, όπως έχει αναφερθεί και άλλες φορές στην εργασία αυτή μπορεί να οφείλεται στις μεθόδους μείωσης του φωτισμού και του θορύβου. Ενδέχεται η μέθοδος LogAbout να μην ήταν εν τέλει η καταλληλότερη ή οι πειραματικές τιμές που δόθηκαν στις παραμέτρους a, b, c (βλ. κεφ. 2) να μην ήταν εξίσου κατάλληλες και να χρειαζόταν κάποιος αλγόριθμος εκπαίδευσης για την όρισή τους. Στη συνέχεια πρέπει να σημειώσουμε ότι οι αλγόριθμοι που έχουν υλοποιηθεί λειτουργούν για τον εντοπισμό των αντικειμένων από συγκεκριμένη απόσταση από την κάμερα. Έτσι το φορτηγό (βλ. αντικείμενο #4) έχει εντοπιστεί ένα μικρό κομμάτι του μόνο. Έτσι, συμπεραίνουμε ότι όσο μεγαλύτερη είναι η απόσταση από την κάμερα

τόσος περισσότερος θόρυβος προστίθεται στα καρέ. Γενικά τα παραπάνω αποτελέσματα θεωρούνται και για τα δύο βίντεο άρτια.

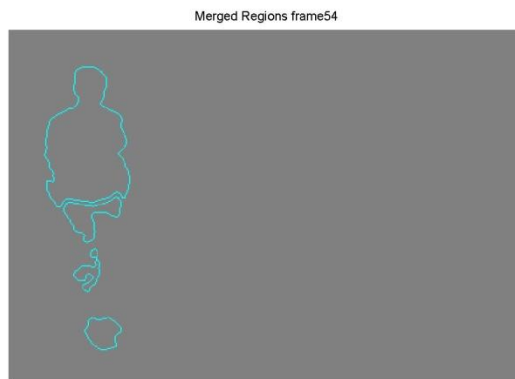
4.4 Εκτεταμένα αποτελέσματα

Όσον αφορά το κομμάτι του Blob Tracking από πολλαπλές κάμερες, τα αποτελέσματα του αλγορίθμου ήταν αρκετά ικανοποιητικά. Ενδεικτικά ας δούμε μερικές εικόνες. Οι πρώτες αριθμητικά εικόνες αφορούν την πρώτη κάμερα ενώ οι δεύτερες τη δεύτερη.



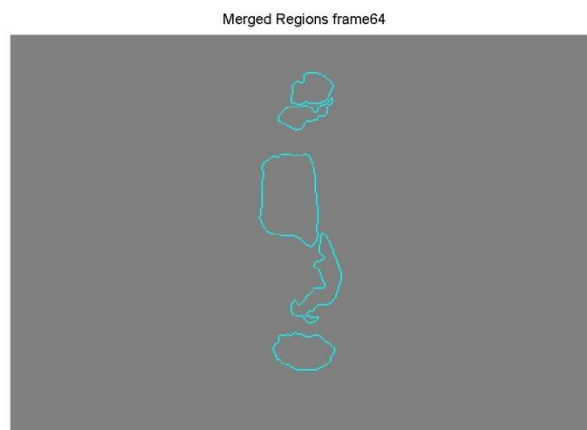
Εικόνα 53

Blob Tracking για το καρέ 44 από πρώτη κάμερα



Εικόνα 54

Blob Tracking για το καρέ 54 από πρώτη κάμερα



Εικόνα 55

Blob Tracking του καρέ 64 από πρώτη κάμερα

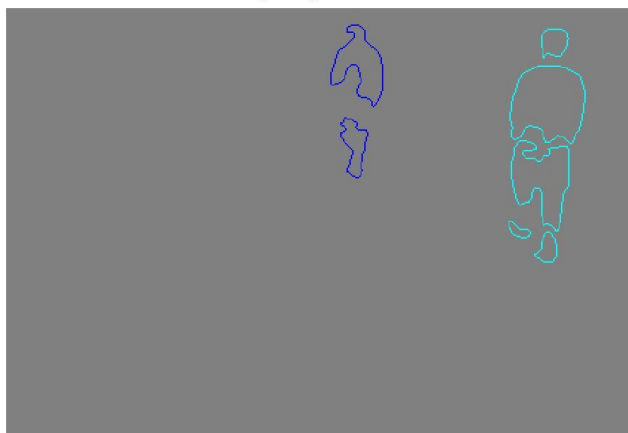
Merged Regions frame74



Εικόνα 56

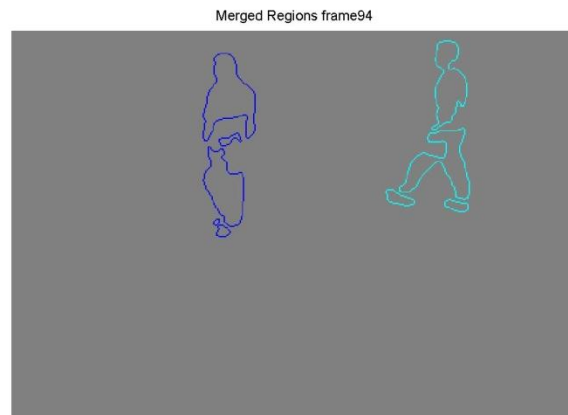
Blob Tracking για το καρέ 74 από πρώτη κάμερα

Merged Regions frame84



Εικόνα 57

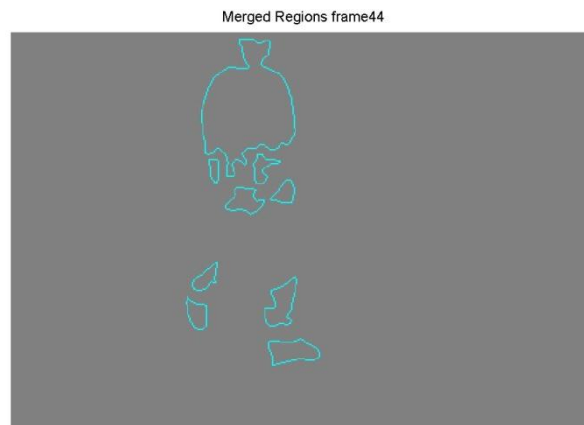
Blob Tracking για το καρέ 84 από πρώτη κάμερα



Εικόνα 58

Blob Tracking για το καρέ 94 από πρώτη κάμερα

Παρακάτω ακολουθούν τα αποτελέσματα της δεύτερης κάμερας. Όπως μπορεί κανείς να παρατηρήσει η μία κάμερα στο χώρο βρίσκεται στην αριστερή γωνία. Η επόμενη κάμερα στον ίδιο χώρο βρίσκεται στην δεξιά γωνία.



Εικόνα 59

Blob Tracking του καρέ 44 από δεύτερη κάμερα

Merged Regions frame54



Εικόνα 60

Blob Tracking του καρέ 54 από δεύτερη κάμερα

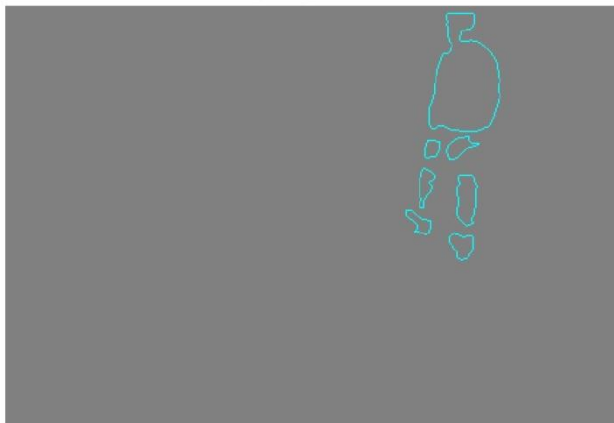
Merged Regions frame64



Εικόνα 61

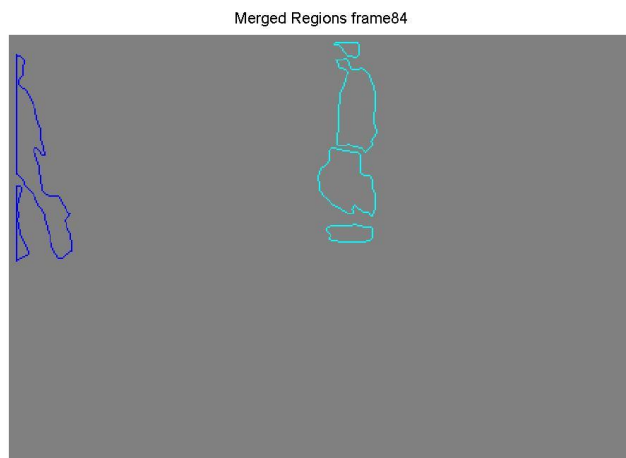
Blob Tracking του καρέ 64 από δεύτερη κάμερα

Merged Regions frame74



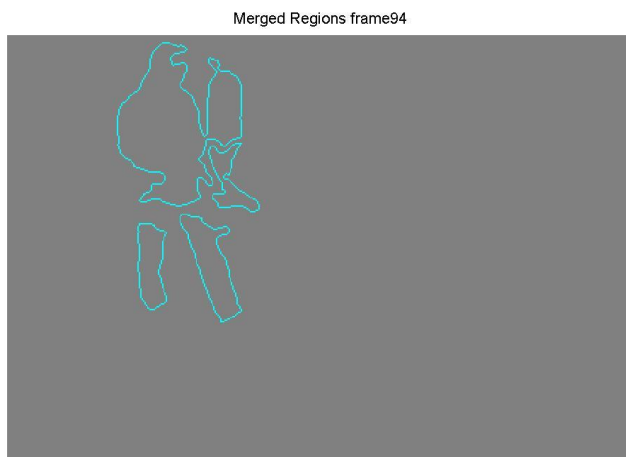
Εικόνα 62

Blob Tracking του καρέ 74 από δεύτερη κάμερα



Εικόνα 63

Blob Tracking του καρέ 84 από δεύτερη κάμερα

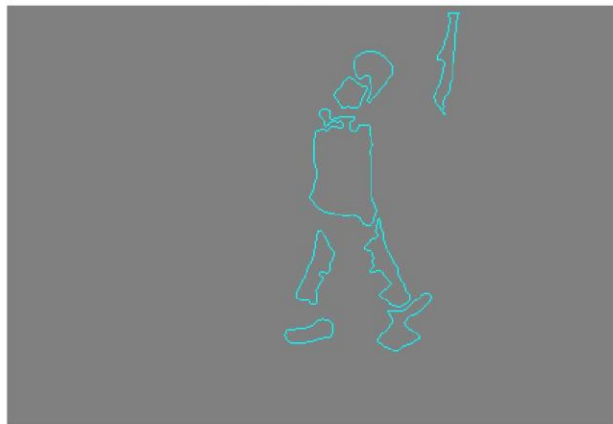


Εικόνα 64

Blob Tracking του καρέ 94 από δεύτερη κάμερα

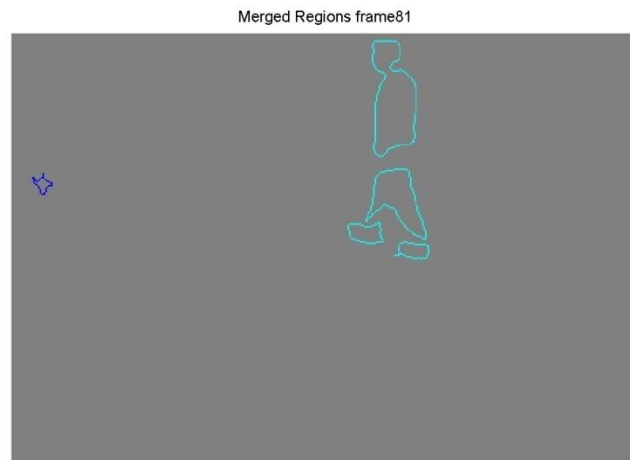
Από τα παραπάνω αποτελέσματα αρχικά παρατηρείται ότι το merging των segmented regions είναι αρκετά ικανοποιητικό πλην των frames που υπάρχει επικάλυψη, όπως φαίνεται στα frames 74, της πρώτης κάμερας και 94 της δεύτερης. Να σημειωθεί ότι δεν υπάρχει συγχρονισμός των frames , όπως φαίνεται παρακάτω ακριβώς τη χρονική στιγμή που εισέρχεται το δεύτερο αντικείμενο στο χώρο και για τις δύο κάμερες (αποτελέσματα από πρώτη και δεύτερη κάμερα αντίστοιχα) . Αυτό μπορεί να οφείλεται είτε σε κακό συγχρονισμό των καμερών είτε στην κακή τοποθέτησή τους.

Merged Regions frame66



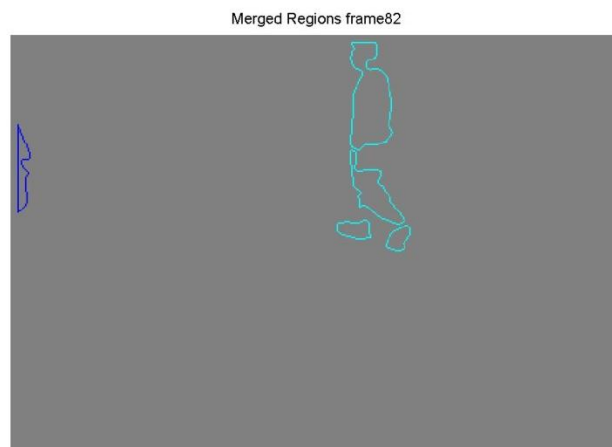
Εικόνα 65

Blob Tracking εισαγωγής του δεύτερου αντικειμένου από πρώτη κάμερα



Εικόνα 66

Blob Tracking εισαγωγής του δεύτερου αντικειμένου από δεύτερη κάμερα



Εικόνα 67

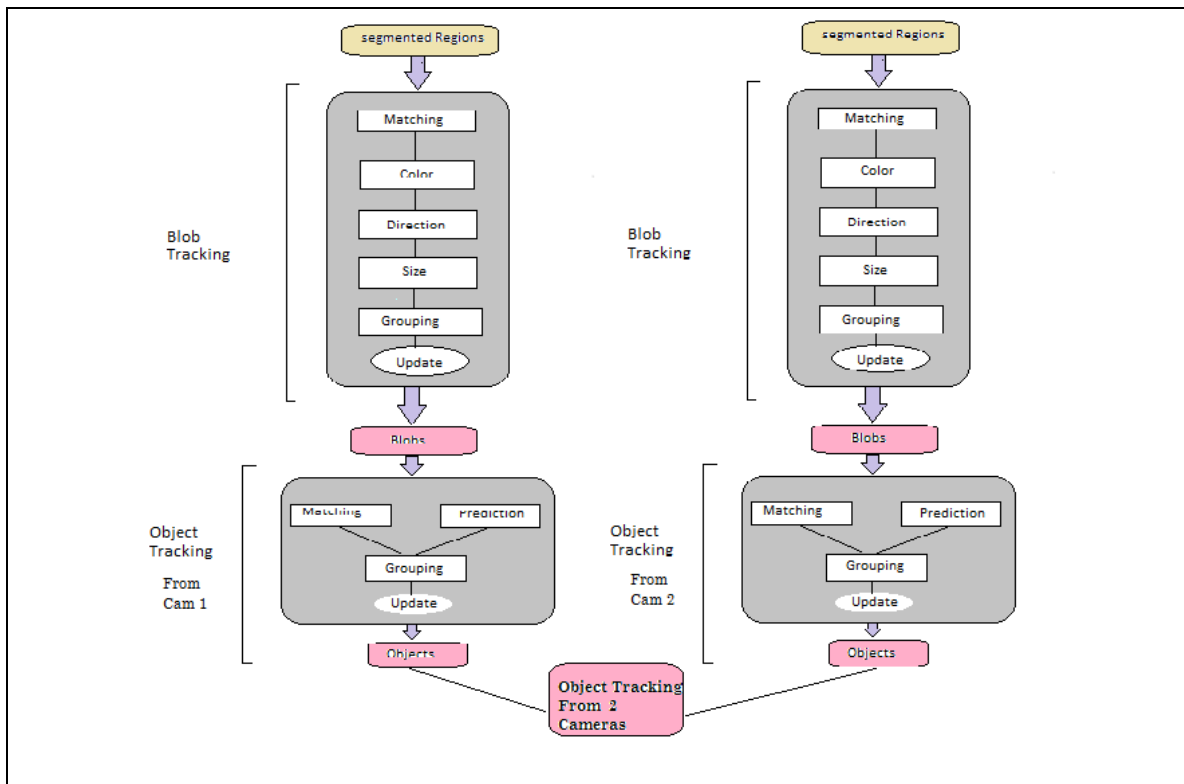
Blob Tracking εισαγωγής του δεύτερου αντικειμένου από δεύτερη κάμερα

Η έρευνά οδηγήθηκε σε άτοπο στο επόμενο κομμάτι του θεωρητικού τρόπου που θα ακολουθούταν έτσι ώστε να πραγματοποιηθεί το αρχικό ζητούμενο. Τα προβλήματα τα οποία οδήγησαν στην μη πραγματοποίηση , λοιπόν, του ζητουμένου ήταν τα παρακάτω.

5^ο Κεφάλαιο: Εντοπισμός αντικειμένων από πολλαπλές κάμερες

Στο τελευταίο στάδιο της εργασίας σκοπός ήταν να πραγματοποιηθεί το Object Tracking , δηλαδή ο εντοπισμός και η ταυτοποίηση των αντικειμένων από ένα σύστημα δύο καμερών. Όταν αναφερόμαστε σε ένα σύστημα δύο και παραπάνω καμερών

εννοούμε πια εντοπισμό αντικειμένων όχι σε δύο διαστάσεις αλλά περνάμε στον τρισδιάστατο χώρο. Πιο συγκεκριμένα, όταν ένα αντικείμενο χάνεται από το οπτικό πεδίο της μίας κάμερας να εντοπίζεται από το οπτικό πεδίο της άλλης και να ταυτοποιείται ταυτόχρονα. Η διαδικασία που ακολουθήθηκε είναι ακριβώς η ίδια με αυτήν που έχει αναλυθεί έως τώρα, απλά θα γινόταν για δύο αρχικά και παραπάνω μετέπειτα κάμερες. Στο παρακάτω διάγραμμα του σχήματος 12, παρατηρείται η διαδικασία που πραγματοποιήθηκε αλλά έφτασε σε άτοπο.

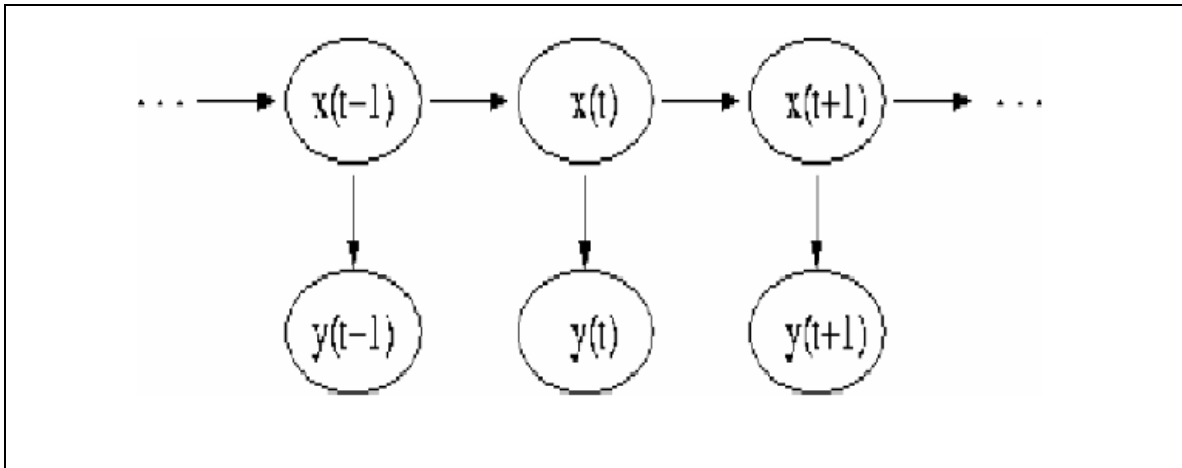


Σχήμα 13: Object tracking from multi cameras diagram

5.1 Hidden Markov Μοντέλο (HMM)

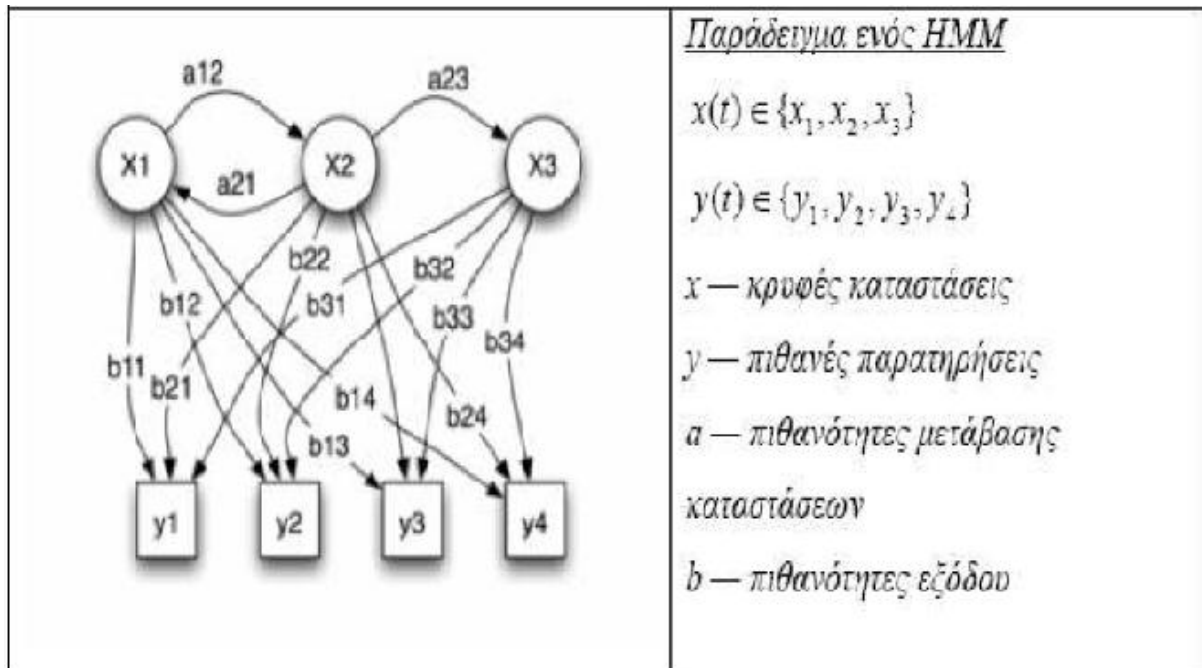
Το HMM είναι το πιο δημοφιλές στατιστικό εργαλείο που μοντελοποιεί ένα ευρύ φάσμα μίας χρονοσειράς δεδομένων. Πιο συγκεκριμένα μπορεί να συνδέσει το

δισδιάστατο χώρο με το χρόνο, δηλαδή μεταφορά στον τρισδιάστατο χώρο. Μπορεί να χρησιμοποιηθεί και για την ανίχνευση αντικειμένων καθώς και την αναγνώριση γεγονότων σε μία βίντεο ακολουθία επιτήρησης. Ένα HMM είναι ένα είδος στοχαστικής μηχανής καταστάσεων. Επιτρέπει μία πιο σύνθετη ανάλυση σε δεδομένα με χωροχρονική μεταβλητότητα [1],[3],[9],[10]. Η δομή ενός HMM είναι η εξής:



Σχήμα 14: Hidden Markov Model (HMM)

Κάθε κύκλος αναπαριστά μία τυχαία μεταβλητή με ένα σύνολο πιθανών τιμών. Η τυχαία μεταβλητή $x(t)$ είναι η κρυφή κατάσταση την χρονική στιγμή t . Τα βέλη δηλώνουν εξαρτήσεις μετάβασης καταστάσεων. Από το διάγραμμα είναι προφανές ότι μια κρυφή τυχαία μεταβλητή $x(t)$ (την χρονική στιγμή t) εξαρτάται μόνο από την τιμή της κρυφής μεταβλητής $x(t-1)$. Οι τιμές των χρονικών στιγμών $t-2$ και νωρίτερα δεν έχουν καμία επίδραση. Αυτό ονομάζεται Μαρκοβιανή ιδιότητα (Markov property). Όμοια, η τιμή της παρατηρούμενης μεταβλητής εξαρτάται μόνο από την τιμή της κρυφής μεταβλητής $x(t)$ (την χρονική στιγμή t). Ας δούμε ένα παράδειγμα ενός HMM:



Σχήμα 15: Παράδειγμα χρήσης ενός HMM

Η χρήση των HMMs αποτελείται από δύο στάδια: την εκπαίδευση και την κατηγοριοποίηση. Στο στάδιο της εκπαίδευσης καθορίζεται ο αριθμός των καταστάσεων του HMM και βελτιστοποιούνται οι πιθανότητες μετάβασης καταστάσεων και εξόδου με τρόπο ώστε τα παραγόμενα σύμβολα να ανταποκρίνονται στα παρατηρούμενα χαρακτηριστικά των εικόνων των παραδειγμάτων μέσα σε μία συγκεκριμένη κατηγορία κίνησης. Στο στάδιο της κατηγοριοποίησης υπολογίζεται η πιθανότητα με την οποία το HMM παράγει την άγνωστη ακολουθία συμβόλων με βάση τα παρατηρούμενα χαρακτηριστικά της. Η πιθανότητα να παρατηρηθεί η ακολουθία $Y = y(0), y(1), \dots, y(L-1)$ μήκους L είναι ίση με:

$$P(Y) = \sum_x P(Y | X) P(X)$$

για όλες τις πιθανές ακολουθίες κρυφών καταστάσεων $X = x(0), x(1), \dots, x(L-1)$. Γενικά το HMM έχει πολύ καλή απόδοση.

6^ο Κεφάλαιο: Μελλοντική εργασία

Σκοπός αυτής της εργασίας ήταν να υλοποιηθούν και να ελεγχθούν οι διαδικασίες που χρειάζονται για την παρακολούθηση των μαζών των αντικειμένων. Σε μελλοντική δουλειά έχοντας δείγματα από τέσσερις κάμερες και όντας γνωστά τα προβλήματα που παρουσιάστηκαν όσον αφορά το Object Tracking, μπορεί να γίνει ο εντοπισμός αντικειμένων από δύο και παραπάνω κάμερες. Για την πραγματοποίηση όμως αυτού του γεγονότος πρέπει να αποφευχθούν τα προβλήματα που αναφέρθηκαν στο κεφάλαιο 5 και να χρησιμοποιηθεί και το Hidden Markov Model (HMM).

6.1 Συμπεράσματα για μελλοντική εργασία

1^ο Πρόβλημα: Εντοπισμός Αντικειμένων

Αρχικά παρουσιάστηκε πρόβλημα στον εντοπισμό των αντικειμένων ατομικά για κάθε μία κάμερα. Ο βασικός λόγος αυτού ήταν το γεγονός ότι στα προηγούμενα δείγματα videos ο ήδη υπάρχων Object Tracking αλγόριθμος λειτουργεί σωστά για συγκεκριμένη απόσταση της κάμερας από τα αντικείμενα, απόσταση η οποία ούτε είναι γνωστή ούτε μπορεί να γνωστοποιηθεί. Αυτό είχε σαν αποτέλεσμα να μη μπορεί να πραγματοποιηθεί πλήρως το δεύτερο κομμάτι του παραπάνω θεωρητικού αλγορίθμου.

2^ο Πρόβλημα: Ταυτοποίηση Αντικειμένων

Όσον αφορά τώρα το τρίτο και τελευταίο κομμάτι του θεωρητικού αυτού αλγορίθμου το βασικότερο πρόβλημα που συναντήθηκε αφορούσε την ταυτοποίηση ,δηλαδή τον καθορισμό κάθε αντικειμένου ως μοναδικό καθ'όλη τη διάρκεια του video. Τα ουσιαστικότερα θέματα που υπήρξαν, που στην ουσία φτάσαμε σε άτοπο , ήταν τα εξής:

Θέμα 1^ο : Οι αλγόριθμοι έχουν δημιουργηθεί με παράγοντα καθορισμού της τροχιάς των αντικειμένων το motion vector των αντικειμένων και όχι το optical flow.

Όπως έχει αναφερθεί στο δεύτερο κεφάλαιο, ο optical flow είναι χρήσιμος για την παρακολούθηση της τροχιάς ενός αντικειμένου σε διάστημα πολλών frames και όχι απλά διαδοχικών, όπως ο motion vector.

Αναφορά σε αυτή τη διαφορά γίνεται διότι, συμπερασματικά με τη χρήση του optical flow θα μπορούσε να αποθηκευτεί σε ένα vector όλος ο όγκος της τροχιάς κάθε αντικειμένου και στη συνέχεια με κατάλληλο calibration να γίνει αντιστοιχία για κάθε frame την ίδια χρονική στιγμή.

Θέμα 2^ο : Το δεύτερο μείζον θέμα που συναντήθηκε ήταν το πρόβλημα του calibration. Με τον όρο calibration εννοούμε τον εντοπισμό των πραγματικών συνθηκών κάτω από τις οποίες πραγματοποιήθηκε ένα σύστημα videos.

Συγκεκριμένα, αυτά τα videos είναι δεδομένα ξένου πανεπιστημίου, που σημαίνει ότι αρχικά και κύρια δε γίνεται να γίνει γνωστή η χρονική καθυστέρηση που είχαν οι κάμερες μεταξύ τους, έτσι ώστε για κάθε frame_t να μην έχω αντιστοιχία δεδομένων. Εκτός αυτού βέβαια, δε γίνεται να γνωρίζουμε τις αποστάσεις των καμερών από το ταβάνι, το έδαφος και μεταξύ των καμερών.

Τα παραπάνω προβλήματα οδήγησαν τον αρχικό στόχο για εντοπισμό των αντικειμένων από πολλαπλές κάμερες σε άτοπο. Ενώ λοιπόν αρχικά είχαμε ξεκινήσει την εργασίας μας με βάση τον εντοπισμό των αντικειμένων από δύο κάμερες διαφορετικού οπτικού πεδίου, καταλήξαμε να εντοπίσουμε τα αντικείμενα από μία και σταθερή κάμερα το οποίο βέβαια είναι εξίσου δύσκολο. Οι λόγοι, πέρα από τα

παραπάνω προβλήματα που έγινε αυτή η αλλαγή της πορείας μας είναι επειδή έπρεπε να αλλάξουμε επίσης όλο το θεωρητικό υπόβαθρο της εργασίας μας, δηλαδή έπρεπε να αλλάξουν οι μεθοδολογίες και οι αλγόριθμοι εξαγωγής της κίνησης και των περιοχών κάτι το οποίο θα έκανε την εργασία μας διπλά δύσκολη και χρονοβόρα.

Βιβλιογραφία

[1] Nhat-Tan Nguyen, Alexandra Branzan-Albu, and Denis Laurendeau: “From optical flow to tracking objects on movie videos”, Computer Vision and Systems Laboratory, Laval University, Quebec (Quebec) G1K 7P4 Canada.

[2] Xiaochen Dai and Shahram Payandeh: “Multi-Level Relay Tracker Utilizing Color-Based Particle Filtering”, Experimental Robotics and Imaging Laboratory, School of Engineering Science, Simon Fraser University, 8888 University Drive, Burnaby, BC, V3J 2W4, Canada.

[3] Chun-hao Wang, Yongjin Wang, Ling Guan: “Event detection and recognition using Histogram of Oriented Gradients and Hidden Markov Models”, Ryerson University, Electrical and Computer Engineering, Toronto, Ontario, Canada M5B 2K3.

[4] Gerard de Haan, Paul W.A.C : “True-Motion Estimation with 3-D Recursive Search Block Matching”, IEEE Transactions on circuits and systems for video technology, vol3, No.5, October 1993.

[5] Yao Nie, Kai-Kuang Ma: “Adaptive Rood Pattern Search for Fast Block-Matching Motion Estimation”, IEEE Transactions on image processing, Vol.11, No.12, December 2002 .

- [6] Hong Liu, Wen Gao, Jun Miao, Jintoo Li: "A Novel Method to Compensate Variety of Illumination in Face Detection", ICT-YCNC FRTJDL, Institute of Computing Technology Chinese Academy of Sciences, Beijing 100080, China.
- [7] a review. Massimo Piccardi: "Background subtraction techniques", Computer Vision Research Group (CVRG) University of Technology, Sydney (UTS)
- [8] Mei Han, Wei Xu, Hai Tao, Yihong Gong: "An Algorithm for Multiple Object Trajectory Tracking", NEC Laboratories America, Cupertino, CA, USA {meihan, xw, ygong@sv.nec-labs.com #University of California at Santa Cruz, Santa Cruz, CA, USA
- [9]» Διπλωματική εργασία Αρβανίτη Γεράσιμου: «Ανίχνευση και παρακολούθηση κίνησης (motion detection and tracking).
- [10] Phil Blunsom: "Hidden Markov Models", August 19, 2004
- [11] Weiming Hu, Tieniu Tan, IEEE, Liang Wang, and Steve Maybank: "A survey on Visual Surveillance of Object Motion and Behaviors".
- [12] Eric Thul: "Adaptive Background Mixture Models for Real-Time Tracking", March 30, 2007
- [13] Sijun Lu, Jian Zhang, David Dagan Feng: "Detection unattended packages through human activity recognition and object association", National ICT Australia (NICTA), 223 Anzac Parade, Kensington, NSW 2052, Australia, School of Information Technology, University of Sydney, Madsen Building F09, University of Sydney, NSW 2006, Australia