



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΕΧΝΙΚΕΣ ΓΙΑ ΑΞΙΟΠΙΣΤΑ ΣΥΣΤΗΜΑΤΑ
ΠΑΡΑΚΟΛΟΥΘΗΣΗΣ

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

ΚΩΝΣΤΑΝΤΙΝΟΣ Γ. ΛΑΛΟΣ
Διπλωματούχου Ηλεκτρονικού Μηχανικού

ΕΠΙΒΛΕΠΩΝ:

Ε. ΠΡΩΤΟΝΟΤΑΡΙΟΣ

Ομότιμος Καθηγητής Ε.Μ.Π.

ΑΘΗΝΑ, Ιούνιος 2012



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ

ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

**ΤΕΧΝΙΚΕΣ ΓΙΑ ΑΞΙΟΠΙΣΤΑ ΣΥΣΤΗΜΑΤΑ
ΠΑΡΑΚΟΛΟΥΘΗΣΗΣ**

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

ΚΩΝΣΤΑΝΤΙΝΟΣ Γ. ΛΑΛΟΣ

Διπλωματούχου Ηλεκτρονικού Μηχανικού

ΤΡΙΜΕΛΗΣ ΣΥΜΒΟΥΛΕΥΤΙΚΗ

ΕΠΙΤΡΟΠΗ:

1. Ε. ΠΡΩΤΟΝΟΤΑΡΙΟΣ, Ομοτ. Καθ. Ε.Μ.Π. (Επιβλέπων)
2. Β. ΛΟΥΜΟΣ, Καθ. Ε.Μ.Π.
3. Θ. ΒΑΡΒΑΡΙΓΟΥ, Καθ. Ε.Μ.Π.

ΕΠΤΑΜΕΛΗΣ ΕΞΕΤΑΣΤΙΚΗ

ΕΠΙΤΡΟΠΗ:

1. Ε. ΠΡΩΤΟΝΟΤΑΡΙΟΣ, Ομοτ. Καθ. Ε.Μ.Π. (Επιβλέπων)
2. Β. ΛΟΥΜΟΣ, Καθ. Ε.Μ.Π.
3. Θ. ΒΑΡΒΑΡΙΓΟΥ, Καθ. Ε.Μ.Π.
4. Α. ΣΤΑΦΥΛΟΠΑΤΗΣ, Καθ. Ε.Μ.Π.
5. Σ. ΠΑΠΑΒΑΣΙΛΕΙΟΥ, Αν. Καθ. Ε.Μ.Π.
6. Δ. ΑΣΚΟΥΝΗΣ, Επικ. Καθ. Ε.Μ.Π.
7. Α. ΔΟΥΛΑΜΗΣ, Επικ. Καθ. Πολυτεχνείου Κρήτης

ΑΘΗΝΑ, Ιούνιος 2012

.....

Κωνσταντίνος Γ. Λάλος

Διδάκτωρ Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Copyright © Κωνσταντίνος Γ. Λάλος, 2012.

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

Πρόλογος

Η διδακτορική διατριβή που παρουσιάζεται στις επόμενες σελίδες εκπονήθηκε από το Ιανουάριο του 2006 μέχρι τον Μάιο του 2012, στο εργαστήριο Τηλεπικοινωνιών του τομέα Επικοινωνιών, Ηλεκτρονικής και Συστημάτων Πληροφορικής, στη Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Εθνικού Μετσόβιου Πολυτεχνείου υπό την επίβλεψη του κ. κ. Εμμανουήλ Πρωτονοτάριου.

Θα ήθελα να ευχαριστήσω από τα βάθη της καρδιάς μου τον καθηγητή μου κ. Εμμανουήλ Πρωτονοτάριο για την υποστήριξη, και την καθοδήγηση που μου παρείχε από την αρχή ως το τέλος της προσπάθειάς μου, καθώς επίσης τους καθηγητές της τριμελούς συμβουλευτικής επιτροπής κ. κ. Βασίλειο Λούμο και Θεοδώρα Βαρβαρίγου.

Επίσης, θα ήθελα να ευχαριστήσω όλους τους συναδέλφους με τους οποίους συνεργάστηκα κατά την διάρκεια της εκπόνησης της διατριβής μου. Ιδιαίτερες ευχαριστίες ωστόσο θα ήθελα να απευθύνω στους συναδέλφους και φίλους Κλεοπάτρα Κωνσταντέλη και Πέτρο Κοσμά.

Τέλος, θα ήθελα να ευχαριστήσω την οικογένειά μου και τους φίλους μου για την στήριξη τους όλα αυτά τα χρόνια.

Κωνσταντίνος Γ. Λάλος

Ιούνιος 2012

Πίνακας περιεχομένων

1	Εισαγωγή	9
1.1	Οργάνωση του εγγράφου.....	11
2	Μέθοδοι Μηχανών Μάθησης.....	15
2.1	Περιγραφή Μηχανών Μάθησης.....	16
2.1.1	Πιθανοτικά Μοντέλα Ταξινόμησης	17
2.1.2	Κατηγοριοποίηση Τεχνικών Μάθησης.....	18
2.2	Μέθοδος Boosting.....	19
2.3	Μέθοδος Boosting μη πραγματικού χρόνου για επιλογή χαρακτηριστικών	23
2.3.1	Χαρακτηριστικά εικόνας	24
2.4	Φίλτρα Σωματιδίων (Particle Filters)	26
2.4.1	Ανάλυση αλγορίθμου φίλτρων σωματιδίων	27
3	Συστήματα Παρακολούθησης.....	32
3.1	Σύστημα Πολυμέσων-Ψυχαγωγίας	35
3.2	Σύστημα Παρακολούθησης και Αναγνώρισης Συμπεριφορών.....	38
3.3	Σύστημα Έλεγχου Βιομηχανικών Διαδικασιών.....	42
4	Παρακολούθηση Αντικειμένων	46
4.1	Τεχνικές παρακολούθησης αντικειμένων.....	47
4.1.1	Διαχωριστικοί αλγόριθμοι για τεχνικές παρακολούθησης	53
4.1.2	Παραγωγικοί αλγόριθμοι για τεχνικές παρακολούθησης.....	57
4.2	Συσχέτιση Δεδομένων	59
5	Ροή αντικειμένου	64
5.1	Διατύπωση προβλήματος μάθησης	65
5.2	Εκτίμηση ροής	70
5.3	Παραδείγματα ροής αντικειμένου.....	73
5.3.1	Ροή αντικειμένου για πεζούς	73
5.3.2	Ροή αντικειμένου για διαφορετικά αντικείμενα.....	79
5.4	Ποσοτική αξιολόγηση	84

6	Προτεινόμενο Σύστημα Παρακολούθησης	89
6.1	Προτεινόμενο Σύστημα Παρακολούθησης.....	92
6.1.1	Φίλτρο Σωματιδίων	93
6.1.2	Ροή Αντικειμένου για Αντικείμενα σε Βιομηχανικό Περιβάλλον	94
6.1.3	Μοντέλο περιγραφής αντικειμένων βάσει Παραγωγικού prior	99
6.1.4	Ενσωμάτωση Ροής Αντικειμένου σε Φίλτρο Σωματιδίων	102
6.1.5	Μηχανισμός Έλεγχου Δειγμάτων Εκπαίδευσης	106
6.2	Πειραματικά Αποτελέσματα.....	110
6.2.1	Πειραματικά Αποτελέσματα για Σύνθετο Βιομηχανικό Περιβάλλον	110
6.2.2	Πειραματικά Αποτελέσματα για τον Μηχανισμό Έλεγχου Δειγμάτων Εκπαίδευσης	122
7	Μέθοδος Συσχέτισης Δεδομένων	127
7.1	Κλιμακούμενη Δεικτοδότηση Πολυμεσικού Υλικού.....	130
7.1.1	Τοπολογία Κελιού.....	130
7.1.2	Χαρτογράφηση Περιεχομένου στην Τοπολογία Κελιών.....	133
7.2	Δυναμική Ομαδοποίηση Πολυμεσικών Στοιχείων	137
7.2.1	Μηχανισμός Δυναμικής Ανανέωσης Περιεχομένου	140
7.3	Δυναμική Ανάκτηση Πολυμεσικών Δεδομένων.....	141
7.4	Πειραματικά Αποτελέσματα.....	143
7.4.1	Παρουσίαση Αντικειμενικών Κριτηρίων	143
7.4.2	Παρουσίαση Αποτελεσμάτων.....	145
8	Σύνοψη	149
8.1	Συνεισφορά – Καινοτομία	149
8.2	Μελλοντική Εργασία.....	152
	Βιβλιογραφικές Αναφορές.....	155

Σχήματα

Σχήμα 1. Αναπαράσταση της μεθόδου Boosting.....	20
Σχήμα 2. (α) Παράδειγμα ορθογωνίου χαρακτηριστικού εικόνας το οποίο παρουσιάζεται εντός του παραθύρου ανίχνευσης. Συγκεκριμένα, το άθροισμα των <i>pixels</i> που βρίσκεται μέσα στα λευκά τετράγωνα αφαιρείται από το άθροισμα των <i>pixels</i> που βρίσκεται μέσα στα γκρι τετράγωνα. Τα χαρακτηριστικά εικόνας με δύο ορθογώνια εμφανίζονται στο σχήμα (A) και (B). Το σχήμα (C) δείχνει ένα χαρακτηριστικό εικόνας με τρία ορθογώνια, και το (D) με τέσσερα. (β) Το άθροισμα των <i>pixels</i> στο ορθογώνιο D μπορεί να υπολογιστεί χρησιμοποιώντας τις τέσσερις αναφορές του πίνακα. Η τιμή της εικόνας <i>integral</i> στη θέση 1 είναι το άθροισμα των <i>pixels</i> στο ορθογώνιο A. Η τιμή της εικόνας <i>integral</i> στη θέση 2 υπολογίζεται ως A+B, στην θέση 3 ως A+C και στην θέση 4 ως A+B+C+D. Τέλος το σύνολο εντός της περιοχής D υπολογίζεται ως 4+1-(2+3)...	25
Σχήμα 3. Αναπαράσταση της γενικής αρχιτεκτονικής του συστήματος POLYMNIA.....	36
Σχήμα 4. Αναπαράσταση της αρχιτεκτονικής του υπό-συστήματος πραγματικού χρόνου.....	36
Σχήμα 5. Η Αρχιτεκτονική του Υποσυστήματος Πραγματικού Χρόνου <i>SemVeillance</i>	39
Σχήμα 6. Αρχιτεκτονική του Υποσυστήματος Μη Πραγματικού Χρόνου.....	40
Σχήμα 7. Εικόνες από το εσωτερικό περιβάλλον της αυτοκινητοβιομηχανίας.....	42
Σχήμα 8. Παρουσίαση του υπό-συστήματος όρασης υπολογιστών	44
Σχήμα 9. Παρουσίαση βίντεο που συλλέχθηκε στο Abbey Road του Λονδίνου (α). Ένας ανιχνευτής βάσει εμφάνισης χρησιμοποιήθηκε για τον εντοπισμό συγκεκριμένων ανθρώπων (β). Ωστόσο αυτή η προσέγγιση δεν δίνει πληροφορίες για την κίνηση του ανθρώπου. Η οπτική ροή δεν μπορεί να προσδιορίσει με ακρίβεια την κίνηση του συγκεκριμένου ανθρώπου αφού υπάρχουν διαφορετικές κινήσεις μέσα στην σκηνή.....	48
Σχήμα 10. Εκτίμηση κίνησης ενός συγκεκριμένου ανθρώπου σε δύο εικόνες βίντεο, όπου παραβιάζεται η χρονική συνέχεια (<i>violation of temporal continuity</i>). Η εκτίμηση γίνεται με την μέθοδο <i>gradient based multi-resolution estimation</i>	49
Σχήμα 11. Παρουσίαση των δυο βασικών κλίσεων ακμών (κεντρική και δεξιά εικόνα) που προέρχονται από μια αρχική εικόνα (αριστερή εικόνα). Ο περιγραφέας <i>HOG</i> δημιουργείται με την βοήθεια αυτής της πληροφορίας.....	52
Σχήμα 12. Παρουσίαση της μεθόδου Boosting πραγματικού χρόνου [4].....	56
Σχήμα 13. Η ανίχνευσης αντικειμένων (α) συνήθως διατυπώνεται ως ένα δυαδικό πρόβλημα ταξινόμησης μιας και ο στόχος είναι διακριθεί το αντικείμενο του ενδιαφέροντος από τον περιβάλλοντα χώρο. Αντίθετα, η μέθοδος ροής αντικειμένου θεωρεί ως πρόβλημα μάθησης την μετατόπιση του αντικειμένου σε τοπικό επίπεδο.....	66
Σχήμα 14. Η εκπαίδευση του ταξινομητή πάνω στην μετατόπιση αντικειμένου επιτυγχάνεται χρησιμοποιώντας θετικά και αρνητικά δείγματα, τα οποία εξάγονται τοπικά και περιέχουν χρονικές πληροφορίες.....	68
Σχήμα 15. Χαρακτηριστικά δείγματα για την εκπαίδευση ενός ταξινομητή στην ροή αντικειμένου.....	69
Σχήμα 16. Αποκρίσεις ταξινομητή για τις περιοχές $\Omega(1)$ και $\Omega(2)$, (α). Χαμηλές τιμές απόκρισης λαμβάνονται στην περίπτωση όπου δεν βρίσκεται το	

αντικείμενο στην περιοχή (γ). Αντίθετα, οι κορυφές δείχνουν την μετατόπιση του αντικείμενου (δ). Το τελικό πεδίο ροής αντικειμένου (β) βασίζεται σε αυτές τις τοπικές αποκρίσεις.....	72
Σχήμα 17. Σε αυτό το πείραμα παρουσιάζονται τα οφέλη της ροής αντικειμένου. Οι προσεγγίσεις οπτικής ροής αποπροσανατολίζονται, όταν παρόμοια αντικείμενα κινούνται προς την ίδια ή διαφορετική κατεύθυνση με το αντικείμενο-του-ενδιαφέροντος (τρίτη γραμμή). Επιπλέον, οι προσεγγίσεις ανθρώπινης ανίχνευσης δεν έχουν σταθερό ρυθμό ανίχνευσης (τέταρτη γραμμή). Επίσης, ο εντοπισμός αντικειμένων δυσκολεύει σε σύνθετα περιβάλλοντα (πέμπτη γραμμή). Αντιθέτως η ροή αντικειμένου (δεύτερη σειρά) μπορεί να προσομοιώσει το πεδίο κίνησης ενός αντικειμένου με επιτυχία, ενώ παράλληλα είναι σε θέση να επικεντρώνετε μόνο στο αντικείμενο υπό μελέτη.....	76
Σχήμα 18. Σε αυτό το πείραμα εκπαιδεύσαμε τον ταξινομητή σε διαφορετικές κατηγορίες αντικειμένων. Η πρώτη και η τρίτη σειρά απεικονίζουν καρέ από διαφορετικές σειρές δοκιμών (κίνηση της κάμερας, θόλωση λόγω κίνησης και πολλαπλά αντικείμενα).....	82
Σχήμα 19. Παραδείγματα εικόνων από βίντεο που μαγνητοσκοπήθηκε σε μία βιομηχανική εγκατάσταση [69]. Οι απλές προσεγγίσεις παρακολούθησης αντικειμένων δεν μπορούν να εφαρμοστούν με ευκολία, λόγω του δύσκολου περιβάλλοντος.....	90
Σχήμα 20. Παραδείγματα εικόνων που χρησιμοποιούνται για την εκπαίδευση του ταξινομητή.....	95
Σχήμα 21. Παρουσίαση της ροής αντικειμένου για το βιομηχανικό περιβάλλον. Οι προσεγγίσεις της οπτικής ροής αποπροσανατολίζονται, από εξωγενείς παράγοντες (π.χ. σπίθες από μηχανή συγκόλλησης), ειδικά όταν κινούνται προς την ίδια κατεύθυνση με το αντικείμενο-του-ενδιαφέροντος (δεύτερη γραμμή). Η τρίτη σειρά παρουσιάζει την εκτιμώμενη ροή αντικειμένου, η οποία επικεντρώνεται μόνο στο αντικείμενο-του-ενδιαφέροντος.....	97
Σχήμα 22. Παρουσίαση της διαδικασίας εκτίμησης για ένα συγκεκριμένο σωματίδιο (<i>particle</i>). Η δειγματοληψία γίνεται με την βάσει των περιοχών απόκρισης(ε) και (ζ).....	103
Σχήμα 23. Ενδεικτικά παραδείγματα του μηχανισμού επαλήθευσης του δείγματος εκπαίδευσης. Χαμηλές τιμές κύρτωσης λαμβάνονται όταν ο άνθρωπος που παρακολουθείται επικαλύπτεται από άλλον άνθρωπο (ε) ή από ένα εμπόδιο (η). Αντίθετα, υψηλότερες τιμές κύρτωσης (β), (δ) παρατηρούνται στις περιοχές όπου το αντικείμενο δεν περιβάλλεται από θορυβώδη χαρακτηριστικά. Μάθηση πραγματικού χρόνου γίνεται με τα δείγματα που έχουν θετικές τιμές κύρτωσης. Η προσέγγιση για τον έλεγχο του δείγματος εκπαίδευσης εκμεταλλεύεται τις πληροφορίες από τα διαγράμματα απόκρισης (δηλαδή, τις αποκρίσεις ταξινομητή, βλ. Σχήμα 24 (β), (δ), (ζ) και (θ)) για την περιοχή που περικλείει ένα δείγμα εκπαίδευσης. Αυτά τα διαγράμματα παρέχονται από ένα προ-εκπαιδευμένο ταξινομητή επάνω στην μετατόπιση αντικειμένων. Στα παραπάνω διαγράμματα, οι τιμές απόκρισης δείχνουν την πιθανότητα για το αντικείμενο που παρακολουθείται να βρίσκεται εντός μίας συγκεκριμένης περιοχής. Ωστόσο, χαμηλότερες τιμές απόκρισης παρατηρούνται, όταν το αντικείμενο που παρακολουθείται επικαλύπτεται από άλλο (βλ. Σχήμα 25 (ε)) ή από ένα εμπόδιο (βλ. Σχήμα 26 (η)). Ως εκ τούτου, η ύπαρξη περιοχών που συσσωρεύονται υψηλές τιμές σημαίνει την ύπαρξη του αντικειμένου εντός αυτής, χωρίς όμως αυτό να περιβάλλεται από θορυβώδη χαρακτηριστικά.....	109

Σχήμα 27. Ενδεικτικά αποτελέσματα. Τα συστήματα <i>OAB1</i> και <i>MIL</i> (δεύτερη και τέταρτη γραμμή) παρασύρονται μακριά από το στόχο, δεδομένου ότι επηρεάζονται από το βιομηχανικό περιβάλλον. Το <i>OAB5</i> (τρίτη γραμμή) δεν παραμένει απολύτως εστιασμένο στο στόχο για όλη την ακολουθία εικόνων. Ο <i>Color tracker</i> (πέμπτη γραμμή) έχει εντελώς ασταθή απόδοση, δεδομένου ότι επηρεάζεται από άλλα χρώματα (π.χ. σπινθήρες από την μηχανή συγκόλλησης). Η προτεινόμενη προσέγγισή (έκτη γραμμή) παραμένει επικεντρωμένη στον κινούμενο στόχο για όλη την ακολουθία.....	115
Σχήμα 28. Ποσοτικά αποτελέσματα. Η προτεινόμενη προσέγγιση παρακολούθησης έχει σταθερή απόδοση σε όλες τις ακολουθίες. Οι άλλες προσεγγίσεις παρασύρονται μακριά από τον στόχο, διότι επηρεάζονται από το σύνθετο βιομηχανικό περιβάλλον.	120
Σχήμα 29. Ποσοτικά αποτελέσματα της προτεινόμενης προσέγγισης παρακολούθησης σε σχέση με άλλες μεθόδους. Η προτεινόμενη μέθοδος λειτουργεί παράλληλα με τον μηχανισμό έλεγχου δειγμάτων εκπαίδευσης με αρκετά μεγάλη σταθερότητα.	124
Σχήμα 30. Παράδειγμα χρησιμοποιούμενης τοπολογίας κελιού.....	133
Σχήμα 31. Παρουσίαση του λόγου βελτίωσης έναντι του μεγέθους της βάσης δεδομένων για διαφορετικά κατώφλια T : 10, 20, 30, 50, 100, 200, 300 και 400	145
Σχήμα 32. Ο λόγος βελτίωσης έναντι του κατωφλιού T για διαφορετικά μεγέθη βάσης δεδομένων.....	146
Σχήμα 33. Η απόδοση ανάκτησης έναντι του μεγέθους της βάσης δεδομένων.....	147
Σχήμα 34. Η απόδοση ανάκτησης έναντι του κατωφλιού για διαφορετικά μεγέθη της βάσης δεδομένων.....	148

Πίνακες

Πίνακας 1. Αλγόριθμος Adaboost (Freund και Schapire [9])	21
Πίνακας 2. Αλγόριθμος Φίλτρων Σωματιδίων (Particle Filters [13])	29
Πίνακας 3. Η μέση τιμή και η διακύμανση σφάλματος μετατόπισης για δύο διαφορετικές μεθόδους εκτίμησης για σενάριο με κινούμενη κάμερα.	85
Πίνακας 4. Η μέση τιμή και η διακύμανση σφάλματος μετατόπισης για δύο διαφορετικές μεθόδους εκτίμησης. Ο υπολογισμός έγινε με ένα σύνολο δεδομένων από μια κάμερα που βρίσκεται σε ένα βιομηχανικό περιβάλλον.	86
Πίνακας 5. Αριθμητική παρουσίαση του μέσου σφάλματος (σε pixel) για τις 5 διαφορετικές εικονοσειρές. Το προτεινόμενο σύστημα παρακολούθησης συγκρίνεται με τις υπόλοιπες μεθόδους.	121
Πίνακας 6. Αριθμητική παρουσίαση του μέσου σφάλματος (σε pixel) για τις 3 διαφορετικές εικονοσειρές. Η προτεινόμενη μέθοδος λειτουργεί παράλληλα με τον μηχανισμό έλεγχου δειγμάτων εκπαίδευσης και συγκρίνεται με τις υπόλοιπες μεθόδους.	125
Πίνακας 7. Αλγόριθμος για την χαρτογράφηση κελιών.	134
Πίνακας 8. Αλγόριθμος για την δημιουργία λογικών διανυσμάτων	136
Πίνακας 9. Αλγόριθμος για την δημιουργία αρχικών ομάδων.	138
Πίνακας 10. Αλγόριθμος για την δυναμική ανανέωση των ομάδων.	140

Περίληψη

Τα συστήματα παρακολούθησης έχουν σημαντικό ρόλο στις εφαρμογές ανίχνευσης συμβάντων και ανάλυσης. Ωστόσο, οι τεχνικές παρακολούθησης αντικειμένων εξακολουθούν να είναι μία από τις μεγαλύτερες προκλήσεις στον τομέα της όρασης υπολογιστών, παρόλο που έχουν μελετηθεί εκτενώς κατά τη διάρκεια των προηγούμενων ετών. Στην παρούσα διατριβή, παρουσιάζεται η ροή αντικείμενου, η οποία χρησιμοποιείται για την εκτίμηση της μετατόπισης και της κατεύθυνσης ενός συγκεκριμένου αντικείμενου. Σε σύγκριση με τις τεχνικές ανίχνευσης και παρακολούθησης, η προτεινόμενη προσέγγιση μπορεί να υπολογίσει την κατεύθυνση του αντικείμενου με απευθείας τρόπο, ενώ παράλληλα μπορεί να αγνοεί άλλες άσχετες κινήσεις μέσα στην σκηνή. Ως εκ τούτου, η ροή αντικείμενου έχει τη δυνατότητα να εστιάζει συνεχώς σε ένα συγκεκριμένο αντικείμενο και να υπολογίζει το πεδίο της κίνησης του. Επίσης, η ροή αντικείμενου ενσωματώνεται σε μία αξιόπιστη προσέγγιση παρακολούθησης, η οποία χρησιμοποιεί ένα παραγωγικό *prior*, με στόχο να αντιμετωπίσει τις αλλαγές της εμφάνισης ενός αντικείμενου. Επιπλέον, προτείνεται ένας μηχανισμός επαλήθευσης δείγματος πραγματικού χρόνου. Η αποτελεσματικότητα της προτεινόμενης μεθόδου παρουσιάζεται σε ποικίλες ακολουθίες, όπως σε ένα σύνθετο βιομηχανικό περιβάλλον. Τα πειραματικά αποτελέσματα δείχνουν ότι η προτεινόμενη τεχνική είναι αξιόπιστη και έχει καλύτερη απόδοση από άλλες μοντέρνες τεχνικές παρακολούθησης αντικειμένων.

Abstract

Camera based supervision is a critical part of event detection and analysis applications. However, visual tracking still remains one of the biggest challenges in the area of computer vision, although it has been extensively discussed during in the previous years. In this thesis, object flow is introduced, for estimating both the displacement and the direction of an object-of-interest. Compared to the detection and tracking techniques, the proposed approach obtains the object displacement directly, while ignoring other irrelevant movements in the scene. Hence, object flow has the ability to continuously focus on a specific object and calculate its motion field. A robust tracking approach based on object flow and an observation model that utilizes a generative prior are adopted to tackle the pitfalls that derive from the appearance changes of the object-of-interest. In addition, an on-line sample verification mechanism is proposed. The efficiency of the proposed technique is demonstrated on various sequences, such as a complex industrial environment. The experimental results show that the proposed technique is sound, yielding improved performance in comparison with other tracking approaches.

1

Εισαγωγή

Η όραση υπολογιστών είναι η επιστήμη που δίνει την δυνατότητα σε μηχανήματα να βλέπουν και να αντιλαμβάνονται τον πραγματικό κόσμο. Συγκεκριμένα, η όραση υπολογιστών ασχολείται με την δημιουργία συστημάτων, τα οποία αντιλαμβάνονται τις πληροφορίες από διάφορες εικόνες ή βίντεο. Αυτές οι πληροφορίες μπορούν να έχουν ποικίλες μορφές, όπως μια σειρά από εικόνες, πλάνα από πολλαπλές κάμερες, ή πολυδιάστατα δεδομένα από μια ιατρική σάρωση.

Η αυτόματη παρακολούθηση αντικειμένων μέσω καμερών είναι ένα απαραίτητο συστατικό των συστημάτων όρασης υπολογιστών, έτσι ώστε να εξαχθούν συμπεράσματα για την κίνηση των αντικειμένων σε μία σκηνή. Ωστόσο, ο τομέας της παρακολούθησης αντικειμένων παραμένει μία από τις μεγαλύτερες προκλήσεις στο ερευνητικό πεδίο της όρασης υπολογιστών, παρόλο που έχει μελετηθεί εκτενώς τα προηγούμενα χρόνια. Συγκεκριμένα, οι τεχνικές παρακολούθησης στοχεύουν στην εκτίμηση της κίνησης ενός ή περισσοτέρων αντικειμένων. Αυτές οι τεχνικές συχνά βασίζονται στις πληροφορίες που προέρχονται από τα κινούμενα αντικείμενα μέσα σε μια σκηνή (π.χ. αυτοκίνητα, ανθρώπους, μηχανήματα). Αυτά τα αντικείμενα συνήθως αλληλεπιδρούν με άλλα αντικείμενα, έτσι ένας από τους βασικούς στόχους της

έρευνας επάνω σε αυτό τον τομέα είναι η μοντελοποίηση της δυναμικής τους συμπεριφοράς.

Στην παρούσα διατριβή προτείνεται μια καινοτόμος μέθοδος για την εκτίμηση του πεδίου κίνησης ενός συγκεκριμένου αντικειμένου σε μια εικονοσειρά. Η τεχνική αυτή ονομάζεται ροή αντικειμένου (*object flow*) και έχει την ικανότητα να προσδιορίζει το πεδίο κίνησης, με παρόμοιο τρόπο με την οπτική ροή. Παράλληλα μπορεί να απορρίπτει άλλες άσχετες κινήσεις στο εσωτερικό μιας σκηνής (π.χ. άλλα κινούμενα αντικείμενα ή κινούμενο φόντο).

Επίσης, παρουσιάζεται ένα σύστημα παρακολούθησης αντικειμένων, το οποίο είναι ικανό να αντιμετωπίσει τις δυσκολίες που προκύπτουν σε ακολουθίες εικόνων από σύνθετα περιβάλλοντα. Πιο αναλυτικά, η τεχνική της ροής αντικειμένου ενσωματώνεται σε ένα φίλτρο σωματιδίων, τόσο ως μοντέλο κίνησης όσο ως μοντέλο περιγραφής αντικειμένου. Επίσης, εισάγεται ένα επιπλέον μοντέλο περιγραφής αντικειμένου, με σκοπό να αντιμετωπίσει προβλήματα που σχετίζονται με τις αναξιόπιστες ανιχνεύσεις αντικειμένου. Αυτό το μοντέλο περιγραφής χρησιμοποιεί μια παραγωγική *prior* μέθοδο, η οποία βασίζεται στο *GPLVM*. Σε αυτή την προσέγγιση το αντικείμενο υπό μελέτη (δηλαδή, ο άνθρωπος που μετακινείται) υποβάλλεται σε διάφορες αλλαγές στην εμφάνιση του. Η *GPLVM* ως μη γραμμική μέθοδος μπορεί να παρέχει μία αποτελεσματικότερη περιγραφή της κατάστασης της εμφάνισης του. Επίσης, παρουσιάζεται ένας μηχανισμός ελέγχου δειγμάτων εκπαίδευσης πραγματικού χρόνου. Αυτός ο μηχανισμός χρησιμοποιεί την κύρτωση ενός διαγράμματος απόκρισης με σκοπό να ελέγξει εάν το αποτέλεσμα του συστήματος παρακολούθησης είναι κατάλληλο για την εκπαίδευση του ταξινομητή σε πραγματικό χρόνο. Από τα πειραματικά αποτελέσματα διαπιστώνεται ότι η προτεινόμενη μέθοδος έχει τη δυνατότητα να παραμείνει συγκεντρωμένη στο

κινούμενο στόχο σε σύγκριση με άλλες προσεγγίσεις από την τρέχουσα βιβλιογραφία που είναι ασταθείς σε σύνθετα περιβάλλοντα.

Το πρόβλημα της συσχέτισης δεδομένων είναι κομβικής σημασίας για τα σύγχρονα συστήματα παρακολούθησης. Θα πρέπει να σημειωθεί ότι, οι τεχνικές συσχέτισης δεδομένων χρησιμοποιούνται κυρίως στις εφαρμογές ανάκτησης και οργάνωσης πολυμέσων (*multimedia content-based retrieval*). Στην παρούσα διατριβή προτείνεται μία μέθοδος συσχέτισης δεδομένων, η οποία μπορεί να οργανώνει δυναμικά διάφορες εικόνες που προέρχονται από ένα σύστημα παρακολούθησης. Συγκεκριμένα, ο προτεινόμενος αλγόριθμος δεικτοδότησης εικόνων έχει την ευελιξία να είναι ανεξάρτητος από το περιεχόμενο της βάσης δεδομένων και τη δομή, ένα γεγονός το οποίο τον κάνει κατάλληλο για εφαρμογές πραγματικού χρόνου. Το προτεινόμενο σχήμα δουλεύει ως ακολούθως. Αρχικά δημιουργείται μια τοπολογία κύβου η οποία αποτελείται από κελιά. Η δομή αυτή χρησιμοποιείται για την αντιστοίχιση του περιεχομένου της εικόνας και/ή του τμήματος βίντεο. Η δομή κελιού χρησιμοποιείται για τον ορισμό συστοιχιών που μοιράζονται κοινό οπτικό περιεχόμενο. Στην συγκεκριμένη εφαρμογή, το οπτικό περιεχόμενο περιγράφεται χρησιμοποιώντας το πρότυπο *MPEG-7*.

1.1 Οργάνωση του εγγράφου

Η παρούσα διατριβή αποτελείται από οκτώ (8) κεφάλαια. Στις ενότητες των κεφαλαίων αυτών παρουσιάζεται με αναλυτικό τρόπο το αντικείμενο της διδακτορικής διατριβής.

Πιο συγκεκριμένα στο Κεφάλαιο 2 γίνεται μία περιγραφή των μηχανών μάθησης (*machine learning*). Συγκεκριμένα, στο Κεφάλαιο 2.1 αναλύονται τα πιθανοτικά μοντέλα ταξινόμησης, τα οποία χωρίζονται σε παραγωγικά και διαχωριστικά. Επίσης,

η μέθοδος *Boosting* παρουσιάζεται αναλυτικά. Στο Κεφάλαιο 2.3 γίνεται η ανάλυση της μεθόδου *Boosting* για επιλογή χαρακτηριστικών εικόνας. Η τεχνική των φίλτρων σωματιδίων αναλύεται σε βάθος στο Κεφάλαιο 2.4.

Στην συνέχεια το Κεφάλαιο 3 εξηγεί τους λόγους για τους οποίους τα συστήματα παρακολούθησης διερευνούνται σε βάθος κατά τη διάρκεια των τελευταίων ετών για την υποστήριξη μιας σειράς εφαρμογών για τις οποίες η κλασική χειροκίνητη παρακολούθηση από ανθρώπους δεν είναι κατάλληλη. Σε αυτήν την κατεύθυνση, παρουσιάζονται συγκεκριμένα παραδείγματα τέτοιων εφαρμογών από διαφορετικές τεχνολογικές περιοχές, όπως συστήματα πολυμέσων-ψυχαγωγίας (*POLYMNIA*) (βλ. Κεφάλαιο 3.1), συστήματα παρακολούθησης και αναγνώρισης συμπεριφορών (*SemVeillance*) (βλ. Κεφάλαιο 3.2) και συστήματα έλεγχου βιομηχανικών διαδικασιών (*SCOVIS*) (βλ. Κεφάλαιο 3.3).

Στο Κεφάλαιο 4 γίνεται μία λεπτομερής περιγραφή της τρέχουσας βιβλιογραφίας. Η παρακολούθηση αντικείμενων (*object tracking*) είναι ένα από τα περισσότερο ενεργά ερευνητικά θέματα στον τομέα της όρασης υπολογιστών. Συγκεκριμένα, στο Κεφάλαιο 4.1 γίνεται μία αναλυτική καταγραφή των τεχνικών για την παρακολούθηση αντικειμένων. Στα Κεφάλαια 4.1.1 και 4.1.2 γίνεται μία παρουσίαση των τεχνικών παρακολούθησης αντικειμένων που βασίζονται σε διαχωριστικούς και παραγωγικούς αλγόριθμους ταξινόμησης. Επίσης, στο Κεφάλαιο 4.1.1.1 γίνεται η ανάλυση αλγόριθμου *Boosting* πραγματικού χρόνου. Τέλος στο Κεφάλαιο 4.2 γίνεται μία παρουσίαση ποικίλων δημοσιεύσεων επάνω στο πρόβλημα της συσχέτισης δεδομένων.

Στο Κεφάλαιο 5 προτείνεται μια καινοτόμος μέθοδος για τον υπολογισμό της μετατόπισης ενός αντικειμένου. Η προτεινόμενη μέθοδος ονομάζεται ροή αντικειμένου (*object flow*) και έχει την ικανότητα να προσδιορίζει το πεδίο κίνησης

ενός συγκεκριμένου κινούμενου αντικειμένου. Παράλληλα μπορεί να απορρίπτει άλλες άσχετες κινήσεις στο εσωτερικό μιας σκηνής. Συγκεκριμένα, στο Κεφάλαιο 5.1 γίνεται η διατύπωση του προβλήματος μάθησης. Επίσης, στο Κεφάλαιο 5.2 παρουσιάζεται ο τρόπος που εκτιμάται η ροή αντικειμένου. Τα αποτελέσματα για πεζούς σε μια πολυσύχναστη σκηνή και για διαφορετικά κινούμενα αντικείμενα παρουσιάζονται στο Κεφάλαιο 5.3.1 και 5.3.2 αντίστοιχα. Τέλος στο Κεφάλαιο 5.4 γίνεται μία αριθμητική αξιολόγηση της προτεινόμενης προσέγγισης.

Στο Κεφάλαιο 6 παρουσιάζεται μια αξιόπιστη προσέγγιση παρακολούθησης αντικειμένου. Η προτεινόμενη μέθοδος χρησιμοποιεί την ροή αντικειμένου τόσο ως μοντέλο κίνησης όσο ως μοντέλο περιγραφής αντικειμένου (βλ. Κεφάλαιο 6.1.2). Στο Κεφάλαιο 6.1.3, εισάγεται ένα επιπλέον μοντέλο περιγραφής αντικειμένου, με σκοπό να αντιμετωπίσει διάφορα προβλήματα που σχετίζονται με αναξιόπιστες ανιχνεύσεις αντικειμένου. Αυτό το μοντέλο περιγραφής χρησιμοποιεί μια παραγωγική *prior* μέθοδο, η οποία βασίζεται στο *GPLVM* (βλ. Κεφάλαιο 6.1.3.1). Επίσης παρουσιάζεται ένας μηχανισμός ελέγχου δειγμάτων εκπαίδευσης πραγματικού χρόνου (βλ. Κεφάλαιο 6.1.5).

Στο κεφάλαιο 7 γίνεται η παρουσίαση μίας μεθόδου συσχέτισης δεδομένων. Συγκεκριμένα, στο Κεφάλαιο 7.1 παρουσιάζεται η προσέγγιση συστοιχιών η οποία χρησιμοποιείται για την βελτίωση του χρόνου απόκρισης. Κατόπιν στο Κεφάλαιο 7.2 παρουσιάζεται η μέθοδος ομαδοποίησης πολυμεσικών δεδομένων. Στο Κεφάλαιο 7.3 γίνεται η ανάλυση της προτεινόμενης μεθόδου ανάκτησης. Τέλος στο Κεφάλαιο 7.4 παρουσιάζονται τα πειραματικά αποτελέσματα όπου αποδεικνύεται ότι ο προτεινόμενος αλγόριθμος έχει γραμμική υπολογιστική πολυπλοκότητα $\mathcal{O}(n)$ για τη

διαγραφή ή εισαγωγή εικόνων ή τμημάτων βίντεο από τις συστοιχίες αντιστοίχως. Η

διατριβή

ολοκληρώνεται με

βιβλιογραφικές

αναφορές.

Η σελίδα αυτή είναι σκόπιμα λευκή.

2

Μέθοδοι Μηχανών Μάθησης

Η παρούσα διατριβή βασίζεται πάνω σε μεθόδους μηχανών μάθησης (*machine learning*). Αυτό το κεφάλαιο εστιάζει σε κάποιες από τις βασικότερες μεθόδους που υπάρχουν στην βιβλιογραφία. Συγκεκριμένα, στο κεφάλαιο 2.1 γίνεται μία περιγραφή των μηχανών μάθησης. Επίσης, αναλύονται τα πιθανοτικά μοντέλα ταξινόμησης, τα οποία χωρίζονται σε παραγωγικά και διαχωριστικά. Κατόπιν γίνεται μία αναλυτική παρουσίαση της μεθόδου *Boosting*, που είναι ένας μετά-αλγόριθμος ο οποίος συνδυάζει πολλαπλούς απλούς ταξινομητές (*classifiers*) σε μια μορφή επιτροπής. Στο Κεφάλαιο 2.3 γίνεται η ανάλυση της μεθόδου *Boosting* για επιλογή χαρακτηριστικών εικόνας. Αυτή η μέθοδος είναι μία από τις σημαντικότερες στο τομέα της όρασης υπολογιστών, διότι συνδυάζει τα πλεονεκτήματα της *Boosting* με τα χαρακτηριστικά εικόνας *Haar*. Η τεχνική των φίλτρων σωματιδίων αναλύεται σε βάθος στο Κεφάλαιο 2.4. Τα φίλτρα σωματιδίων είναι μία πιθανοτική προσέγγιση και επιτυγχάνουν την εκτίμηση της θέσης ενός κινούμενου αντικειμένου χρησιμοποιώντας μία σειρά από θορυβώδης οπτικές μετρήσεις.

2.1 Περιγραφή Μηχανών Μάθησης

Σε αυτήν την ενότητα, παρουσιάζουμε το γενικό υπόβαθρο για τις μεθόδους ταξινόμησης (*classification methods*). Ο κύριος στόχος αυτών των μεθόδων είναι η εκπαίδευση ενός μηχανισμού που είναι σε θέση να κατατάσσει ένα διάνυσμα εισόδου (*input vector*) \mathbf{x} σε μια από τις προκαθορισμένες κατηγορίες-κλάσεις \mathbf{y} , όπου $\mathbf{y} = \mathbf{C}_1, \dots, \mathbf{C}_K$. Η εκπαίδευση του μηχανισμού γίνεται με την κατασκευή ενός συνόλου δειγμάτων εκπαίδευσης (*training set*) $\mathcal{S}_{fix} = \{(x_1, y_1), (x_2, y_2), \dots, (x_L, y_L)\}$ με $Y \in \{-1, +1\}$ και $X \in \mathbb{R}^d$ όπου d είναι η διάσταση του διανύσματος εισόδου.

Μάθηση μηχανών (*machine learning*) ονομάζεται η διαδικασία μέσω της οποίας εκτιμάται η βέλτιστη συνάρτηση $f: X \rightarrow Y$. Ειδικότερα, ο στόχος της επιβλεπόμενης μάθησης μηχανών (*supervised machine learning*) είναι να προβλεφτεί η ετικέτα της πραγματικής κλάσης \mathbf{y} μέσω της συνάρτησης $f(\mathbf{x})$, δηλαδή $\hat{\mathbf{y}} = f(\mathbf{x})$.

Συχνά θεωρείται ότι τα δείγματα εκπαίδευσης εκπορεύονται από μία κατανομή $p(\cdot)$, η οποία ορίζεται στο σύνολο των δειγμάτων $\mathcal{S} = X \times Y$, όπου $\mathcal{S}_{fix} \subseteq \mathcal{S}$. Μια υπόθεση (*Hypothesis*) H κατασκευάζεται προεκτείνοντας την γνώση που προσφέρεται από την συνάρτηση f , η οποία και φέρει την πληροφορία για το \mathcal{S}_{fix} , σε ολόκληρο το χώρο του \mathcal{S} . Αυτή η υπόθεση (*Hypothesis*) περιγράφει όσο το δυνατόν καλύτερα το σύνολο των δειγμάτων \mathcal{S} που εκπορεύονται από μία κατανομή $p(\cdot)$.

Τέλος, ένας αλγόριθμος μάθησης A είναι μία συνάρτηση $A: (X \times Y)^* \rightarrow Y^X$, όπου $(X \times Y)^* = \bigcup_{i=1}^{\infty} (X \times Y)^i$ είναι το σύνολο όλων των δειγμάτων και Y^X είναι το σύνολο των συναρτήσεων $X \rightarrow Y$.

Σε αυτή την διατριβή θα επικεντρώσουμε την προσοχή μας κυρίως στα πιθανοτικά μοντέλα ταξινόμησης (*probabilistic classification models*), όπου οι προβλέψεις για συγκεκριμένες κλάσεις λαμβάνουν την μορφή πιθανοτικών κατανομών. Ωστόσο, θα πρέπει να σημειωθεί ότι αυτές οι τεχνικές εγγενώς περιλαμβάνουν κάποιο βαθμό αβεβαιότητας (*uncertainty*) ως προς τις πιθανοτικές προβλέψεις που γίνονται για την κατάταξη των διανυσμάτων εισόδου \mathbf{x} σε μια από τις διαφορετικές κλάσεις \mathcal{C}_k . Συγκεκριμένα, ο στόχος αυτών των μεθόδων είναι να γίνουν προβλέψεις με τρόπο που να αντικατοπτρίζεται ο βαθμός αβεβαιότητας.

2.1.1 Πιθανοτικά Μοντέλα Ταξινόμησης

Το σημείο εκκίνησης για να αναλυθούν τα πιθανοτικά μοντέλα ταξινόμησης είναι η μελέτη της από κοινού κατανομής $p(\mathbf{y}, \mathbf{x})$, όπου η μεταβλητή \mathbf{y} εκφράζει την αντίστοιχη ετικέτα της κλάσης και \mathbf{x} το διάνυσμα εισόδου, το οποίο η μέθοδος ταξινόμησης καλείται να κατηγοριοποιήσει. Χρησιμοποιώντας το θεώρημα *Bayes* η κατανομή $p(\mathbf{y}, \mathbf{x})$ μπορεί να εκφραστεί και ως $p(\mathbf{y}) p(\mathbf{x}|\mathbf{y}) = p(\mathbf{x}) p(\mathbf{y}|\mathbf{x})$. Αυτό αναδεικνύει δύο διαφορετικές προσεγγίσεις των μεθόδων ταξινόμησης.

- **Παραγωγική (Generative):** Η πρώτη προσέγγιση ονομάζεται παραγωγική (*generative*) και ο στόχος της είναι η ξεχωριστή μοντελοποίηση της κατανομής $p(\mathbf{x}|\mathbf{y})$ (*likelihood*) και της *prior* κατανομής $p(\mathbf{y})$ για όλες τις διαφορετικές κλάσεις \mathbf{y} , όπου $\mathbf{y} = \mathcal{C}_1, \dots, \mathcal{C}_K$. Έτσι η *posterior* κατανομή μπορεί να υπολογιστεί μέσω του θεωρήματος *Bayes* ως $p(\mathbf{y}|\mathbf{x}) \propto p(\mathbf{y}) p(\mathbf{x}|\mathbf{y})$.

- **Διαχωριστική(Discriminative):** Η εναλλακτική προσέγγιση, η οποία ονομάζεται διαχωριστική ή διακριτική (*discriminative*), εστιάζεται στη μοντελοποίηση της *posterior* κατανομής $p(\mathbf{y}|\mathbf{x})$ με τρόπο άμεσο.

Για να γίνουν πρακτικώς εφαρμόσιμες τόσο οι παραγωγικές όσο και οι διαχωριστικές μέθοδοι ταξινόμησης, θα πρέπει να εκτιμηθούν μέσω αλγορίθμων μάθησης και κατάλληλων δειγμάτων, οι βέλτιστες συναρτήσεις είτε για την κατανομή $p(\mathbf{x}|\mathbf{y})$ και $p(\mathbf{y})$ (*generative*) είτε για την $p(\mathbf{y}|\mathbf{x})$ (*discriminative*), αντίστοιχα.

2.1.2 Κατηγοριοποίηση Τεχνικών Μάθησης

Σε αυτή την ενότητα παρουσιάζουμε τις βασικές κατηγορίες των τεχνικών μάθησης.

Αυτές είναι,

- **Μάθηση παρτίδας (batch learning):** Μετά την αρχικοποίηση του μοντέλου, ο μηχανισμός μάθησης επεξεργάζεται όλα τα δείγματα εκπαίδευσης και ενημερώνει συνολικά το μοντέλο. Θα πρέπει να σημειωθεί ότι ο μηχανισμός αυτός επαναλαμβάνεται έως ότου να τερματιστεί από ένα συγκεκριμένο κριτήριο (π.χ. έπειτα από ένα συγκεκριμένο ο αριθμό επαναλήψεων). Οι επαναλήψεις αυτές ονομάζονται και εποχές (*epochs*).
- **Αυξητική μάθηση (incremental learning):** Μετά την αρχικοποίηση του μοντέλου, η διαδικασία μάθησης επαναλαμβάνεται για κάθε δείγμα εκπαίδευσης ξεχωριστά και στη συνέχεια εκτελείται άμεσα η ανανέωση του μοντέλου.
- **Μάθηση σε μη-πραγματικό χρόνο (off-line learning):** Με αυτή την προσέγγιση όλα τα δείγματα εκπαίδευσης είναι εκ των προτέρων διαθέσιμα. Επίσης η πρόσβαση σε αυτά τα δεδομένα μπορεί να γίνει κατ'επανάληψη.

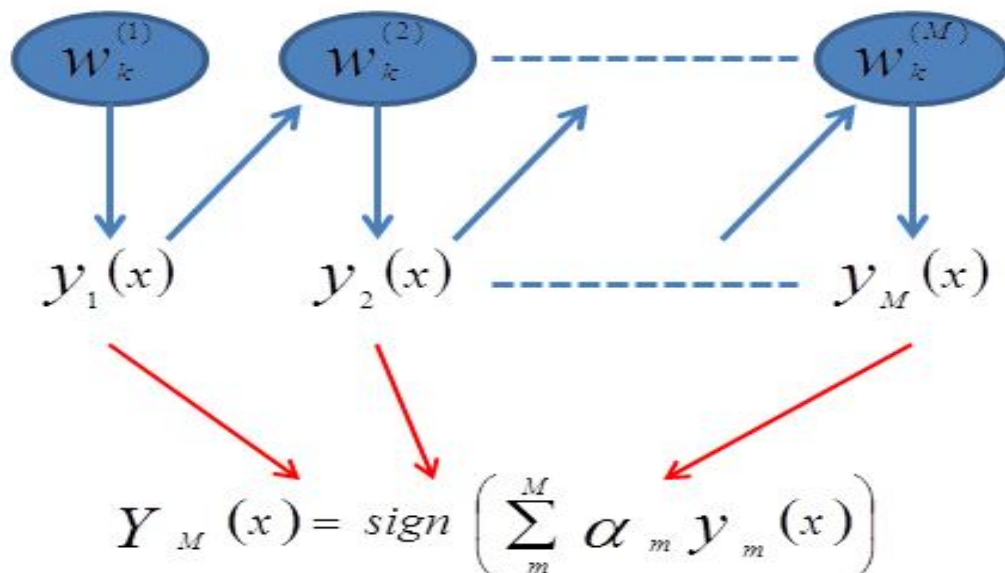
- **Μάθηση σε πραγματικό χρόνο (on-line learning)**: Στην μάθηση πραγματικού χρόνου κάθε δείγμα εκπαίδευσης απορρίπτεται κατόπιν επεξεργασίας και ενημέρωσης του μοντέλου. Επιπλέον, δεδομένου ότι ένας αλγόριθμος μάθησης πραγματικού χρόνου έχει να κάνει με περιορισμένους υπολογιστικούς πόρους, η μνήμη που καταναλώνεται θα πρέπει να παραμένει σταθερή κατά την διάρκεια της εκπαίδευσης

2.2 Μέθοδος *Boosting*

Η μέθοδος *Boosting* είναι ένας μετά-αλγόριθμος, ο οποίος συνδυάζει πολλαπλούς ταξινομητές «βάσης» (*base classifiers*) με σκοπό να δημιουργηθεί μια μορφή επιτροπής, της οποίας η απόδοση μπορεί να είναι σημαντικά καλύτερη από τον κάθε ταξινομητή «βάσης» ξεχωριστά. Η διατριβή αυτή επικεντρώνεται στην προσέγγιση *AdaBoost*, η οποία προτάθηκε από τους *Freund* και *Schapire* [9]. Η μέθοδος *Boosting* μπορεί να δώσει καλά αποτελέσματα, ακόμη και αν οι ταξινομητές «βάσης» (*base classifiers*) έχουν απόδοση, η οποία είναι ελαφρώς καλύτερη από την τυχαία (~50%). Ως εκ τούτου οι ταξινομητές «βάσης» ονομάζονται και *weak learners*.

Η κύρια διαφορά μεταξύ των μεθόδων *Boosting* και άλλων διαφορετικών μεθόδων (π.χ. μέθοδοι επιτροπής *ensemble*) είναι ότι οι ταξινομητές «βάσης» (*base classifiers*) εκπαιδεύονται σε σειρά. Συγκεκριμένα, ο κάθε ταξινομητής εκπαιδεύεται χρησιμοποιώντας μία σταθμισμένη μορφή (*weighted form*) του συνόλου των δειγμάτων εκπαίδευσης, κατά την οποία ο συντελεστής στάθμισης (*weighting coefficient*) που συνδέεται με κάθε δείγμα εκπαίδευσης εξαρτάται μόνο από την απόδοση του προηγούμενου ταξινομητή «βάσης». Ειδικότερα, για τα δείγματα που ταξινομούνται σε λάθος κλάση από έναν από τους ταξινομητές βάσης, δίνεται μεγαλύτερο βάρος όταν εκείνα χρησιμοποιούνται για την εκπαίδευση του επόμενου ταξινομητή «βάσης» στην ακολουθία. Αφού όλοι οι ταξινομητές «βάσης» έχουν

εκπαιδευτεί καταλλήλως, οι προβλέψεις τους συνδυάζονται με την βοήθεια ενός σταθμισμένου συστήματος ψηφοφορίας με πλειοψηφία (*weighted majority voting scheme*). Η αναπαράσταση της μεθόδου *Boosting* απεικονίζεται στο Σχήμα 1.



Σχήμα 1. Αναπαράσταση της μεθόδου Boosting.

Σε αυτό το σημείο θα γίνει ανάλυση ενός δυαδικού προβλήματος ταξινόμησης, στο οποίο τα K δείγματα εκπαίδευσης (*training samples*) εκφράζονται από τα διανύσματα εισόδου x_1, \dots, x_K , μαζί με τις αντίστοιχες μεταβλητές που εκφράζουν την ετικέτα της κλάσης t_1, \dots, t_K , όπου $t_K \in \{-1, +1\}$. Αρχικά, για κάθε δείγμα εκπαίδευσης δίνεται μια παράμετρος στάθμισης (*weighting parameter*) w_K , η οποία έχει την τιμή $1/K$ για όλα τα δείγματα εκπαίδευσης. Σε αυτό το σημείο, γίνεται η υπόθεση ότι υπάρχει μια μέθοδος διαθέσιμη για την εκπαίδευση του κάθε ταξινομητή «βάσης» (*base classifier*), με σκοπό να διαμορφωθεί η συνάρτηση $y(x) \in \{-1, +1\}$. Σε κάθε νέο στάδιο του αλγορίθμου *AdaBoost* εκπαιδεύεται ένας νέος ταξινομητής χρησιμοποιώντας το σύνολο δεδομένων, στο οποίο οι συντελεστές στάθμισης

(*weighting parameters*) προσαρμόζονται ανάλογα με την απόδοση του προηγούμενου εκπαιδευμένου ταξινομητή. Αυτό έχει ως αποτέλεσμα να δίνεται μεγαλύτερο βάρος στα δείγματα εκπαίδευσης, τα οποία δεν ταξινομήθηκαν σωστά.

Τέλος, όταν ο επιθυμητός αριθμός των ταξινομητών «βάσης» έχει εκπαιδευτεί κατάλληλα, αυτοί συνδυάζονται για να σχηματίσουν μια επιτροπή, η οποία χρησιμοποιεί τους συντελεστές στάθμισης με σκοπό να δώσει διαφορετικό βάρος σε κάθε ταξινομητή «βάσης». Η ακριβής μορφή του αλγορίθμου *AdaBoost* παρουσιάζεται παρακάτω.

Πίνακας 1. Αλγόριθμος *Adaboost* (Freund και Schapire [9])

1. Αρχικοποίηση των συντελεστών στάθμισης $\{w_k\}$ θέτοντας $w_k = 1/K$ για

$$k = 1, \dots, K.$$

2. Για $m = 1, \dots, M$:

a. Εκπαίδευση του ταξινομητή $y_m(x)$ με τρόπο που να ελαχιστοποιείται

η παρακάτω συνάρτηση σφάλματος,

$$J_m = \sum_{k=1}^K w_k^{(m)} I(y_m(x_k) \neq t_k) \quad (2.1)$$

όπου $I(y_m(x_k) \neq t_k)$ είναι μία λειτουργική συνάρτηση (*Indicator function*) και ισούται με 1 όταν $y_m(x_k) \neq t_k$ και 0 διαφορετικά.

b. Υπολογισμός της ποσότητας,

$$s_m = \frac{\sum_{k=1}^K w_k^{(m)} I(y_m(x_k) \neq t_k)}{\sum_{k=1}^K w_k^{(m)}} \quad (2.2)$$

και μετά υπολογισμός της ποσότητας,

$$\alpha_m = \ln \left\{ \frac{1 - \epsilon_m}{\epsilon_m} \right\} \quad (2.3)$$

c. Ενημέρωση των συντελεστών στάθμισης,

$$w_k^{(m+1)} = w_k^{(m)} \exp\{\alpha_m I(y_m(x_k) \neq t_k)\} \quad (2.4)$$

3. Η τελική πρόβλεψη γίνεται χρησιμοποιώντας το μοντέλο,

$$Y_M(x) = \text{sign} \left(\sum_{m=1}^M \alpha_m y_m(x) \right) \quad (2.5)$$

Όπως γίνεται αντιληπτό, ο πρώτος ταξινομητής «βάσης» $y_1(x)$ εκπαιδεύεται με τους συντελεστές στάθμισης $w_k^{(1)}$ να είναι όλοι ίσοι. Αυτό το γεγονός αντιστοιχεί με τη συνήθη διαδικασία για την εκπαίδευση ενός ενιαίου ταξινομητή (*single classifier*). Από την Εξίσωση (2.4), παρατηρείται ότι στις επόμενες επαναλήψεις οι συντελεστές στάθμισης $w_k^{(m)}$ αυξάνονται για τα δείγματα εκπαίδευσης που ταξινομούνται εσφαλμένα και μειώνονται για τα δείγματα εκπαίδευσης που ταξινομούνται σωστά. Ως εκ τούτου, οι διαδοχικοί ταξινομητές «βάσης» αναγκάζονται να δώσουν μεγαλύτερη έμφαση στα δείγματα εκπαίδευσης, τα οποία έχουν ταξινομηθεί λάθος από τους προηγούμενους ταξινομητές και τα δείγματα εκπαίδευσης που εξακολουθούν να ταξινομούνται εσφαλμένα από διαδοχικές ταξινομητές λαμβάνουν το μεγαλύτερο βάρος αντίστοιχα. Η ποσότητα ϵ_m αντιπροσωπεύει την σταθμισμένη

μέτρηση των ποσοστών σφάλματος του κάθε ταξινομητή «βάσης» για το σύνολο των δεδομένων. Επίσης, οι συντελεστές στάθμισης α_m που ορίζονται από την Εξίσωση (2.3) δίνουν μεγαλύτερο βάρος στον ταξινομητή «βάσης» που έχει την μεγαλύτερη ακρίβεια. Αυτό παίζει ιδιαίτερο ρόλο κατά τον υπολογισμό του συνολικού αποτελέσματος ταξινόμησης, το οποίο δίνεται από την Εξίσωση (2.5).

2.3 Μέθοδος *Boosting* μη πραγματικού χρόνου για επιλογή χαρακτηριστικών

Οι *Tieu* και *Viola* [9] παρουσίασαν μια μέθοδο *Boosting* που έχει ως στόχο την επιλογή χαρακτηριστικών εικόνας (*image features*). Η ιδέα είναι ότι κάθε μοναδικό χαρακτηριστικό εικόνας αντιστοιχεί σε ένα και μόνο ταξινομητή «βάσης» ή *weak learner* και ο αλγόριθμος *Boosting* ουσιαστικά κάνει επιλογή των κατάλληλων χαρακτηριστικών μέσα από μια μεγάλη συλλογή χαρακτηριστικών F . Δεδομένου ότι αυτή η συλλογή χαρακτηριστικών εικόνας F μπορεί πρακτικά να είναι πολύ μεγάλη, για λόγους υπολογιστικής αποδοτικότητας ο αλγόριθμος επικεντρώνεται μόνο σε ένα υποσύνολο της $F_{sub} = \{f_1, \dots, f_k\} \subseteq F$.

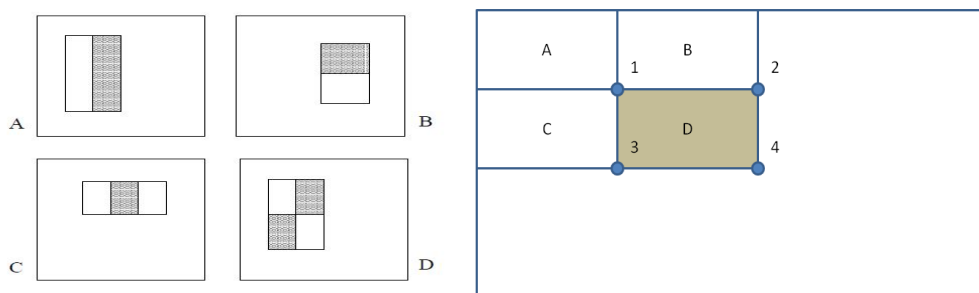
Η διαδικασία της εκπαίδευσης γίνεται με παρόμοιο τρόπο με κλασική μέθοδο *Boosting* (βλ. Κεφ. 2.2). Σε κάθε επανάληψη m του αλγορίθμου, εκείνος επιλέγει ένα νέο χαρακτηριστικό εικόνας (*image feature*) (βλ. Κεφάλαιο 2.3.1) και το προσθέτει (με τον αντίστοιχο συντελεστή στάθμισης w_k^m) στην «επιτροπή» των ταξινομητών «βάσης». Όλα τα χαρακτηριστικά εικόνας αξιολογούνται και η καλύτερη επιλογή γίνεται με βάση την ποσότητα $y_m(x)$, ενώ παράλληλα το α_m ορίζεται σύμφωνα με το

ε_{opt} . Τέλος, η τελική πρόβλεψη του ταξινομητή γίνεται ως σταθμισμένος γραμμικός συνδυασμός των ταξινομητών «βάσης». (βλ. Εξίσωση (2.5)). Θα πρέπει να σημειωθεί ότι το σφάλμα εκπαίδευσης μειώνεται εκθετικά κατά τη διάρκεια των επαναλήψεων του αλγορίθμου *Boosting*, οι οποίες είναι πλέον ίσες ως προς τον αριθμό των επιλεγμένων χαρακτηριστικών εικόνας.

2.3.1 Χαρακτηριστικά εικόνας

Σε αυτό το σημείο θα γίνει παρουσίαση των χαρακτηριστικών εικόνας (*image features*) *Haar* τα οποία χρησιμοποιούνται κυρίως από τον αλγόριθμο *Boosting* (βλ. Κεφάλαιο 2.3). Θα πρέπει να σημειωθεί ότι ως είσοδο στους αλγορίθμους μάθησης τα χαρακτηριστικά εικόνων συχνά προτιμούνται από τις τιμές των *pixel*. Αυτό συμβαίνει διότι βοηθούν στην μείωση της ενδό-μεταβλητότητας εντός μιας συγκεκριμένης κλάσης εικόνων, ενώ παράλληλα αυξάνουν την μεταβλητότητα ανάμεσα σε διαφορετικές κλάσεις εικόνων.

Η μέθοδος *Boosting* μη πραγματικού χρόνου για επιλογή χαρακτηριστικών (*off-line boosting for feature selection*) χρησιμοποιήθηκε στην εργασία του *Viola* και *Jones* [3] προκειμένου να δημιουργηθεί ένας μηχανισμός ανίχνευσης προσώπων. Η βασική ιδέα αυτής της προσέγγισης είναι ότι ένας μικρός αριθμός από χαρακτηριστικά εικόνων μπορεί να είναι επαρκής για τον διαχωρισμό ενός αντικειμένου ενδιαφέροντος από το περιβάλλοντα χώρο. Συγκεκριμένα χρησιμοποιήθηκαν τα χαρακτηριστικά τύπου *Haar* τα οποία παρουσιάζονται στο Σχήμα 2(α) παρακάτω.



(α) χαρακτηριστικά εικόνας Haar (β) παράδειγμα λειτουργίας εικόνας *integral*

Σχήμα 2. (α) Παράδειγμα ορθογωνίου χαρακτηριστικού εικόνας το οποίο παρουσιάζεται εντός του παραθύρου ανίχνευσης. Συγκεκριμένα, το άθροισμα των *pixels* που βρίσκεται μέσα στα λευκά τετράγωνα αφαιρείται από το άθροισμα των *pixels* που βρίσκεται μέσα στα γκρι τετράγωνα. Τα χαρακτηριστικά εικόνας με δύο ορθογώνια εμφανίζονται στο σχήμα (A) και (B). Το σχήμα (C) δείχνει ένα χαρακτηριστικό εικόνας με τρία ορθογώνια, και το (D) με τέσσερα. (β) Το άθροισμα των *pixels* στο ορθογώνιο D μπορεί να υπολογιστεί χρησιμοποιώντας τις τέσσερις αναφορές του πίνακα. Η τιμή της εικόνας *integral* στη θέση 1 είναι το άθροισμα των *pixels* στο ορθογώνιο A. Η τιμή της εικόνας *integral* στη θέση 2 υπολογίζεται ως A+B, στην θέση 3 ως A+C και στην θέση 4 ως A+B+C+D. Τέλος το σύνολο εντός της περιοχής D υπολογίζεται ως $4+1-(2+3)$.

Τα τέσσερα αυτά πρωτότυπα (βλ. Σχήμα 2(α)) μπορούν να διαμορφωθούν σε διαφορετικά μεγέθη και γωνίες με σκοπό να δημιουργήσουν μια ποικιλία χαρακτηριστικών. Η τιμή του χαρακτηριστικού εικόνας *Haar* υπολογίζεται ως το άθροισμα των *pixels* μέσα στις ορθογώνιες περιοχές, οι οποίες είναι είτε θετικές (μαύρες περιοχές) ή αρνητικές (λευκές περιοχές). Ο υπολογισμός αυτής της τιμής του χαρακτηριστικού εικόνας μπορεί να γίνει αποτελεσματικά με τις εικόνες *integral* (βλ. Σχήμα 2(β)). Οι εικόνες *integral*, οι οποίες συμβολίζονται ως *I* συνοψίζουν όλες τις τιμές των *pixel* από το άνω αριστερό άκρο μέχρι το κάτω δεξιά άκρο. Συγκεκριμένα, σε μια εικόνα *I* αυτό ορίζεται ως,

$$I(x,y) = \sum_{x'=1}^x \sum_{y'=1}^y I(x',y') \quad (2.6)$$

Ο προ-υπολογισμός της εικόνας *integral* μπορεί να εφαρμοστεί αποτελεσματικά για όλα τα *pixel* με μια απλή σάρωση της εικόνας. Κατόπιν, το οποιοδήποτε σύνολο περιοχών (π.χ. περιοχές *A*, *B*, *C* και *D* στο Σχήμα 2(β)) μπορεί να υπολογιστεί με 4 προσβάσεις στην μνήμη και με 3 απλές προσθέσεις. Αυτό επιτρέπει στο να γίνεται εξαντλητική αντιστοίχιση βάσει κάποιου πρότυπου (*template matching*) κατά την διαδικασία σάρωσης της εικόνας. Θα πρέπει να σημειωθεί ότι η συγκεκριμένη προσέγγιση παρουσιάζεται για ορθογώνια χαρακτηριστικά εικόνας *Haar*. Ωστόσο αυτό το γεγονός δεν απαγορεύει στο να χρησιμοποιηθούν χαρακτηριστικά εικόνας *Haar* με συγκεκριμένες γωνίες [11].

2.4 Φίλτρα Σωματιδίων (*Particle Filters*)

Η τεχνική των φίλτρων σωματιδίων (*PF*) [13] έχει αποδειχθεί ότι είναι μια καλή λύση για την παρακολούθηση κινούμενων αντικειμένων σε διάφορα συστήματα παρακολούθησης. Η παρακολούθηση αντικειμένων με την χρήση φίλτρων σωματιδίων είναι μια προσέγγιση που έχει συζητηθεί εκτενώς στη βιβλιογραφία κατά τη διάρκεια των προηγούμενων ετών. Εμφανίστηκε στη βιβλιογραφία με διάφορα ονόματα όπως *Bootstrap filtering*, *survival of the fittest*, όπως και ως αλγόριθμος συμπύκνωσης (*condensation algorithm*). Τα φίλτρα σωματιδίων μπορούν να δώσουν λύσεις στα *Hidden Markov Models (HMM)* (βλ. [12] για περισσότερες πληροφορίες), όπου η «κρυφή» κατάσταση (*hidden state*) \mathbf{x}_t των *HMM* μπορεί για παράδειγμα να αντιπροσωπεύει την θέση, ταχύτητα ή επιτάχυνση του κινούμενου αντικειμένου και \mathbf{y}_t τις οπτικές μετρήσεις (*visual measurements*) πάνω στην περιοχή του κινούμενου αντικειμένου. Συγκεκριμένα, μέσα από μια Μπεϋζιανή (*Bayesian*) προοπτική, το

πρόβλημα παρακολούθησης αντικειμένων θα μπορούσε να αντιμετωπιστεί με τον αναδρομικό υπολογισμό της «κρυφής» συνάρτησης πυκνότητας $p(x_t | y_{1:t})$, η οποία σχετίζεται με μία σειρά από οπτικές μετρήσεις $y_{1:t}$. Τα φίλτρα σωματιδίων μπορούν να δώσουν μια καλή λύση σε αυτό το πρόβλημα αντιπροσωπεύοντας την απαιτούμενη συνάρτηση πυκνότητας $p(x_t | y_{1:t})$ με μια σειρά από τυχαία δείγματα με τα κατάλληλα βάρη (δηλαδή με *Monte-Carlo simulations*).

2.4.1 Ανάλυση αλγορίθμου φίλτρων σωματιδίων

Όπως αναφέρθηκε σε προηγούμενη ενότητα, ο στόχος ενός συστήματος παρακολούθησης είναι η αναδρομική εκτίμηση της θέσης ενός κινούμενου αντικειμένου με μια σειρά από θορυβώδης οπτικές μετρήσεις επάνω σε αυτό, χρησιμοποιώντας τα κατάλληλα βήματα πρόβλεψης και ανανέωσης (*prediction and update steps*). Συγκεκριμένα, ο στόχος είναι ο υπολογισμός της *posterior* κατανομής $p(x_t | y_{1:t})$ της «κρυφής» κατάστασης x_t , για την οποία υπάρχουν μια σειρά από μετρήσεις $y_{1:t} = (y_1, \dots, y_t)$ έως την στιγμή t . Στα φίλτρα σωματιδίων η *posterior* κατανομή υπολογίζεται χρησιμοποιώντας την ιδέα της δειγματοληψίας σημασίας (*importance sampling*).

Πιο αναλυτικά, ο αλγόριθμος των φίλτρων σωματιδίων λειτουργεί ως εξής. Αρχικά τα τυχαία δείγματα x_t^i εκπορεύονται από μία προτεινόμενη κατανομή (*proposal distribution*) $q(\cdot)$, δηλ. $x_t^i \sim q(x_t | x_{t-1}, y_{1:t})$ και κατόπιν για κάθε τυχαίο δείγμα x_t^i υπολογίζεται η αντίστοιχη εξίσωση ενημέρωσης (*update equation*) έτσι ώστε να γίνουν οι κατάλληλες μετρήσεις. Θα πρέπει να σημειωθεί ότι γίνονται οι εξής παραδοχές. Συγκεκριμένα, οι μετρήσεις που γίνονται πάνω στα τυχαία δείγματα x_t^i είναι υπό όρους ανεξάρτητες (*conditional independent*) και ότι η κατάσταση x_t

ακολουθεί μια πρώτης τάξεως αλυσίδα Μαρκόφ (*first order Markov series*). Με βάση τις παραπάνω παραδοχές, η εξίσωση ενημέρωσης για το βάρος ενός τυχαίου δείγματος x_t^i μπορεί να εκφραστεί ως εξής,

$$w_t^i = w_{t-1}^i \frac{p(y_t | x_t^i) p(x_t^i | x_{t-1}^i)}{q(x_t^i | x_{t-1}^i, y_t)} \quad (2.7)$$

Ωστόσο, η απόδοση ενός συστήματος παρακολούθησης, το οποίο βασίζεται στα φίλτρα σωματιδίων, εξαρτάται σε μεγάλο βαθμό από την επιλογή της κατανομής $q(\cdot)$. Μια επιλογή που συχνά χρησιμοποιείται είναι η *prior* κατανομή, δηλ. $p(x_t^i | x_{t-1}^i) = q(x_t^i | x_{t-1}^i, y_t)$. Αυτό έχει ως αποτέλεσμα μια νέα εξίσωση ενημέρωσης για το βάρος ενός τυχαίου δείγματος x_t^i . Συγκεκριμένα,

$$w_t^i = w_{t-1}^i p(y_t | x_t^i) \quad (2.8)$$

Η κατανομή $p(y_t | x_t^i)$ συμβολίζει τις οπτικές μετρήσεις (*visual measurements*) πάνω στην περιοχή του κινούμενου αντικειμένου. Αυτό μπορεί να γίνει με την σύγκριση της πιθανής περιοχής με κάποιο πρότυπο (π.χ. *Template matching*). Επίσης, θα πρέπει να σημειωθεί ότι η *prior* κατανομή περιέχει πληροφορίες για τον τρόπο μεταβολής των «κρυφών» καταστάσεων x_t και συνήθως εκφράζεται μέσω κάποιου συγκεκριμένου μοντέλου κίνησης (π.χ. *Autoregressive motion model*). Ωστόσο, η δειγματοληψία από ένα συγκεκριμένο μοντέλο κίνησης μπορεί να οδηγήσει σε εκτροπή του συστήματος παρακολούθησης, ιδιαίτερα όταν ο στόχος κάνει απότομες κινήσεις. Περισσότερες λεπτομέρειες επάνω σε αυτό το θέμα παρουσιάζονται στην αναφορά [13]. Ακολουθεί ο αλγόριθμος για τα φίλτρα σωματιδίων.

Πίνακας 2. Αλγόριθμος Φίλτρων Σωματιδίων (*Particle Filters [13]*)

$$\left[\{x_t^i, w_t^i\}_{i=1}^{N_S} \right] = PF \left[\{x_{t-1}^i, w_{t-1}^i\}_{i=1}^{N_S}, y_t \right]$$

1. Για $t = 1, \dots, N_S$:

a. Εκπόρευση τυχαίων δειγμάτων \mathbf{x}_t^i από την προτεινόμενη κατανομή

$$q(\mathbf{x}_t^i | \mathbf{x}_{t-1}^i, y_t), \text{ δηλ. } x_t^i \sim q(x_t | x_{t-1}, y_{1:t})$$

b. Για κάθε τυχαίο δείγμα \mathbf{x}_t^i υπολογισμός του βάρους w_t^i βάσει της

Εξίσωσης (2.7)

2. Υπολογισμός του ολικού βάρους, $\mathbf{t} = \sum_{i=1}^{N_S} w_t^i$

3. Για $t = 1, \dots, N_S$:

a. Εξομάλυνση του βάρους $w_t^i = \mathbf{1}/N_S w_t^i$

4. Υπολογισμός του $\tilde{N}_{eff} = \frac{1}{\sum_{i=1}^{N_S} (w_t^i)^2}$

5. Εάν $\tilde{N}_{eff} \leq N_S$ (N_S είναι σταθερά που ορίζεται στην αρχή του αλγορίθμου).

Δειγματοληψία με τον παρακάτω αλγόριθμο.

a. Αρχικοποίηση του CDF: $c_1 = 0$

b. Για $t = 2, \dots, N_S$:

– Κατασκευή του CDF: $c_t = c_{t-1} + w_t^t$

c. Ξεκίνημα από την αρχή του CDF: $t = 1$

d. Σχηματισμός ενός αρχικού σημείου: $\mathbf{u}_1 \sim \mathbf{U}[0, N_S^{-1}]$

e. Για $j = 1, \dots, N_S$:

- Μετακίνηση επάνω στο CDF $u_j = u_1 + N_S^{-1}(j - 1)$

- Όταν $u_j > c_i$

i. $t = t + 1$

f. Εκχώρηση δείγματος $\mathbf{x}_t^{j*} = \mathbf{x}_t^j$

g. Εκχώρηση βάρους $w_t^i = 1/N_S$

Η σελίδα αυτή είναι σκόπιμα λευκή.

3

Συστήματα

Παρακολούθησης

Το κεφάλαιο αυτό εξηγεί τους λόγους για τους οποίους τα συστήματα παρακολούθησης διερευνούνται σε βάθος, με στόχο την υποστήριξη μιας σειράς μοντέρνων εφαρμογών, για τις οποίες μπορεί με ευκολία να αντικατασταθεί η κλασική χειροκίνητη παρακολούθηση από ανθρώπους. Οι στόχοι της όρασης υπολογιστών στα αυτόματα συστήματα παρακολούθησης είναι η αυτόματη ανίχνευση και παρακολούθηση συγκεκριμένων αντικειμένων (π.χ. άνθρωποι, μηχανήματα), είτε σε πραγματικό είτε σε μη πραγματικό χρόνο (π.χ. κατά την διάρκεια μιας μαγνητοσκοπημένης ακολουθίας βίντεο). Συγκεκριμένα, έχει πραγματοποιηθεί αρκετή έρευνα σε αυτό τον τομέα, ιδιαίτερα για την παρακολούθηση αντικειμένων όπως πεζοί, πρόσωπα και αυτοκίνητα. Επίσης η έρευνα για εφαρμογές σε βοηθητικά περιβάλλοντα (*assistive environments*) έχει ένα ισχυρό αντίκτυπο σε πολλές ιατρικές εφαρμογές, όπως είναι η παρακολούθηση ασθενών σε χώρους νοσοκομείων.

Ωστόσο, ένα μεγάλο κομμάτι της έρευνας επάνω στα αυτόματα συστήματα παρακολούθησης περιορίστηκε μόνο σε εργαστηριακά πειράματα (π.χ. ακολουθίες εικόνων με συνθετικά δεδομένα) και δεν ασχολήθηκε με προβλήματα που

παρουσιάζονται κυρίως σε συστήματα πραγματικού χρόνου, όπως είναι οι πολλαπλές αλλαγές εμφάνισης του αντικειμένου, επικαλύψεις από άλλα αντικείμενα και αλλαγές φωτισμού.

Θα πρέπει να σημειωθεί ότι οι εφαρμογές παρακολούθησης (*tracking*) είναι ένα κρίσιμο συστατικό για την αυτόματη εξαγωγή συμπερασμάτων. Ειδικότερα, η παρακολούθηση της κίνησης των ασθενών σε χώρους νοσοκομείων έχει λάβει ιδιαίτερη προσοχή, δεδομένου ότι θα μπορούσε να χρησιμοποιηθεί για να παραχθούν πληροφορίες για τα συστήματα αναγνώρισης συμπεριφορών. Επιπρόσθετα, η οπτική παρακολούθηση θα μπορούσε να θεωρηθεί ως μια ασφαλής λύση για την παρακολούθηση των ασθενών, διότι δεν χρησιμοποιούνται ηλεκτρικά καλώδια ή ασύρματοι πομποί στο σώμα του ασθενούς.

Η όραση υπολογιστών είναι η επιστήμη που δίνει την δυνατότητα στους υπολογιστές να βλέπουν και να αντιλαμβάνονται τον πραγματικό κόσμο. Συγκεκριμένα, η όραση υπολογιστών ασχολείται με την δημιουργία συστημάτων τα οποία λαμβάνουν πληροφορίες από εικόνες ή βίντεο. Αυτές οι πληροφορίες μπορούν να έχουν ποικίλες μορφές, όπως μια σειρά από εικόνες, πλάνα από πολλαπλές κάμερες, ή πολυδιάστατα δεδομένα από μια ιατρική σάρωση. Ειδικότερα η σημασία της όρασης υπολογιστών βρίσκεται στο γεγονός ότι έχει πολλαπλές διαφορετικές εφαρμογές όπως,

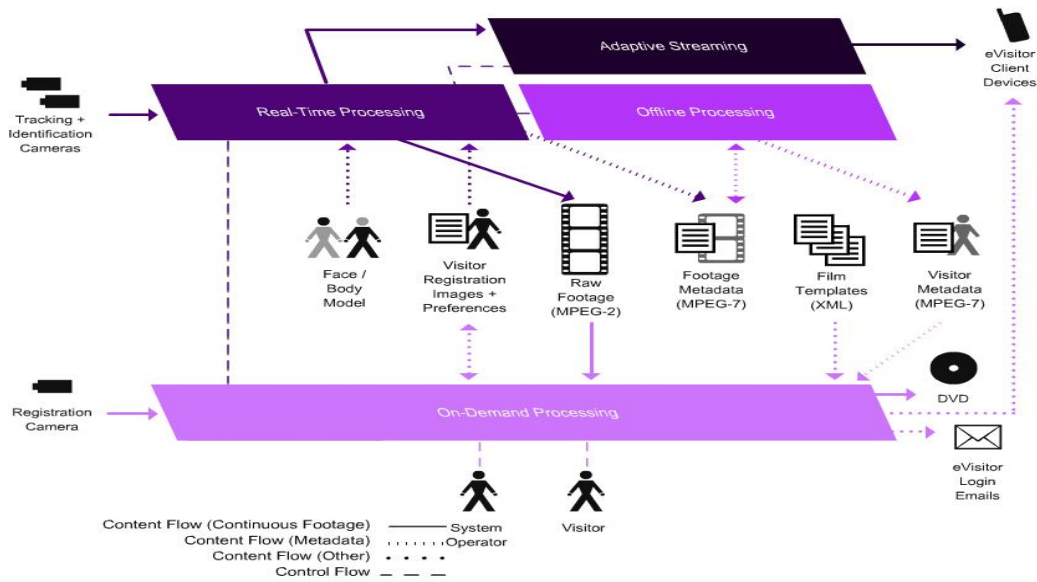
- Έλεγχος διαδικασιών (αυτόνομων οχημάτων και βιομηχανικά ρομπότ),
- Ανίχνευση συμβάντων (π.χ. την ανίχνευση τροχαίων ατυχημάτων),
- Οργάνωση πληροφοριών (π.χ. βάσεις δεδομένων εικόνας ευρετηρίασης),
- Μοντελοποίηση αντικειμένων και περιβάλλοντα χώρου (για την πρόληψη σφαλμάτων σε πυρηνικούς αντιδραστήρες όπου η ανθρώπινη παρέμβαση είναι ιδιαίτερα δύσκολη ή για την ανάλυση ιατρικών εικόνων με σκοπό των εντοπισμό όγκων),

- Αλληλεπίδραση ανθρώπου-υπολογιστή (π.χ. ευφυείς διεπαφές σε ηλεκτρονικά παιχνίδια),
- Παραγωγή βίντεο (για τη βελτίωση της ποιότητας των ταινιών ή για την δημιουργία κινηματογραφικών εφέ),

Τα πολλαπλά οικονομικά οφέλη της επιστήμης της όρασης υπολογιστών σε διάφορες πρακτικές εφαρμογές προκάλεσαν την εκτεταμένη έρευνα, η οποία πρόσφατα επιταχύνθηκε από την σημαντική βελτίωση των ηλεκτρονικών υπολογιστών. Επιπλέον, η αμφίδρομη αλληλεπίδραση της όρασης υπολογιστών με την βιολογία απέφερε κέρδη και για τις δύο επιστήμες. Συγκεκριμένα, τα νέα υπολογιστικά μοντέλα που δημιουργήθηκαν έδωσαν την δυνατότητα στους υπολογιστές στο να αντιλαμβάνονται εικόνες και βίντεο με βάση την ανθρώπινη αντίληψη. Σε αυτήν την κατεύθυνση, παρουσιάζονται παρακάτω μερικές επιστημονικές μελέτες, οι οποίες αναπτύχθηκαν σε διαφορετικά πλαίσια πρακτικών εφαρμογών, όπως συστήματα πολυμέσων-ψυχαγωγίας (*POLYMNIA*) (βλ. Κεφάλαιο 3.1), συστήματα παρακολούθησης και αναγνώρισης συμπεριφορών (*SemVeillance*) (βλ. Κεφάλαιο 3.2) και συστήματα έλεγχου βιομηχανικών διαδικασιών (*SCOVIS*) (βλ. Κεφάλαιο 3.3).

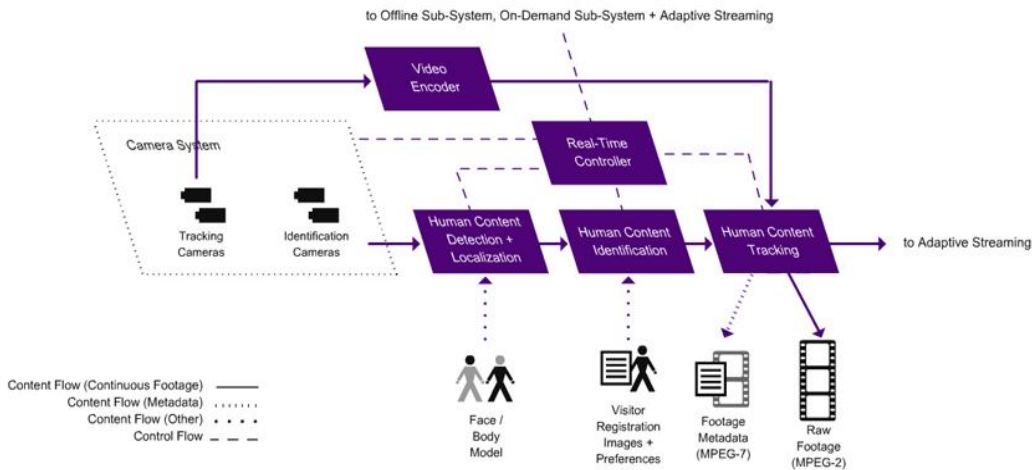
3.1 Σύστημα Πολυμέσων-Ψυχαγωγίας

Η ανάγκη για ένα σύστημα πολυμέσων-ψυχαγωγίας έφερε την δημιουργία του επιστημονικού έργου *POLYMNIA* [14] [15], το οποίο αναπτύχθηκε με σκοπό την κατασκευή μίας έξυπνης *cross - media* πλατφόρμας για προσωπική ψυχαγωγία και διασκέδαση σε θεματικά πάρκα (π.χ. *Allou Fun Park*). Συγκεκριμένα, οι επισκέπτες του θεματικού πάρκου εντοπίζονται αυτόματα, απομονώνονται από το περιβάλλον και κατόπιν καταγράφονται οι κινήσεις τους ξεχωριστά. Η καταγραφή γίνεται με την βοήθεια ενός συστήματος όρασης υπολογιστών που διαθέτει πολλές κάμερες, οι οποίες έχουν τοποθετηθεί σε διάφορα σημεία του θεματικού πάρκου. Η καταγραφή αυτή γίνεται προκειμένου να παραχθεί ένα υψηλής ποιότητας σουβενίρ (*DVD*) προσαρμοσμένο στον επισκέπτη, μέσα στο οποίο εκείνος θα εμφανίζεται ως ο πραγματικός πρωταγωνιστής στον χώρο του θεματικού πάρκου. Επιπλέον, η πλατφόρμα *POLYMNIA* παρέχει την δυνατότητα εμπλουτισμού του σουβενίρ με εξατομικευμένο ηλεκτρονικό περιεχόμενο. Συγκεκριμένα, ο επισκέπτης μπορεί να επιλέξει μεταξύ κάποιων προκαθορισμένων σεναρίων επαυξημένης πραγματικότητας (*augmented reality*). Επίσης, το έργο *POLYMNIA* ανέπτυξε μια έξυπνη πλατφόρμα *cross - media*, της οποίας ο σκοπός είναι η διανομή ψηφιακού περιεχομένου σε πολλαπλούς απομακρυσμένους χρήστες, οι οποίοι έχουν πρόσβαση στο ιντερνέτ από διαφορετικές τερματικές συσκευές, όπως υπολογιστές, κινητά τηλέφωνα και *PDA*. Η τεχνολογία αυτή δίνει την ευκαιρία στους επισκέπτες του θεματικού πάρκου να μοιραστούν την εμπειρία τους με φίλους ή μέλη της οικογένειας τους που βρίσκονται σε απόσταση. Μια γενική επισκόπηση της αρχιτεκτονικής *POLYMNIA* παρουσιάζεται στο Σχήμα 3.



Σχήμα 3. Αναπαράσταση της γενικής αρχιτεκτονικής του συστήματος POLYMNIA.

Το σύστημα POLYMNIA αποτελείται από πολλαπλά υπό-συστήματα, τα οποία έχουν διαφορετικές λειτουργίες και αλληλεπιδρούν μεταξύ τους. Ωστόσο, για την καταγραφή και τον εντοπισμό των επισκεπτών μέσα στο θεματικό πάρκο, το σύστημα POLYMNIA βασίστηκε σε ένα καινοτόμο υπό-σύστημα όρασης υπολογιστών πραγματικού χρόνου (*Real-Time Processing Unit*). Η αρχιτεκτονική του υπό-συστήματος πραγματικού χρόνου παρουσιάζεται στο Σχήμα 4.



Σχήμα 4. Αναπαράσταση της αρχιτεκτονικής του υπό-συστήματος πραγματικού χρόνου.

Αυτή η μονάδα αποτελείται ένα σύστημα αναγνώρισης, εντοπισμού συγκεκριμένων ανθρώπων και από ένα σύστημα διεπαφής (*interface*), το οποίο αρχικοποιεί το σύστημα καμερών. Συγκεκριμένα, το υπό-σύστημα πραγματικού χρόνου επεξεργάζεται τα δεδομένα που συλλέγονται από τις κάμερες που είναι τοποθετημένες στο χώρο του θεματικού πάρκου.

Αρχικά όταν ένας επισκέπτης εισέρχεται στο χώρο του θεματικού πάρκου, οι εικόνες από το πρόσωπο και το σώμα του (μέσω *face* και *body registration*) κατατίθενται σε μια βάση δεδομένων (*POLYMNIA Database*). Έπειτα με την βοήθεια του συστήματος εντοπισμού ανθρώπων (*human and face detection*) παρέχονται οι συντεταγμένες προσώπου και σώματος του στον χώρο που βρίσκεται. Για να γίνει ταχύτερη η αναγνώριση, χρησιμοποιείται η μέθοδος του επίκεντρου προσοχής (*FOA*), στην οποία τίθενται σε λειτουργία κλιμακούμενοι ταξινομητές (*cascade classifiers*). Με την βοήθεια συγκεκριμένων καμερών η μονάδα αναγνώρισης (*identification*) προσδιορίζει την ταυτότητα του επισκέπτη. Θα πρέπει να σημειωθεί ότι οι κάμερες αναγνώρισης είναι τοποθετημένες σε συγκεκριμένες θέσεις, έτσι ώστε να διευκολύνεται η διαδικασία αναγνώρισης.

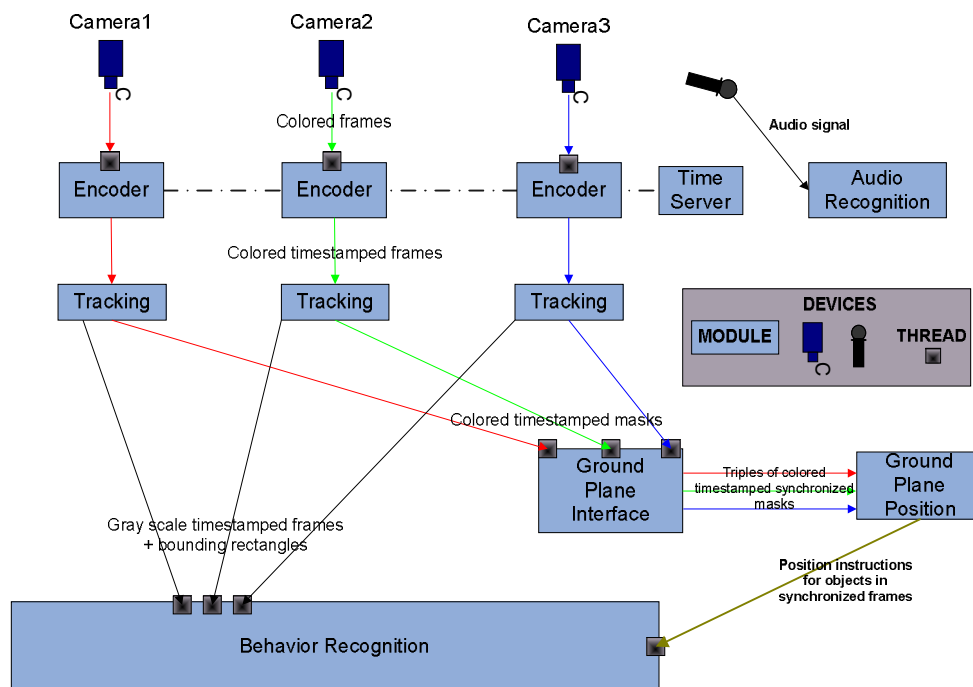
Κατόπιν τίθενται σε λειτουργία οι κάμερες που έχουν ως σκοπό την συνεχή παρακολούθηση (*tracking*) των επισκεπτών. Σε αυτό τον μηχανισμό χρησιμοποιείται η τεχνική αφαίρεσης περιβάλλοντος (*background subtraction*) που αποσκοπεί στην αναγνώριση των κινούμενων περιοχών. Ειδικότερα, οι κινούμενες περιοχές αναλύονται με σκοπό να καθοριστούν οι θέσεις για το κεφάλι και το σώμα του επισκέπτη. Κάθε μονάδα παρακολούθησης (*tracking*) επικοινωνεί με τις υπόλοιπες κάμερες, έτσι ώστε να μεταδώσει τα κατάλληλα δεδομένα για την μετατόπιση των επισκεπτών. Αυτό γίνεται ειδικά όταν οι επισκέπτες εισέρχονται σε νέες περιοχές του θεματικού πάρκου.

3.2 Σύστημα Παρακολούθησης και Αναγνώρισης

Συμπεριφορών

Για την ανάπτυξη ενός συστήματος παρακολούθησης και αναγνώρισης συμπεριφορών δημιουργήθηκε το έργο *SemVeillance*. Συγκεκριμένα, ο στόχος αυτού ήταν η δημιουργία ενός συστήματος εντοπισμού της παραβατικής συμπεριφοράς, αλλά και η ανάπτυξη μηχανισμών αναζήτησης, εύρεσης και πλοήγησης οπτικών δεδομένων με βάση τα στοιχεία της συμπεριφοράς. Ένα τέτοιο σύστημα όρασης υπολογιστών θα ήταν πολύ σημαντικό σε μία πλειάδα από εφαρμογές και οργανισμούς παροχής υπηρεσιών τόσο στον ευρύτερο τομέα των μαζικών μεταφορών όσο και στον τομέα του ελέγχου της ποιότητας διεργασιών. Τέτοια συστήματα είναι σε παγκόσμιο επίπεδο σε πρωτόγονη μορφή και οποιαδήποτε συμβολή σε αυτά αποτελεί ένα σημαντικό βήμα στην διαδικασία ανάπτυξης της ασφάλειας.

Το σύστημα όρασης υπολογιστών *SemVeillance* χωρίζεται στο υπό-σύστημα πραγματικού χρόνου και στο υποσύστημα μη πραγματικού χρόνου. Συγκεκριμένα το υποσύστημα πραγματικού χρόνου αποτελείται από συγκεκριμένες υπό-μονάδες. Το παρακάτω σχήμα (Σχήμα 5) παρουσιάζει την αρχιτεκτονική του υπό-συστήματος καθώς και το πώς οι διάφορες υπό-μονάδες συνεργάζονται μεταξύ τους.



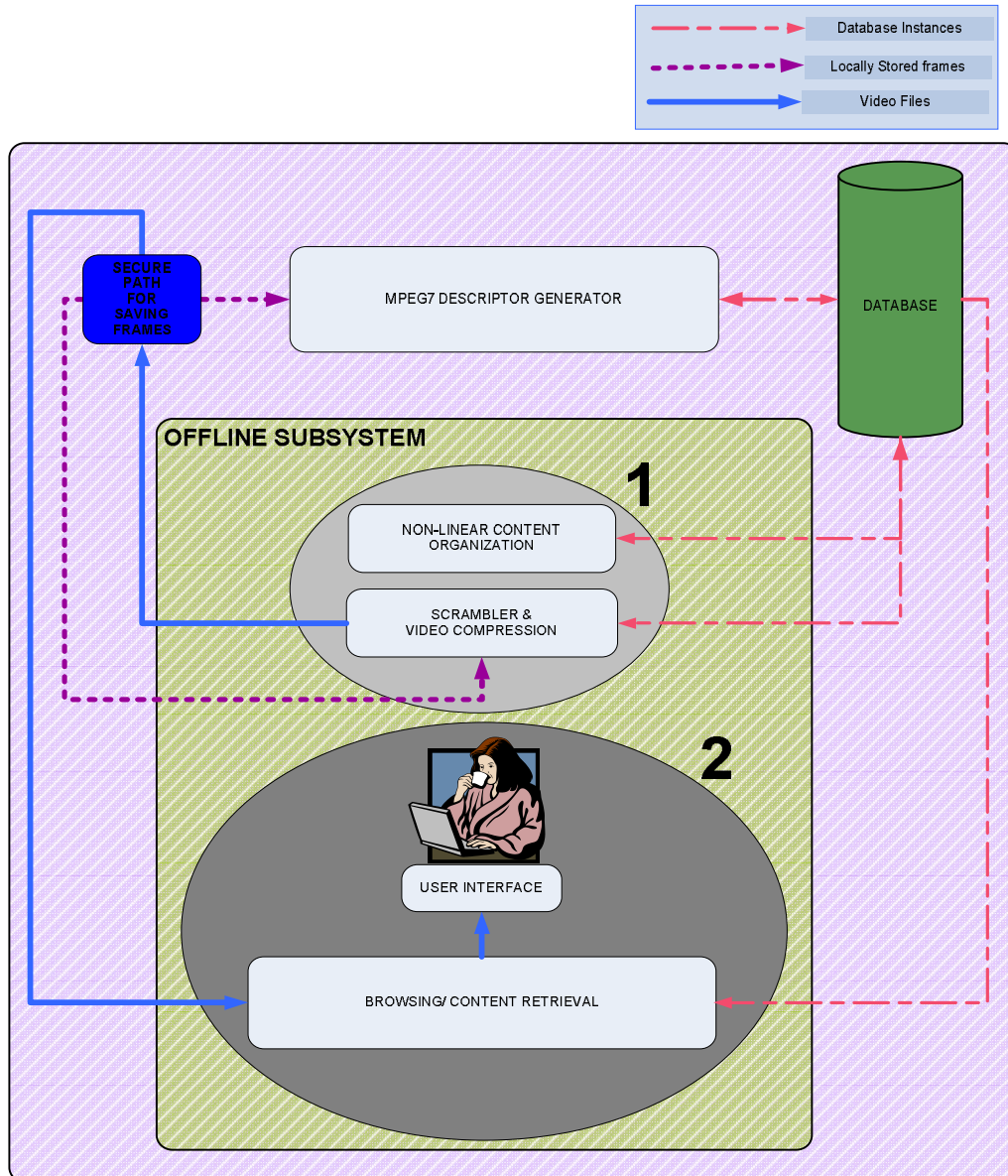
Σχήμα 5. Η Αρχιτεκτονική του Υποσυστήματος Πραγματικού Χρόνου *SemVeillance*

Ο βασικός στόχος του υποσυστήματος πραγματικού χρόνου είναι ο εντοπισμός και η αναγνώριση βραχυπρόθεσμων συμπεριφορών (στάση, τρέξιμο, περπάτημα, απότομη κίνηση) από τα άτομα που βρίσκονται μέσα στο χώρο έλεγχου του συστήματος *SemVeillance*. Η διαδικασία αναγνώρισης συμπεριφορών λειτουργεί ως εξής.

Αρχικά θα πρέπει να σημειωθεί ότι ο χώρος καλύπτεται από **3** κάμερες *AXIS PTZ 213*¹, οι οποίες έχουν επικαλυπτόμενο πεδίο φάσματος. Οι εικόνες (*frames*) που λαμβάνονται από τις κάμερες προωθούνται στον *encoder*. Συγκεκριμένα, η υπό-μονάδα *encoder* παραλαμβάνει τις εικόνες, τους αποδίδει χρονική ταυτότητα για λόγους συγχρονισμού (*timestamps*) και τις αποθηκεύει τοπικά. Επιπλέον, προωθεί την κάθε εικόνα στην υπό-μονάδα αναγνώρισης συμπεριφορών (*Behavior Recognition*). Κάθε εικόνα (*frame*) που λαμβάνεται από την υπό-μονάδα αναγνώρισης

¹ http://www.axis.com/products/cam_213/

συμπεριφορών (*Behavior Recognition*), χρησιμοποιείται για τη δημιουργία και την συντήρηση ενός απλού μοντέλου περιβάλλοντος (*background model*), από την μονάδα *tracking*. Το μοντέλο αυτό χρησιμοποιείται για την ανίχνευση και την παρακολούθηση των ατόμων σε πραγματικό χρόνο.



Σχήμα 6. Αρχιτεκτονική του Υποσυστήματος Μη Πραγματικού Χρόνου

Επίσης, το σύστημα *SemVeillance* αποτελείται και από ένα υποσύστημα μη πραγματικού χρόνου, το οποίο εξάγει διάφορους περιγραφείς *MPEG-7*² για τις εικόνες (*frames*) που έχουν καταχωρηθεί στην βάση δεδομένων του συστήματος, οι οποίες περιέχουν τις συμπεριφορές που εντοπίστηκαν από το υποσύστημα πραγματικού χρόνου. Συγκεκριμένα, οι εξαγόμενοι περιγραφείς *MPEG-7* οργανώνονται σε τοπολογία κελιών [16]. Κατά αυτό τον τρόπο επιτυγχάνουμε με γραμμική πολυπλοκότητα την αναδιοργάνωση των στοιχείων σε μία δυναμικά μεταβαλλόμενη πλατφόρμα. Μέσω μίας διαπροσωπεία χρήσης (*user interface*) η μονάδα ανάκτησης και αναζήτησης δεδομένων παρουσιάζει μια λίστα στην οποία ο χρήστης μπορεί να επιλέξει τον περιγραφέα *MPEG-7* που επιθυμεί. Κατόπιν η διαπροσωπεία χρήσης (*user interface*) παρουσιάζει τα βίντεο στα οποία βρέθηκε ο συγκεκριμένος περιγραφέας *MPEG-7*. Το προτεινόμενο σύστημα μπορεί επίσης να χρησιμοποιηθεί σε διάχυτα υπολογιστικά συστήματα (*pervasive computer systems*).

Επίσης, το υποσύστημα μη πραγματικού χρόνου λαμβάνει ως είσοδο ακολουθίες βίντεο και τις κωδικοποιεί ώστε να μην είναι ορατό το περιεχόμενο στην αρχική του μορφή και συνεπώς διασφαλίζονται τα προσωπικά δεδομένα.

² <http://mpeg.chiariglione.org/standards/mpeg-7/mpeg-7.htm>

3.3 Σύστημα Έλεγχου Βιομηχανικών Διαδικασιών

Στα πλαίσια του επιστημονικού έργου *SCOVIS*³ δημιουργήθηκε ένα σύστημα ελέγχου βιομηχανικών διαδικασιών που αποσκοπεί στην βελτίωση της ευελιξίας και της απόδοσης των σημερινών συστημάτων παρακολούθησης για τον έλεγχο της ροής εργασίας σε διάφορες υποδομές. Το *SCOVIS* επιτρέπει την εύκολη και αποτελεσματική εγκατάσταση ευφών συστημάτων εποπτείας, κάτι το οποίο δεν κατέστη δυνατό μέχρι σήμερα λόγω των πολύπλοκων οπτικών διεργασιών. Για αυτό το λόγο επιλέχθηκε το εσωτερικό περιβάλλον μιας αυτοκινητοβιομηχανίας (βλ. Σχήμα 7) με σκοπό την αξιολόγηση του συστήματος *SCOVIS* κάτω από πραγματικές συνθήκες εργασίας. Συγκεκριμένα, το σύστημα *SCOVIS* υποστηρίζει την αυτόματη ανίχνευση συμπεριφορών, την παραβίαση ροών εργασίας και τον εντοπισμό των κυριότερων αντικειμένων (κινούμενων ή στατικών).



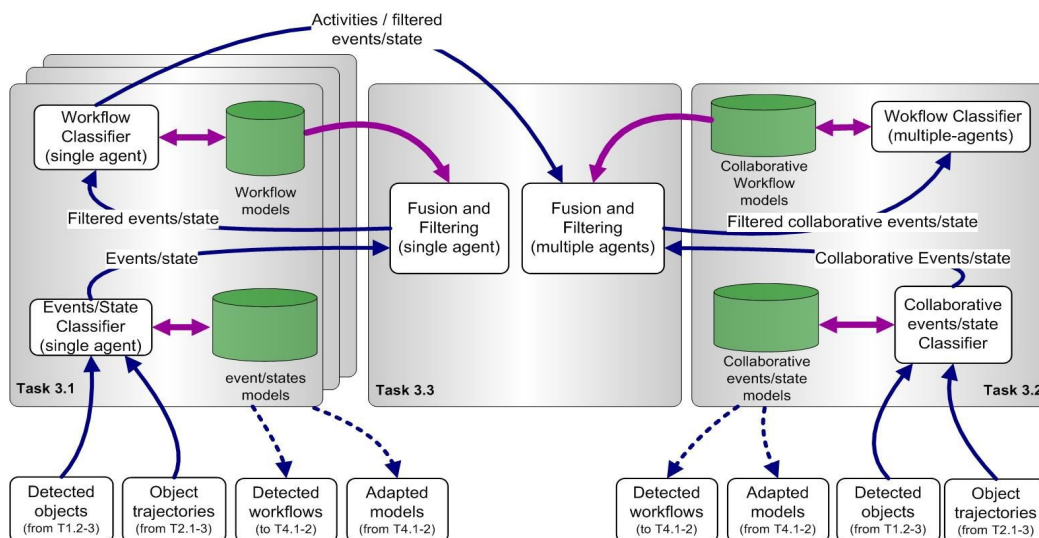
Σχήμα 7. Εικόνες από το εσωτερικό περιβάλλον της αυτοκινητοβιομηχανίας

³ <http://www.scovis.eu/>

Το σύστημα του *SCOVIS* υποστηρίζει πολλαπλές κάμερες δικτύου, έτσι ώστε πολύπλοκες συμπεριφορές να μπορούν να χαρακτηριστούν ως συνδυασμός χώρο-χρονικών (*spatio-temporal*) σχέσεων. Επίσης, η συνεχής παρακολούθηση με πολλαπλές κάμερες απαιτεί σωστή βαθμονόμηση (*camera calibration*), σωστή σύντηξη πληροφοριών (*information fusion*), αποτελεσματική παρακολούθηση σε πραγματικό χρόνο (*real-time tracking*) και συντονισμό κίνησης καμερών στην περίπτωση που υπάρχουν ενεργές κάμερες (*active cameras*). Για αυτό το λόγο, το σύστημα καμερών είναι αυτορυθμιζόμενο, δηλαδή είναι σε θέση να υπολογίσει αυτόματα την θέση των καμερών χωρίς την γνώση χωρικών σχέσεων.

Τα μοντέλα που χρησιμοποιούνται από τις μηχανές ανίχνευσης αντικειμένων μπορούν να προσαρμόζονται και να εμπλουτίζονται αυτόματα στο πέρασμα του χρόνου, μέσω της συλλογής δεδομένων σε πραγματικό χρόνο. Επίσης, για την βελτίωση του μηχανισμού ανίχνευσης αντικειμένων, όπως και για άλλες λειτουργίες, ο χρήστης του συστήματος *SCOVIS* μπορεί να αλληλεπιδρά μέσω των μηχανών ανατροφοδότησης ενδιαφέροντος (*Relevance Feedback*). Με αυτό τον τρόπο, ο χρήστης μπορεί παράλληλα να αξιολογεί και να βελτιώνει την απόδοση του συστήματος.

Σε αυτό το επιστημονικό έργο χρησιμοποιήθηκαν κυρίως ασθενώς εποπτευόμενοι αλγόριθμοι μάθησης (*semi-supervised learning*) και στρατηγικές αυτό-προσαρμογής (*self-adaptation*) με σκοπό την ανάλυση των ροών εργασίας και συμπεριφορών. Συγκεκριμένα, οι ασθενώς εποπτευόμενοι αλγόριθμοι μάθησης κατά το στάδιο της αρχικοποίησης χρησιμοποιούν ένα μικρό αριθμό επισημασμένων δεδομένων (*labeled data*) για την εκπαίδευση του μοντέλου. Κατόπιν, αξιοποιούν ένα μεγάλο αριθμό από δεδομένα χωρίς επισήμανση (*unlabeled data*).



Σχήμα 8. Παρουσίαση του υπό-συστήματος όρασης υπολογιστών

Θα πρέπει να σημειωθεί ότι το πρώτο βήμα για την ερμηνεία μιας σκηνής ή την εξαγωγή συμπερασμάτων και συμπεριφορών γίνεται με την αναγνώριση των αντικειμένων. Ιδιαίτερο ενδιαφέρον σε αυτό το έργο παρουσιάζει η υλοποίηση του υπό-συστήματος όρασης υπολογιστών (βλ. Σχήμα 8) που αποτελείται από μια σειρά εργαλείων, τα οποία στοχεύουν στην αυτοματοποιημένη ανίχνευση και αναγνώριση των αντικειμένων σε μια σκηνή. Στο *SCOVIS* αυτό επιτυγχάνεται με την χρήση χαρακτηριστικών εικόνας (βλ. Κεφάλαιο 2.3.1) τα οποία έχουν υψηλές δυνατότητες εκπροσώπησης. Συγκεκριμένα, μία κατάλληλη προσέγγιση για να καλυφτεί το χάσμα μεταξύ των αντικειμένων και των χαρακτηριστικών εικόνας είναι η εφαρμογή προσεγγίσεων αυτόματης μάθησης.

Η σελίδα αυτή είναι σκόπιμα λευκή.

4

Παρακολούθηση Αντικειμένων

Η παρακολούθηση αντικειμένων (*object tracking*) είναι ένα από τα περισσότερο ενεργά ερευνητικά θέματα στον τομέα της όρασης υπολογιστών. Στην ενότητα αυτή παρουσιάζεται μια λεπτομερής περιγραφή της τρέχουσας βιβλιογραφίας. Συγκεκριμένα, στο Κεφάλαιο 4.1 γίνεται μία αναλυτική καταγραφή των μοντέρνων τεχνικών για την παρακολούθηση αντικειμένων. Κατόπιν, στα Κεφάλαια 4.1.1 και 4.1.2 γίνεται μία παρουσίαση των τεχνικών που βασίζονται σε διαχωριστικούς και παραγωγικούς αλγόριθμους ταξινόμησης. Επίσης, γίνεται η ανάλυση του αλγόριθμου *Boosting* πραγματικού χρόνου (βλ. Κεφάλαιο 4.1.1.1), ο οποίος χρησιμοποιήθηκε στην υλοποίηση του προτεινόμενου συστήματος παρακολούθησης. Τέλος στο Κεφάλαιο 4.2 γίνεται μία παρουσίαση ποικίλων δημοσιεύσεων επάνω στο πρόβλημα της συσχέτισης δεδομένων.

4.1 Τεχνικές παρακολούθησης αντικειμένων

Οι τεχνικές παρακολούθησης στοχεύουν στην εκτίμηση της κίνησης ενός ή περισσότερων αντικειμένων μέσα σε μία σκηνή. Αυτές οι τεχνικές συχνά βασίζονται στις πληροφορίες που προέρχονται από τα κινούμενα αντικείμενα μέσα σε μια σκηνή (π.χ. αυτοκίνητα, ανθρώπους, μηχανήματα). Αυτά τα αντικείμενα συνήθως αλληλεπιδρούν με άλλα αντικείμενα, έτσι ένας από τους βασικούς στόχους είναι η μοντελοποίηση της δυναμικής τους συμπεριφοράς. Η εκτίμηση του πεδίου κίνησης συνήθως πραγματοποιείται με μεθόδους οπτικής ροής (βλ. Σχήμα 9(γ)). Έργα σχετικά με την οπτική ροή προτάθηκαν στις αρχές της δεκαετίας του **80** [35], [8] και ως στόχο είχαν την εύρεση κοινών περιοχών μεταξύ δύο διαδοχικών εικόνων. Με την πάροδο των χρόνων έγινε σημαντική πρόοδος, τόσο στην βελτίωση της υπολογιστικής ταχύτητας [6] όσο και στην αντιμετώπιση του προβλήματος μετατόπισης μίας μεγάλης περιοχής [36]. Πρόσφατα, τεχνικές μάθησης [37] χρησιμοποιήθηκαν προκειμένου να ξεπεραστούν οι περιορισμοί που προκύπτουν από την κλασική μέθοδο οπτικής ροής. Σε γενικές γραμμές, οι τεχνικές οπτικής ροής έχουν πολλές εφαρμογές, όπως η τμηματοποίηση κίνησης (*motion segmentation*) [38], συλλογή στατιστικών στοιχείων από τη σκηνή [39] ή ως διεπαφή ανθρώπου-υπολογιστή (*human-computer interface*) [40].



(α) Ζεύγη Εικόνων



(β) Ανίχνευση Αντικειμένου

(γ) Οπτική Ροή

Σχήμα 9. Παρουσίαση βίντεο που συλλέχθηκε στο Abbey Road του Λονδίνου (α). Ένας ανιχνευτής βάσει εμφάνισης χρησιμοποιήθηκε για τον εντοπισμό συγκεκριμένων ανθρώπων (β). Ωστόσο αυτή η προσέγγιση δεν δίνει πληροφορίες για την κίνηση του ανθρώπου. Η οπτική ροή δεν μπορεί να προσδιορίσει με ακρίβεια την κίνηση του συγκεκριμένου ανθρώπου αφού υπάρχουν διαφορετικές κινήσεις μέσα στην σκηνή.

Αρκετές προσεγγίσεις στην τρέχουσα βιβλιογραφία [17], [18] χρησιμοποιούν τα φίλτρα σωματιδίων (*particle filters*) για την εκτίμηση της κίνησης ενός αντικειμένου σε μια σκηνή. Η τεχνική αυτή εξαρτάται σε μεγάλο βαθμό από την επιλογή της προτεινόμενης κατανομής κίνησης (*proposal distribution*), δηλαδή το μοντέλο κίνησης, (βλ. Κεφάλαιο 2.4.1). Αυτή η κατανομή ουσιαστικά καθοδηγεί τους

μεμονωμένους *trackers* (ή αλλιώς σωματίδια - *particles*) σε περιοχές της σκηνής, όπου θα γίνουν οι μετρήσεις (*visual measurements*) [13], με σκοπό τον εντοπισμό του αντικειμένου. Η επιλογή της κατανομής κίνησης εξαρτάται από το σενάριο στο οποίο εφαρμόζονται τα φίλτρα σωματιδίων. Μία ακατάλληλη επιλογή της κατανομής κίνησης θα μπορούσε να μειώσει ριζικά την αποτελεσματικότητα (βλ. Κεφάλαιο 2.4.1) της δειγματοληψίας (*importance sampling*), με αποτέλεσμα πολλοί *trackers* να χαθούν. Μια συνήθης πρακτική για την διαδικασία της δειγματοληψίας είναι να γίνεται βάσει ενός μοντέλου σταθερής κίνησης [20]. Ωστόσο, η προσέγγιση αυτή μπορεί να επηρεάσει αρνητικά την απόδοση του φίλτρου σωματιδίων, ιδιαίτερα όταν το αντικείμενο που παρακολουθείται παραβιάζει την χρονική συνέχεια (*temporal continuity*), πράγμα το οποίο συμβαίνει στις απότομες ή απρόβλεπτες κινήσεις του αντικειμένου (ή στις περιπτώσεις χαμηλού ρυθμού καρέ, π.χ. *Low Frame Rate*). Αυτός άλλωστε είναι ένας λόγος που το πρόβλημα παρακολούθησης σε περίπλοκες περιστάσεις παραμένει άλυτο.



Σχήμα 10. Εκτίμηση κίνησης ενός συγκεκριμένου ανθρώπου σε δύο εικόνες βίντεο, όπου παραβιάζεται η χρονική συνέχεια (*violation of temporal continuity*). Η εκτίμηση γίνεται με την μέθοδο *gradient based multi-resolution estimation*.

Ωστόσο διάφορες μέθοδοι θα μπορούσαν να συνδυαστούν προκειμένου να αυξηθεί η αποτελεσματικότητα της μεθόδου παρακολούθησης υπό τέτοιες δύσκολες συνθήκες. Μια τεχνική που μπορεί να αντιμετωπίσει ως ένα βαθμό τις χρονικές παραβιάσεις είναι η [23]. Συγκεκριμένα σε αυτήν την εργασία προτείνεται μια μέθοδος παρακολούθησης αντικειμένων, η οποία εισάγει τις μετρήσεις κίνησης σε ένα συγκεκριμένο μοντέλο περιγραφής αντικειμένου μαζί με ένα περιγραφέα χρώματος. Το προτεινόμενο μοντέλο μπορεί να αντιμετωπίσει τις απότομες αλλαγές κίνησης, μέσω ενός μοντέλου κίνησης οπτικής ροής (*gradient based multi-resolution estimation* βλ. Σχήμα 10). Η ίδια φιλοσοφία χρησιμοποιήθηκε και στο [24], όπου μία γεωμετρικά προσδιορισμένη κατανομή κίνησης ενσωματώθηκε σε ένα φίλτρο σωματιδίων. Οι συγγραφείς της δημοσίευσης [66] προτείνουν μια ιδέα για τα σύνθετα μοντέλα κίνησης, Αυτή η νέα παραλλαγή του φίλτρου σωματιδίων ονομάζεται *Multi-Model Particle Filter (MMPF)*. Η γενική αυτή προσέγγιση θα μπορούσε να εφαρμοστεί ως μία λύση σε διάφορα προβλήματα μεγάλων διαστάσεων, όπως η εκτίμηση της ανθρώπινης κίνησης.

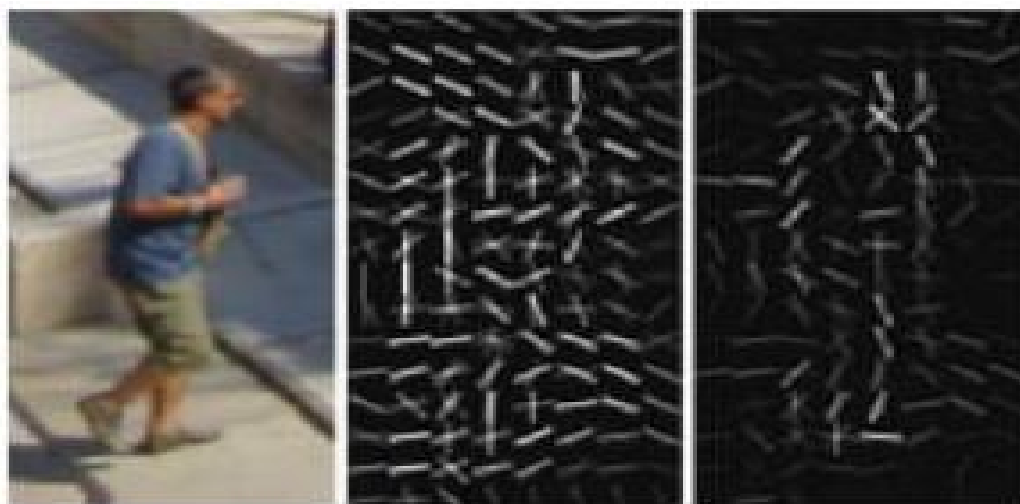
Επίσης, αρκετοί αλγόριθμοι έχουν παρουσιαστεί στον τομέα των *Rao - Blackwell* φίλτρων σωματιδίων. Η εργασία που παρουσιάστηκε στο [33] πρότεινε μια αποτελεσματική μέθοδο εκπροσώπησης ενός υπό-χώρου (*sub-space*) εντός της δομής του φίλτρου σωματιδίων, με σκοπό να ενσωματωθούν οι διαφορετικοί συντελεστές του διανύσματος κατάστασης (*state-vector*). Επιπλέον, στο [34] προτάθηκε ακόμη ένας *Rao - Blackwellized* αλγόριθμος με σκοπό την βελτίωση της αποτελεσματικότητας των φίλτρων σωματιδίων.

Όπως προαναφέρθηκε, η επιλογή του μοντέλου κίνησης έχει σημαντικό ρόλο στην απόδοση του συστήματος παρακολούθησης. Παρόλα αυτά, σημαντικό ρόλο έχει και το μοντέλο εμφάνισης που χρησιμοποιείται για να περιγράψει το αντικείμενο [24].

Αρκετές μέθοδοι παρακολούθησης χρησιμοποιούν ένα συγκεκριμένο μοντέλο χρώματος (είτε *RGB* είτε *HSV*). Σε γενικές γραμμές, αρκετές προσεγγίσεις από αυτές έχουν αποδειχθεί αποτελεσματικές για ορισμένες απλές εφαρμογές (π.χ. *Blob Tracking*). Για παράδειγμα, η παρακολούθηση αντικειμένου βάσει χρώματος είναι ιδιαίτερα χρήσιμη, ιδίως όταν η χρωματική δομή των κινούμενων αντικειμένων παρουσιάζει μικρή διακύμανση. Σε μία από τις πρώτες δημοσιεύσεις σε αυτόν τον τομέα [31] προτάθηκε μια τεχνική *Monte Carlo*, η οποία βασίστηκε στις αποστάσεις μεταξύ δύο ιστογραμμάτων χρώματος (εκφραζόμενες μέσα σε ένα πιθανοτικό πλαίσιο). Θα πρέπει να σημειωθεί ότι η ενσωμάτωση στο σύστημα παρακολούθησης πολλαπλών μοντέλων περιγραφής αντικειμένου, όπως χρώμα, σχήμα, ακμές (*edges*) και υφή (*texture*), μπορεί να συμβάλλει στη περαιτέρω βελτίωση της μεθόδου παρακολούθησης.

Συγκεκριμένα, στο [32] παρουσιάστηκε μια προσέγγιση με στόχο την καλύτερη περιγραφή του αντικειμένου που παρακολουθείται, στην οποία εντάχθηκαν και το χρώμα και η περιγραφή ακμών. Επιπλέον, μια εναλλακτική μέθοδος για την περιγραφή των ακμών μιας εικόνας παρουσιάστηκε στο άρθρο [14]. Ειδικότερα, στο πλαίσιο αυτό ο κινούμενος στόχος αποδίδεται μέσω ενός ιστογράμματος χρώματος (*Color Histogram*) παράλληλα με ένα βελτιωμένο ιστόγραμμα, το οποίο περιέχει πληροφορίες για τον προσανατολισμό της κλίσης των ακμών (βλ. Σχήμα 11, *Histogram of oriented gradients, HOG*). Θα πρέπει να σημειωθεί ότι ο περιγραφέας αντικειμένου *HOG* είναι από τους πιο ανθεκτικούς περιγραφείς, ειδικά στις αλλαγές φωτισμού και στις αλλαγές στο φόντο. Επίσης, το άρθρο [63] παρουσιάζει την ένταξη απλών χρωματικών κατανομών σε φίλτρα σωματιδίων, παράλληλα με μια μέθοδο για το πρότυπο προσαρμογής (*template matching*). Επίσης, η μέθοδος αυτή προτείνει μία

νέα προσέγγιση επαναρχικοποίησης, λόγω του γεγονότος ότι τα αντικείμενα που παρακολουθούνται μπορούν να φύγουν από την σκηνή και να εμφανιστούν εκ νέου.



Σχήμα 11. Παρουσίαση των δυο βασικών κλίσεων ακμών (κεντρική και δεξιά εικόνα) που προέρχονται από μια αρχική εικόνα⁴ (αριστερή εικόνα). Ο περιγραφέας *HOG* δημιουργείται με την βοήθεια αυτής της πληροφορίας.

Ωστόσο, οι παραλλαγές στην εμφάνιση ενός αντικείμενου μπορούν να προσδιοριστούν καλύτερα από μία προσέγγιση ταξινόμησης (βλ. Σχήμα 9(β)). Για αυτό το λόγο μία συνήθης πρακτική είναι ο συνδυασμός ενός ταξινομητή-ανιχνευτή (*detector*) με ένα *tracker* [21], [22]. Τα τελευταία χρόνια, υπήρξε μια τάση θέσπισης τεχνικών ταξινόμησης σε αλγορίθμους παρακολούθησης.

Επίσης, τα γραφικά μοντέλα (*graphical models*) αποτελούν μία σημαντική κατηγορία προσεγγίσεων μάθησης στη σύγχρονη βιβλιογραφία. Ειδικότερα, τα πιθανοτικά γραφικά μοντέλα (*probabilistic graphical models*) μπορούν να χρησιμοποιηθούν για την απεικόνιση της κατανομής πιθανοτήτων. Τα τυχαία πεδία Μαρκόφ (*Markov Random Fields MRF*), είναι ένα σημαντικό υποσύνολο γραφικών μοντέλων, τα οποία μπορούν να χρησιμοποιηθούν για να εκφράσουν τους χαλαρούς περιορισμούς μεταξύ

⁴ <http://people.seas.harvard.edu/~ely/faceparts/serial.html>

κάποιων τυχαίων μεταβλητών [67]. Συγκεκριμένα, η μέθοδος [68] χρησιμοποιεί ένα τυχαίο πεδίο Μαρκόφ για να περιγράψει ένα μοντέλο κίνησης. Αυτό το *MRF* θα μπορούσε να χρησιμοποιηθεί για να περιγράψει τις αλληλεπιδράσεις μεταξύ πολλαπλών στόχων. Ωστόσο οι βασικές τεχνικές ταξινόμησης χωρίζονται σε παραγωγικές (*generative*) και διαχωριστικές (*discriminative*) (βλ. Κεφάλαιο 2.1.1).

4.1.1 Διαχωριστικοί αλγόριθμοι για τεχνικές παρακολούθησης

Οι διαχωριστικοί αλγόριθμοι [26] απλοποιούν την διαδικασία εξαγωγής συμπερασμάτων, εστιάζοντας την οικοδόμηση ενός μοντέλου που προβλέπει απευθείας την ύπαρξη του αντικειμένου (δηλαδή, αν ένα δείγμα εκπαίδευσης είναι το κατάλληλο ή όχι). Οι μέθοδοι παρακολούθησης μέσω ανίχνευσης (*tracking-by-detection*) εμπίπτουν σε αυτήν την κατηγορία, δεδομένου ότι οι διαχωριστικοί ταξινομητές (*discriminative classifiers*) εκπαιδεύονται με σκοπό να διαχωρίσουν το αντικείμενο από τον περιβάλλοντα χώρο. Σε αυτήν την κατεύθυνση [18], προτάθηκε μια αποτελεσματική τεχνική που ονομάζεται *boosted* φίλτρα σωματιδίων (*boosted particle filters*). Η τεχνική αυτή συνδυάζει τα πλεονεκτήματα των φίλτρων σωματιδίων και της μεθόδου *Adaboost* με σκοπό την παρακολούθηση πολλαπλών παιχτών χόκεϊ. Επίσης, ένας διαχωριστικός αλγόριθμος μάθησης, ο οποίος ενσωματώνεται σε ένα σύστημα παρακολούθησης, περιγράφεται στο [20] και χρησιμοποιεί την τεχνική του *Multiple Instance Learning (MIL)*. Επίσης, έχουν προταθεί αρκετές μέθοδοι στην βιβλιογραφία, οι οποίες ενημερώνουν τον ταξινομητή σε πραγματικό χρόνο, για τις αλλαγές που γίνονται στην εμφάνιση του αντικειμένου. Ωστόσο αρκετές από αυτές συναντούν δυσκολίες ιδιαίτερα στην περίπτωση που χρησιμοποιούνται στρατηγικές αυτό-μάθησης [41] (*self-learning strategies*). Ειδικά στον τομέα εφαρμογών πραγματικού χρόνου ιδιαίτερο ενδιαφέρον παρουσιάζει η τεχνική του *Grabner* [4], η οποία αναλύεται παρακάτω.

4.1.1.1 Μέθοδος *Boosting* πραγματικού χρόνου

Σε αυτήν την εργασία παρουσιάστηκε ένας γενικός αλγόριθμος ταξινόμησης, που επιτρέπει την επιλογή χαρακτηριστικών εικόνας (βλ. Κεφάλαιο 2.3.1) σε πραγματικό χρόνο, χρησιμοποιώντας ως βάση την μέθοδο *Boosting*. Επίσης μπορεί να διαχειριστεί μεγάλες συλλογές χαρακτηριστικών εικόνας (*large feature pool*) με μειωμένο υπολογιστικό κόστος. Ως εκ τούτου αυτή η μέθοδος δίνει την δυνατότητα να αντιμετωπίσει μια μεγάλη ποικιλία διαφορετικών εφαρμογών σε πραγματικό χρόνο.

Για να δημιουργηθεί μία τέτοια προσέγγιση πρέπει όλα τα βήματα του αλγορίθμου (βλ. Κεφάλαιο 2.2) να γίνονται σε πραγματικό χρόνο. Αυτό σημαίνει ότι ο αλγόριθμος μπορεί να ανανεώνεται με ένα και μόνο δείγμα εκπαίδευσης και να το απορρίπτει κατόπιν ενημέρωσης του. Θα πρέπει να σημειωθεί, ότι η ενημέρωση των ταξινομητών «βάσης» δεν είναι το πρόβλημα, διότι πολλές μέθοδοι μπορούν να χρησιμοποιηθούν για την δημιουργία υποθέσεων (*Hypothesis* βλ. Κεφάλαιο 2.2). Το δύσκολο βήμα είναι ο υπολογισμός της κατανομής βάρους για τα δείγματα, γιατί δεν μπορεί κάποιος να γνωρίζει εκ των προτέρων τη δυσκολία εκπαίδευσης για ένα συγκεκριμένο δείγμα.

Ο *Grabner* [4] χρησιμοποίησε τις ιδέες του *Oza* [57] και τις πειραματικές συγκρίσεις που έκανε [58]. Η βασική ιδέα αυτής της προσέγγισης είναι ότι η δυσκολία ενός δείγματος εκπαίδευσης μπορεί να μεταδοθεί στο σύνολο των ταξινομητών «βάσης». Συγκεκριμένα, η μοντελοποίηση του κέρδους της νέας πληροφορίας (δηλαδή, του νέου δείγματος) για τον n ταξινομητή «βάσης», μπορεί να χρησιμοποιηθεί για να ανανεώσει τον $n + 1$ ταξινομητή «βάσης». Αυτή η μοντελοποίηση είναι κωδικοποιημένη μέσω του *importance weight* λ . Θα πρέπει να σημειωθεί ότι ο

αλγόριθμος *Boosting* πραγματικού χρόνου προϋποθέτει ότι ο αριθμός των ταξινομητών «βάσης» είναι προκαθορισμένος. Αυτό έρχεται σε αντίθεση με τον αλγόριθμο *Boosting* μη-πραγματικού χρόνου, όπου το σύνολο των δειγμάτων εκπαίδευσης είναι σταθερό και ο αριθμός των ταξινομητών «βάσης» μεταβλητός.

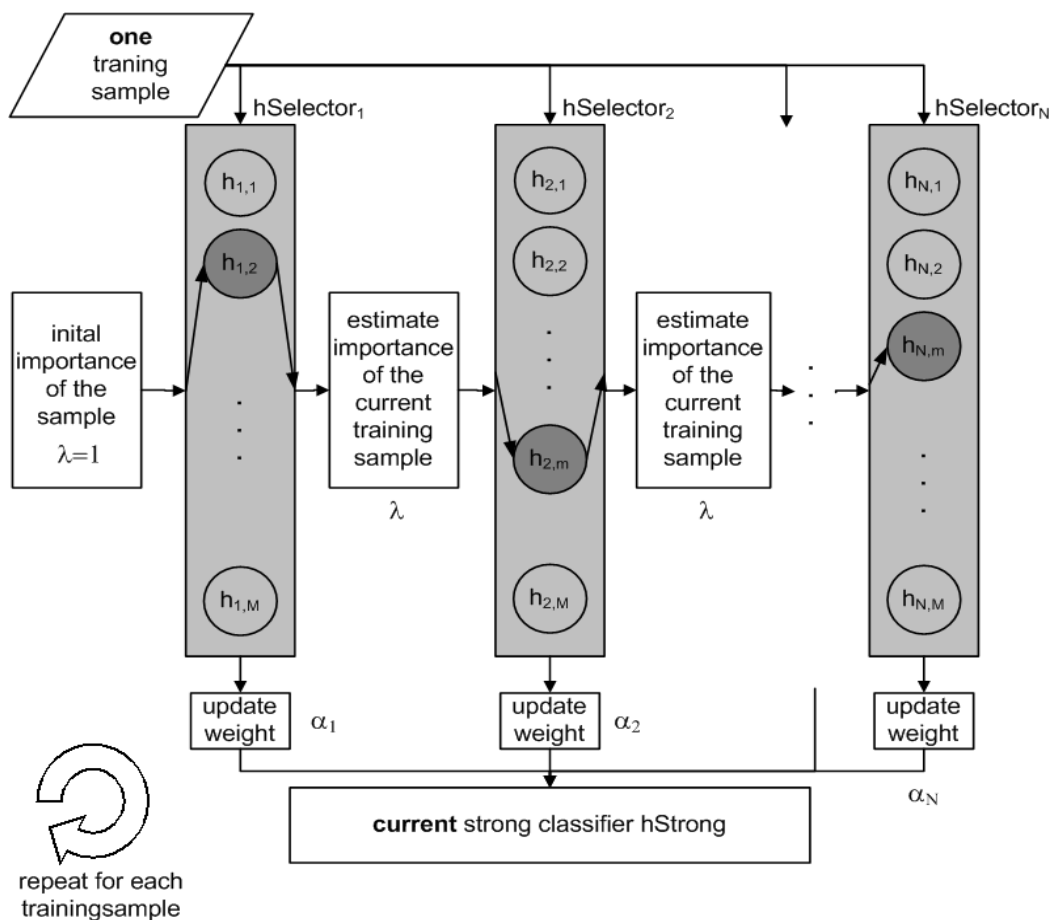
Η βασική καινοτομία αυτής της προσέγγισης είναι ότι προτείνεται ένας αλγόριθμος *Boosting* πραγματικού χρόνου για την εύρεση χαρακτηριστικών εικόνας (βλ. Κεφάλαιο 2.3.1). Για αυτό το λόγο ορίζονται οι παρακάτω έννοιες.

- Επιλογέας (Selector): Έχοντας M ταξινομητές «βάσης» με τις αντίστοιχες υποθέσεις $H^{weak} = \{h_1^{weak}, \dots, h_M^{weak}\}$, ο επιλογέας (*selector*) διαλέγει ακριβώς έναν από αυτούς,

$$h^{sel}(x) = h_m^{weak}(x) \quad (4.1)$$

όπου ο m ταξινομητής «βάσης» επιλέγεται βάσει ενός κριτηρίου βελτιστοποίησης. Συγκεκριμένα, χρησιμοποιείται το προσδοκώμενο σφάλμα e_i του κάθε ταξινομητή «βάσης» $h_i^{weak} \in H^{weak}$ έτσι ώστε $m_i = \underset{i}{\operatorname{argmin}} e_i$.

Θα πρέπει να σημειωθεί ότι ο επιλογέας (*selector*) μπορεί να θεωρηθεί ως ένας ταξινομητής (*classifier*), ο οποίος επιλέγει τους αντίστοιχους ταξινομητές «βάσης». Η κατάρτιση ενός επιλογέα σημαίνει ότι κάθε ταξινομητής «βάσης» εκπαιδεύεται (ή ενημερώνεται) και επιλέγεται ο καλύτερος (δηλαδή, εκείνος με τη χαμηλότερη εκτίμηση σφάλματος). Όπως και με την περίπτωση *Adaboost* [3], οι ταξινομητές «βάσης» H^{weak} αντιστοιχούν σε ένα χαρακτηριστικό εικόνας. Ο επιλογέας μπορεί να διαλέξει μέσα από ένα υπό-σύνολο χαρακτηριστικών εικόνας $F_{sub} = \{f_1, \dots, f_M | f_i \in F\}$. Η βασική ιδέα είναι η εφαρμογή της προσέγγισης *Adaboost* [3] επάνω στους επιλογείς (*Selectors*).



Σχήμα 12. Παρουσίαση της μεθόδου Boosting πραγματικού χρόνου [4].

Τα βήματα για τον παραπάνω αλγόριθμο είναι ως εξής. Αρχικά ένα ορισμένο σύνολο N από επιλογείς $h_1^{set}, \dots, h_N^{set}$ αρχικοποιείται τυχαία, όπου ο κάθε επιλογέας διαθέτει το δικό του σύνολο χαρακτηριστικών εικόνας. Όταν ένα νέο δείγμα εκπαίδευσης μαζί με την ετικέτα του (x, y) εισέρχεται στον αλγόριθμο οι επιλογείς ανανεώνονται. Η ανανέωση γίνεται βάσει του *importance weight* λ του δείγματος εκπαίδευσης. Για την ανανέωση των ταξινομητών «βάσης» οποιαδήποτε τεχνική μάθησης πραγματικού χρόνου μπορεί να χρησιμοποιηθεί. Ο Grabner [4] χρησιμοποίησε τον βασικό *Expectation Maximization (EM)* αλγόριθμο με σκοπό να υπολογίσει τις πιθανοτικές

κατανομές των θετικών και των αρνητικών δειγμάτων για να παράγει μία συγκεκριμένη υπόθεση (*Hypothesis*). Τότε ο επιλογέας διαλέγει τον ταξινομητή «βάσης» με το χαμηλότερο σφάλμα,

$$\mathit{argmin}_m(\mathbf{e}_{n,m}), \mathbf{e}_{n,m} = \frac{\lambda_{n,m}^{wrong}}{\lambda_{n,m}^{corr} + \lambda_{n,m}^{wrong}} \quad (4.2)$$

Θα πρέπει να σημειωθεί ότι $\mathbf{e}_{n,m}$ είναι το σφάλμα του m ταξινομητή βάσης $h_{n,m}^{weak}$ στο n επιλογή, το οποίο έχει υπολογιστεί από τα βάρη των σωστών ταξινομημένων δειγμάτων $\lambda_{n,m}^{corr}$ και των εσφαλμένων ταξινομημένων δειγμάτων $\lambda_{n,m}^{wrong}$. Τέλος, το αντίστοιχο *voting weight* α_n και *importance weight* λ του δείγματος ανανεώνεται και περνάει στον επόμενο επιλογέα h_{n+1}^{sel} . Αυτή η διαδικασία επαναλαμβάνεται για όλους τους επιλογείς. Ο τελικός ταξινομητής αυτής της προσέγγισης υπολογίζεται από τον γραμμικό συνδυασμό των επιλογέων.

$$h^{strong}(x) = \mathit{sign} \left(\sum_{n=1}^N \alpha_n h_n^{sel}(x) \right) \quad (4.3)$$

4.1.2 Παραγωγικοί αλγόριθμοι για τεχνικές παρακολούθησης

Από την άλλη πλευρά, τα συστήματα παρακολούθησης που βασίζονται σε παραγωγικές μεθόδους [27] αντιμετωπίζουν το πρόβλημα της ταξινόμησης κατά τρόπο εποικοδομητικό. Αυτό επιτυγχάνεται με την εκμάθηση ενός μοντέλου (*prior model*) για την εμφάνιση του αντικειμένου πριν τεθεί σε λειτουργία το σύστημα παρακολούθησης. Το συμπέρασμα για την ύπαρξη του αντικειμένου προκύπτει όταν η τρέχουσα εκτίμηση εξακολουθεί να είναι έγκυρη βάσει του μοντέλου (*prior model*) για την εμφάνιση του αντικειμένου. Πιο συγκεκριμένα, ένα μοντέλο εμφάνισης

μπορεί να οριστεί με παραγωγικές προσεγγίσεις όπως είναι η ανάλυση σε κύριες συνιστώσες (*PCA*) [28]. Στη συνέχεια, θορυβώδη χαρακτηριστικά μπορούν να ανιχνευθούν με την προβολή των δεδομένων σε ίδιο-εικόνες (*eigenimages*). Ωστόσο, στη βιβλιογραφία έχουν προταθεί πολύ πιο αποτελεσματικές προσεγγίσεις. Ειδικότερα, ο *Leonardis* [27] πρότεινε ένα εναλλακτικό τρόπο για τον καθορισμό των συντελεστών των ίδιο-εικόνων χρησιμοποιώντας την μέθοδο της υπό-δειγματοληψίας (*sub-sampling*). Επίσης, ο *Woodley* [29] παρουσίασε ένα παραγωγικό μοντέλο υπολογίζοντας τις τοπικές μη-αρνητικές μήτρες (*Local Non-negative Matrix Factorization, LNMF*) με σκοπό να αποφευχθούν οι θορυβώδεις περιοχές σε μία εικόνα (*noisy areas*). Ανάλογη προσέγγιση μπορεί να επιτευχθεί μέσω της διαδικασίας αραιής βελτιστοποίησης (*sparse optimization*) [30]. Θα πρέπει όμως να σημειωθεί ότι οι παραπάνω γραμμικές παραγωγικές μέθοδοι ταξινόμησης δεν μπορούν να αντιμετωπίσουν δραματικές αλλαγές στην εμφάνιση του αντικειμένου που παρακολουθείται.

Μία παραγωγική μέθοδος που μπορεί να αντιμετωπίσει τέτοια προβλήματα είναι η *GPLVM* [59], η οποία ανακαλύπτει χαμηλών διαστάσεων υπό-επίπεδα (*low dimensional manifold*) χρησιμοποιώντας ένα μικρό αριθμό δειγμάτων κατάρτισης. Αυτή η προσέγγιση αναλύεται με περισσότερες λεπτομέρειες στο κεφάλαιο 6.1.3.1.

4.2 Συσχέτιση Δεδομένων

Το πρόβλημα της συσχέτισης δεδομένων είναι κομβικής σημασίας για τα σύγχρονα συστήματα παρακολούθησης. Θα πρέπει να σημειωθεί ότι, οι τεχνικές συσχέτισης δεδομένων χρησιμοποιούνται κυρίως στις εφαρμογές ανάκτησης και οργάνωσης πολυμέσων (*multimedia content-based retrieval*). Επίσης, η συνεχόμενη δημιουργία πολυμεσικών εφαρμογών ολοένα και αυξάνει το ερευνητικό ενδιαφέρον, για την δημιουργία αξιόπιστων τεχνικών οι οποίες ως στόχο έχουν την γρήγορη και αυτόματη οργάνωση των πολυμεσικών δεδομένων. Αυτές οι τεχνικές χρησιμοποιούν διάφορους αλγορίθμους που στόχο έχουν την οργάνωση μιας βάσης δεδομένων που αποτελείται από πολυμεσικά δεδομένα (π.χ. εικόνες) σε ομάδες που μοιράζονται κοινά χαρακτηριστικά.

Συγκεκριμένα, σε ένα σύστημα το οποίο αποτελείται από ένα μεγάλο αριθμό πολυμεσικών δεδομένων, η εφαρμογή αλγορίθμων ομαδοποίησης παίζει ιδιαίτερα σημαντικό ρόλο, ειδικά στην περίπτωση όπου ο χρήστης επιθυμεί να αναζητήσει γρήγορα τα δεδομένα που τον αφορούν. Η αρχική περιγραφή των δεδομένων μπορεί να γίνει είτε με την βοήθεια των πρωτογενών χαρακτηριστικών (*low level descriptors*), όπως χρώμα, μέγεθος αντικειμένων, είτε με την βοήθεια των σημασιολογικών χαρακτηριστικών (*high level descriptors*) όπως για παράδειγμα άνθρωποι, κτήρια, αυτοκίνητα.

Η απλούστερη προσέγγιση για την εύρεση παρόμοιων πολυμεσικών δεδομένων σε μια βάση είναι η σειριακή αναζήτηση. Με τον τρόπο αυτόν, όταν γίνεται μια επερώτηση (*image query*) σε μία εικόνα (ή ένα τμήμα βίντεο), όλοι οι δείκτες της αντίστοιχης βάσης δεδομένων ψάχνουν όλες τις καταχωρήσεις με σειριακό τρόπο. Ωστόσο, αυτή η απλή προσέγγιση εμφανίζει περιορισμούς στον χρόνο απόκρισης

ανάκτησης και στην επεκτασιμότητα, ιδιαίτερα στην περίπτωση μεγάλων και κατανεμημένων πολυμεσικών βάσεων. Έτσι, ο χρόνος που απαιτείται για την εύρεση μιας εικόνας ή ενός τμήματος βίντεο μπορεί να είναι ασύμφορος. Συγκεκριμένα, οι κλασσικές εφαρμογές στατικής δεικτοδότησης (*static indexing*) προσφέρουν χαμηλή απόδοση, όσον αφορά στην υπολογιστική πολυπλοκότητα και την επεκτασιμότητα.

Θα πρέπει να τονιστεί ότι η δεικτοδότηση πολυμέσων (*multimedia indexing*) θα πρέπει να γίνεται με έναν τρόπο έτσι ώστε να υποστηρίζεται *α)* δυναμική ενημέρωση των δεικτοδοτημένων δεδομένων, *β)* χαμηλή υπολογιστική πολυπλοκότητα στη δημιουργία των δεικτών. Οι αλγόριθμοι που ασχολούνται με την οργάνωση του περιεχομένου θα πρέπει να έχουν την ικανότητα να προσθέτουν ή/και να αφαιρούν περιεχόμενο από την βάση δεδομένων δυναμικά. Συνεπώς, η επεκτασιμότητα σε τέτοιες εφαρμογές αποκτά ιδιαίτερη σημασία. Προκειμένου να ξεπεραστεί το πρόβλημα αυτό, διάφορες προσεγγίσεις για οργάνωση πολυμέσων βάσει περιεχομένων έχουν υιοθετηθεί [42], [43], [44], [45]. Ειδικότερα, ο αλγόριθμος που παρουσιάζεται στην [46] χρησιμοποιεί μία *KD* δενδρική δομή, η οποία διαχωρίζει έναν συγκεκριμένο χώρο χαρακτηριστικών σε προκαθορισμένα υπέρ-επίπεδα, ασχέτως της κατανομής των διανυσμάτων χαρακτηριστικών. Επιπροσθέτως, η μέθοδος που περιγράφεται στην [47] χρησιμοποιεί μία δομή *R*-δένδρου, στην οποία ο χώρος χαρακτηριστικών διαχωρίζεται αναλόγως με την κατανομή των αντικειμένων της βάσης δεδομένων. Μια επέκταση της μεθόδου *R*-δένδρου μπορεί να βρεθεί στην [48], η οποία ονομάζεται *R**-δένδρο. Η προσέγγιση οργάνωσης *X*-δένδρου προτείνεται στην [49] και στοχεύσει στον διαχωρισμό δεδομένων πολλών διαστάσεων.

Επιπρόσθετα, εξειδικευμένες μέθοδοι για την γρήγορη αναζήτηση δεδομένων βίντεο έχουν επίσης προταθεί στη βιβλιογραφία. Συγκεκριμένα, μια νέα ιεραρχική μέθοδος

αναζήτησης παρουσιάζεται στην [50]. Στην εργασία αυτή η οργάνωση δεδομένων πραγματοποιείται χρησιμοποιώντας μια προσέγγιση γράφου δενδρικής δομής. Αυτή η τεχνική παρέχει αποδοτικά αποτελέσματα μόνο στην περίπτωση κατά την οποία υπάρχει στη βάση παρόμοιο περιεχόμενο, όπως περιεχόμενο από ακολουθίες βίντεο. Στην [51] προτείνεται μια τεχνική Αρχείου Προσέγγισης Ταξινομημένου Διανύσματος (*Ordered Vector Approximation File*), η οποία καλείται *OVA-File*. Αυτή η προσέγγιση διαχωρίζει τα αρχεία χαρακτηριστικών σε φέτες (*slices*) τέτοιες ώστε μόνο ένας μικρός αριθμός κομματιών είναι προσβάσιμος και ελέγχονται κατά την αναζήτηση του *k*-Κοντινότερου Γείτονα (*k-Nearest Neighbor - kNN*). Η δουλειά στην [51] παρουσιάζει εισαγωγές νέων διανυσμάτων σε ένα *OVA-File* με την ελαχιστοποίηση της μέσης απόστασης ανάμεσα σε νέα διανύσματα και τις προσεγγίσεις τους. Το *VigSig* σύστημα [52] χρησιμοποιεί ένα πρωτότυπο σχήμα μη γραμμικής εξαγωγής χαρακτηριστικών σε αυθαίρετους μετρικούς χώρους που συνδυάζει την τριγωνική ανισότητα με την κλασική ανάλυση σε κύριες συνιστώσες (*Principal Component Analysis – PCA* βλ. Κεφάλαιο 4.1.2). Επίσης προκειμένου να επιτευχθεί περαιτέρω βελτίωση της απόδοσης ανάκτησης, η [52] εισαγάγει έναν νέο γράφο-θεωρητικό αλγόριθμο για συστοιχίες (*clustering*) κατάλληλο για μεγάλες βάσεις δεδομένων. Ένα Ιεραρχικό Κυτταρικό Δένδρο (*Hierarchical Cellular Tree-HCT*) παρουσιάζεται στην [53] με στόχο να προσφέρει μία αποτελεσματική λύση, ειδικά για τη δεικτοδότηση μεγάλων βάσεων πολυμεσικών δεδομένων. Η *HCT* αποτελεί μία δυναμική, βασισμένη σε κελιά, και ιεραρχικά δομημένη μέθοδο δεικτοδότησης, η οποία είναι σκοπίμως σχεδιασμένη για πράξεις επερωτήσεων-σημείων και ικανότητες προηγμένης πλοήγησης σε μεγάλες βάσεις δεδομένων. Στην [54] παρουσιάζεται μία πρωτότυπη μέθοδος δεικτοδότησης βασισμένη σε υπέρ-επίπεδα για την υποστήριξη αποδοτικής επεξεργασίας επερωτήσεων αναζήτησης ομοιοτήτων σε χώρους μεγάλων διαστάσεων. Η κεντρική ιδέα του προτεινόμενου

σχήματος είναι η βελτίωση της απόδοσης του διαχωρισμού δεδομένων σε έναν πολυδιάστατο χώρο χρησιμοποιώντας ένα υπέρ-επίπεδο. Η μέθοδος αυτή παρέχει επιπρόσθετα τμήματα (*partitions*) ενός υπό-χώρου και μπορεί να χρησιμοποιηθεί προς εκμετάλλευση της έννοιας του κόμβου.

Οι *MPEG-7* περιγραφείς χρησιμοποιούνται για την αναπαράσταση του οπτικού περιεχομένου (χρώμα, υφή, σχήμα και κίνηση) σε πολυμεσικές βάσεις δεδομένων. Το πρότυπο *MPEG-7* εκμεταλλεύεται την διαλειτουργικότητα, μεταφερσιμότητα και προσβασιμότητα της αναπαραστώμενης πληροφορίας της εικόνας [56]. Η αποδοτικότητα του κάθε περιγραφέα ποικίλλει ανάλογα με το σύνολο των δεδομένων. Όμως, η υλοποίηση ενός μεγάλου αριθμού *MPEG-7* περιγραφέων είναι μη αποδοτική όσον αφορά την υπολογιστική πολυπλοκότητα όπως και για την ανάγκη ενός συστήματος που θα λειτουργεί σε πραγματικό χρόνο. Συγκεκριμένα, ο *CLD* είναι ιδανικός για την ανάκτηση εικόνων χάρη στο μέγεθός του και στο γεγονός ότι αποδίδει καλά σε διαφορετικά σύνολα δεδομένων. Επιπροσθέτως, αποτελεί έναν από τους πιο έξυπνους περιγραφείς, διότι όχι μόνο περιγράφει χρώματα, αλλά εντοπίζει και τη θέση τους στην εικόνα [55]. Οι *DCD* και *HTD* είναι ιδανικοί για την περιγραφή απλών και περίπλοκων σχημάτων και υφών σε μια εικόνα. Σύμφωνα με το [55] οι καλύτεροι περιγραφείς για χρήση συνδυασμών είναι οι *CLD*, *EHD*, *RSD*, *DCD* και *HTD*.

Η σελίδα αυτή είναι σκόπιμα λευκή.

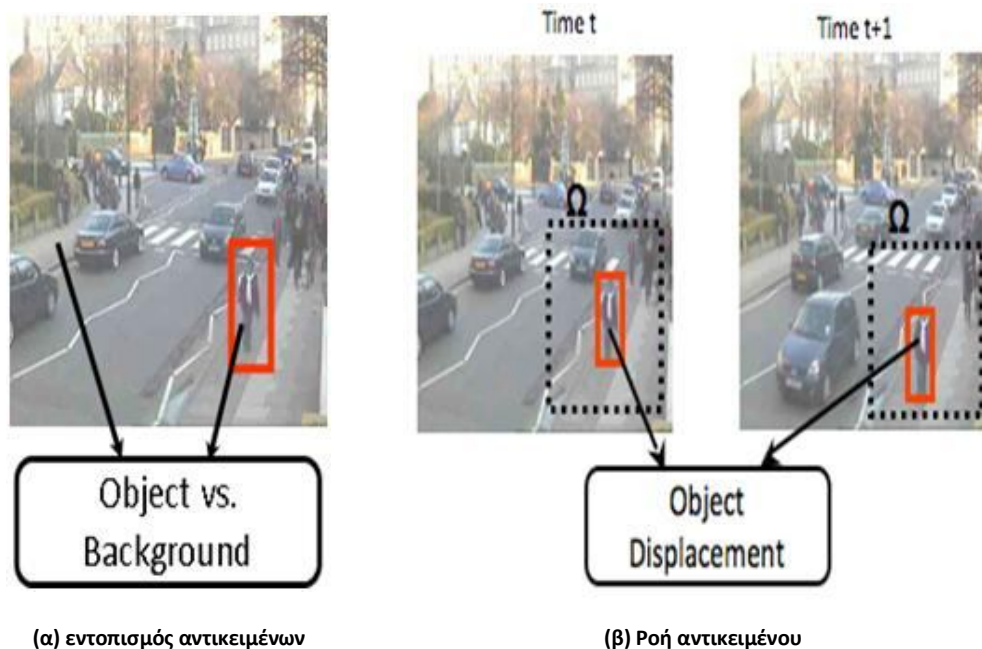
5

Ροή αντικειμένου

Στο κεφάλαιο αυτό προτείνεται μια καινοτόμος μέθοδος για τον υπολογισμό της μετατόπισης ενός αντικειμένου σε μια εικονοσειρά. Η τεχνική αυτή ονομάζεται ροή αντικειμένου (*object flow*) και έχει την ικανότητα να προσδιορίζει το πεδίο κίνησης, με παρόμοιο τρόπο με την οπτική ροή ενός συγκεκριμένου κινούμενου αντικειμένου. Παράλληλα μπορεί να απορρίπτει άλλες άσχετες κινήσεις στο εσωτερικό μιας σκηνής (π.χ. άλλα κινούμενα αντικείμενα ή κινούμενο φόντο). Συγκεκριμένα στο Κεφάλαιο 5.1 γίνεται η διατύπωση του προβλήματος μάθησης. Επίσης, στο Κεφάλαιο 5.2 παρουσιάζεται ο τρόπος που εκτιμάται η ροή αντικειμένου και τα αποτελέσματα για πεζούς σε μια πολυσύχναστη σκηνή παρουσιάζονται στο Κεφάλαιο 5.3.1. Στο Κεφάλαιο 5.3.2 γίνεται μία ποιοτική παρουσίαση των αποτελεσμάτων ροής αντικειμένου για διαφορετικά κινούμενα αντικείμενα. Τέλος στο Κεφάλαιο 5.4 γίνεται μία ποσοτική αξιολόγηση της προτεινόμενης προσέγγισης.

5.1 Διατύπωση προβλήματος μάθησης

Ο στόχος των παραδοσιακών μηχανισμών ανίχνευσης (*object detection*) είναι ο εντοπισμός συγκεκριμένων αντικειμένων σε μια εικόνα. Στις περισσότερες μεθόδους, η εκπαίδευση του ταξινομητή (*classifier*) γίνεται με τα με τα κατάλληλα δείγματα (θετικά και αρνητικά), με σκοπό να εντοπιστεί το αντικείμενο ενδιαφέροντος από παρασκήνιο (*background*). Κάτι τέτοιο θα μπορούσε να διατυπωθεί και ως ένα δυαδικό πρόβλημα ταξινόμησης. Ωστόσο, σε σύγκριση με τους μηχανισμούς ανίχνευσης αντικειμένου, το πρόβλημα της ανίχνευσης της μετατόπισης και της κατεύθυνσης αντίστοιχα γίνεται σε μια συγκεκριμένη περιοχή αναζήτησης Ω (βλ. Σχήμα 13). Μέσα σε αυτή την περιοχή, ζεύγη που αποτελούνται από υπό-εικόνες και βρίσκονται σε διάφορα χρονικά διαστήματα αναγνωρίζονται από τον ταξινομητή (*classifier*) με σκοπό να υπολογιστεί η μετατόπιση και η κατεύθυνση ενός συγκεκριμένου αντικειμένου. Για αυτό το λόγο, το μέγεθος της περιοχής αναζήτησης Ω έχει ιδιαίτερη σημασία. Ειδικά στην περίπτωση απότομης κίνησης ή του χαμηλού ρυθμού καρτέ (*low frame rate*) βέλτιστος υπολογισμός μπορεί να επιτευχθεί μόνο με την βοήθεια μιας αρκετά μεγάλης περιοχής αναζήτησης Ω . Ωστόσο, το μέγεθος αυτής έρχεται σε αντίθεση με την απαιτούμενη υπολογιστική πολυπλοκότητα και θα μπορούσε να δημιουργήσει αρκετά προβλήματα ειδικά όταν περισσότερα από ένα αντικείμενα βρίσκονται μέσα στη σκηνή.



(α) εντοπισμός αντικειμένων

(β) Ροή αντικειμένου

Σχήμα 13. Η ανίχνευσης αντικειμένων (α) συνήθως διατυπώνεται ως ένα δυαδικό πρόβλημα ταξινόμησης μιας και ο στόχος είναι διακριθεί το αντικείμενο του ενδιαφέροντος από τον περιβάλλοντα χώρο. Αντίθετα, η μέθοδος ροής αντικειμένου θεωρεί ως πρόβλημα μάθησης την μετατόπιση του αντικειμένου σε τοπικό επίπεδο.

Διατύπωση προβλήματος. Το πρόβλημα μάθησης διατυπώνεται ως την εκμάθηση μιας συνάρτησης απόστασης, (*distance function* - βλ. [1] για μια πρόσφατη επισκόπηση). Η τεχνική βασίζεται στο έργο του Hertz [2]. Θα πρέπει να σημειωθεί ότι η εκπαίδευση μιας συνάρτησης απόστασης μπορεί να επιτευχτεί με την βοήθεια ενός δυαδικού ταξινομητή (π.χ., *Binary Classifier - Support Vector Machines* ή μέθοδος *Boosting*), χρησιμοποιώντας τα κατάλληλα ζεύγη των δειγμάτων. Τα ζευγάρια που περιέχουν θετικά δείγματα προέρχονται από την ίδια κατηγορία, ενώ τα ζευγάρια που περιέχουν αρνητικά δείγματα προέρχονται από δυο διαφορετικές κατηγορίες. Συγκεκριμένα, το πρόβλημα μάθησης μπορεί να διατυπωθεί σε πεδίο γινομένων (*product space*) ως εξής, $\mathbf{C}: \mathbf{X} \times \mathbf{X} \rightarrow \mathbf{Y} = [-1,1]$. Σε αυτή την περίπτωση ένας εκπαιδευμένος ταξινομητής $\mathbf{C}(x_1, x_2)$ θα παρουσιάσει μεγάλο βαθμό απόκρισης

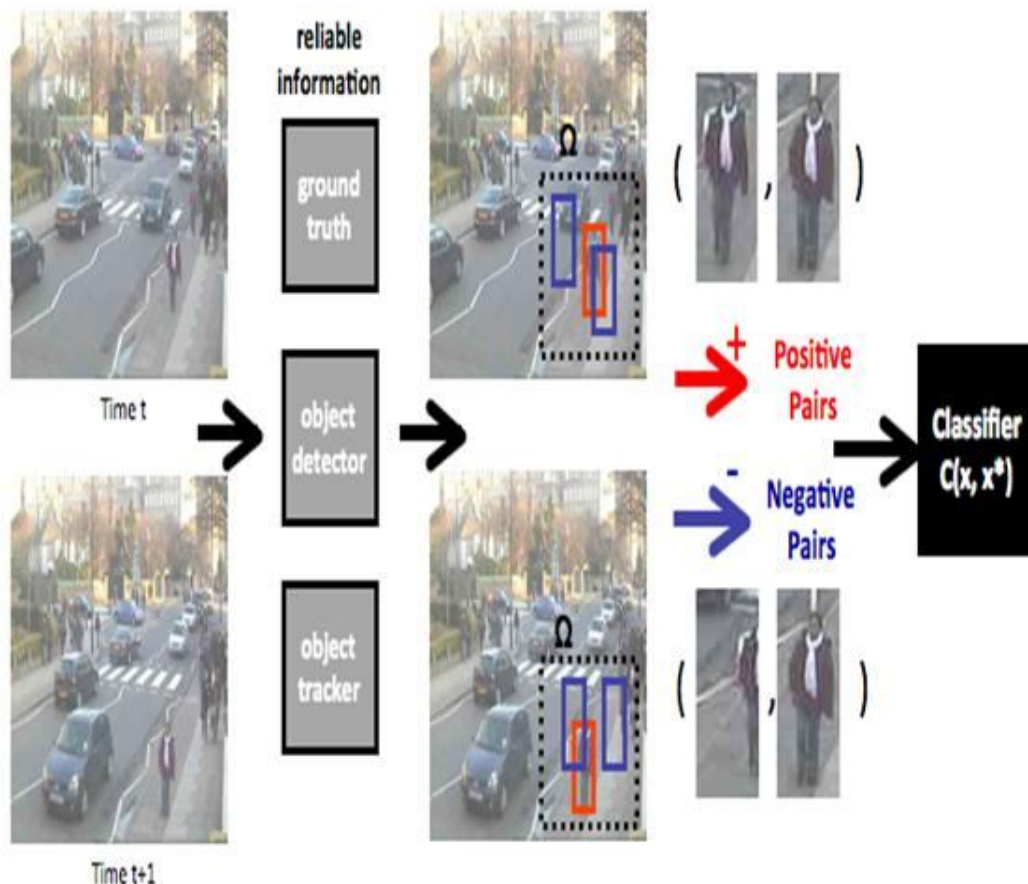
(*confidence value*), όταν τα δυο δείγματα x_1 και x_2 είναι παρόμοια και χαμηλό βαθμό απόκρισης αν τα δυο δείγματα διαφέρουν σημαντικά μεταξύ τους.

Μάθηση Ροής Αντικειμένου: Η συνολική προσέγγιση του μηχανισμού μάθησης απεικονίζεται στο Σχήμα 14. Για την εκπαίδευση ενός ταξινομητή (*classifier*) πάνω στην μετατόπιση αντικειμένου πρέπει να δημιουργηθεί μια ομάδα με τα κατάλληλα δείγματα. Συγκεκριμένα, τα δείγματα εκπαίδευσης αποτελούνται από ζευγάρια εικόνων και περιέχουν την κατάλληλη χρονική πληροφορία για το θετικό X^+ και το αρνητικό σύνολο X^- , αντίστοιχα. Πιο συγκεκριμένα,

- **Θετικό σύνολο X^+ :** Ένα θετικό δείγμα περιέχει πληροφορίες σχετικά με τον τρόπο που ένα κινούμενο αντικείμενο μεταλλάσσεται μέσα στο χρόνο. Ως εκ τούτου, το δείγμα αυτό δημιουργείται από τη ένωση δύο υπό-εικόνων που προέρχονται από δύο διαφορετικές εικόνες χρονικά και περιέχουν το αντικείμενο υπό μελέτη,

$$X^+ = \{(x_t^*, x_{t+1}^*) \mid x_t^*, x_{t+1}^* \in \Omega^{(i)} \text{ και ανήκει σε αντικείμενο}\} \quad (5.1)$$

Η σήμανση του αντικειμένου που αντιπροσωπεύεται από τις περιοχές x_t^* και x_{t+1}^* μπορεί να γίνει με κάποια αξιόπιστη πληροφορία, όπως με την βοήθεια ανθρώπινης σήμανσης (*ground truth*) ή χρησιμοποιώντας την έξοδο από ένα ανιχνευτή αντικειμένων με υψηλή ακρίβεια.



Σχήμα 14. Η εκπαίδευση του ταξινομητή πάνω στην μετατόπιση αντικειμένου επιτυγχάνεται χρησιμοποιώντας θετικά και αρνητικά δείγματα, τα οποία εξάγονται τοπικά και περιέχουν χρονικές πληροφορίες.

– **Αρνητικό σύνολο X^-** : Το σύνολο που περιέχει τα αρνητικά δείγματα

χωρίζεται σε δύο υποσύνολα $X^- = X_{obj}^- \cup X_{back}^-$. Το πρώτο υποσύνολο

αποτελείται από δείγματα τα οποία περιέχουν το αντικείμενο στην παρούσα εικόνα μαζί με ένα κομμάτι του αντικειμένου από προηγούμενη εικόνα.

Συγκεκριμένα,

$$X_{obj}^- = \left\{ (x_t^*, x_{t+1}^{(i)}) \mid x_t^*, x_{t+1}^{(i)} \in \Omega^{(i)} \text{ και } x_{t+1}^* \text{ ανήκει σε αντικείμενο} \right\} \quad (5.2)$$

Αυτά τα δείγματα εκπαίδευσης βοηθούν τον ταξινομητή να καταστείλει τα τοπικά μέγιστα (*min-max suppression*) γύρω από την περιοχή που βρίσκεται το αντικείμενο. Από την άλλη πλευρά, το δεύτερο υποσύνολο των αρνητικών δειγμάτων περιλαμβάνει περιοχές από το περιβάλλοντα χώρο. Τέτοια δείγματα είναι ιδιαίτερα χρήσιμα σε δύσκολες εφαρμογές, διότι μπορούν να βοηθήσουν τον ταξινομητή (*classifier*) έτσι ώστε να εκτιμήσει με χαμηλές τιμές (*classification responses*) τις κενές περιοχές. Συγκεκριμένα,

$$X_{\text{train}}^- = \left\{ (x_z^{(i)}, x_z^{(j)}) \mid x_z^{(i)}, x_z^{(j)} \in \Omega^{(i)} \right\} \quad (5.3)$$

Παραδείγματα θετικών και αρνητικών δειγμάτων παρουσιάζονται στο Σχήμα 15.



Σχήμα 15. Χαρακτηριστικά δείγματα για την εκπαίδευση ενός ταξινομητή στην ροή αντικειμένου.

Ταξινομητής (Classifier): Στην διατριβή αυτή χρησιμοποιήθηκε η προσέγγιση «*Boosting for feature selection*» με τα χαρακτηριστικά *Haar* [3]. Ένας ταξινομητής μπορεί να εκπαιδευτεί σε μη πραγματικό [3], ή σε πραγματικό [4] χρόνο. Για να χρησιμοποιηθούν τα ζεύγη εικόνων ως είσοδο στο ταξινομητή, ακολουθήθηκαν οι εμπειρικές μελέτες που προτείνονται από *Hertz* [5]. Μια προσέγγιση για την εκμάθηση της αλλαγής της εμφάνισης ενός αντικειμένου μπορεί να γίνει με δείγματα που αποτελούνται από την συνένωση των δύο εικόνων που περιέχουν το αντικείμενο.

Μια άλλη προσέγγιση είναι η δημιουργία δειγμάτων που αποτελούνται από την απόλυτη διαφορά των διανυσμάτων που περιέχουν τις δύο εικόνες. Εμπειρικές μελέτες που πραγματοποιήθηκαν έδειξαν ότι ο ταξινομητής αποδίδει καλύτερα με την πρώτη προσέγγιση.

5.2 Εκτίμηση ροής

Η ροή αντικείμενου (*object flow*) είναι ένα διανυσματικό μέγεθος και για μια τοπική υπό-εικόνα \mathbf{x} μπορεί να υπολογιστεί το πλάτος μετατόπισης $D_{obj}(\mathbf{x})$ και η γωνία $\varphi_{obj}(\mathbf{x})$. Συγκεκριμένα, έστω $C(\mathbf{x}, \mathbf{x}')$ η απόκριση του ταξινομητή για ένα ζευγάρι υπό-εικόνων, όπου \mathbf{x} είναι μια υπό-εικόνα στην τρέχουσα εικόνα και \mathbf{x}' είναι μια υπό-εικόνα που ανήκει στην περιοχή Ω των τοπικών υπό-εικόνων στην προηγούμενη εικόνα. Ορίζουμε την μετατόπιση $\Delta x_{obj}(\mathbf{x})$ και $\Delta y_{obj}(\mathbf{x})$ ενός αντικειμένου στις x και y κατευθύνσεις, το σταθμισμένο άθροισμα (*weighted sum*) των αποστάσεων εντός της τοπικής περιοχής Ω . Πιο συγκεκριμένα,

$$\begin{pmatrix} \Delta x_{obj}(\mathbf{x}) \\ \Delta y_{obj}(\mathbf{x}) \end{pmatrix} = \frac{1}{\sum_{\mathbf{x}' \in \Omega} C(\mathbf{x}, \mathbf{x}')} \sum_{\mathbf{x}' \in \Omega} C(\mathbf{x}, \mathbf{x}') \begin{pmatrix} dx \\ dy \end{pmatrix} \quad (5.4)$$

Όπου, dx και dy είναι οι x και y αποστάσεις τις υπό-εικόνας \mathbf{x} από το \mathbf{x}' . Βάσει αυτού, το πλάτος και η γωνία μπορούν να υπολογιστούν ως εξής,

$$\begin{aligned} D_{obj}(\mathbf{x}) &= \sqrt{\Delta x_{obj}(\mathbf{x})^2 + \Delta y_{obj}(\mathbf{x})^2} \\ \varphi_{obj}(\mathbf{x}) &= \tan^{-1} \left(\frac{\Delta y_{obj}(\mathbf{x})}{\Delta x_{obj}(\mathbf{x})} \right) \end{aligned} \quad (5.5)$$

Προκειμένου να μειωθούν οι ακραίες τιμές (*outliers*), οι τοπικές μετατοπίσεις των αντικειμένων εντός της περιοχής Ω θα πρέπει αυτές να ξεπερνούν κάποια συγκεκριμένη τιμή της απόκρισης του ταξινομητή. Συγκεκριμένα,

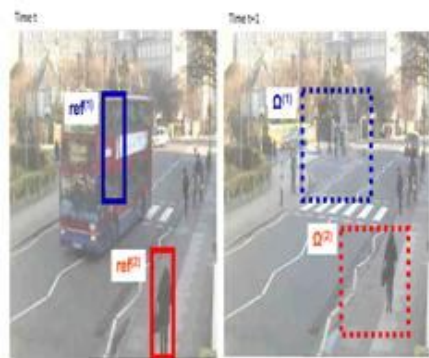
$$\bar{c}_{obj}(\mathbf{x}) = \frac{1}{|\Omega|} \sum_{\mathbf{x} \in \Omega} \tilde{c}(\mathbf{x}, \mathbf{x}')^2 \quad (5.6)$$

$$\text{όπου } \tilde{c}(\mathbf{x}, \mathbf{x}') = \max(0, c(\mathbf{x}, \mathbf{x}'))$$

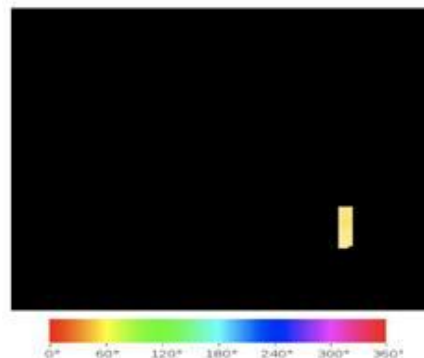
Συνοψίζοντας, η ροή αντικειμένου σημειώνεται μόνον εάν η μέση απόκριση του ταξινομητή είναι μεγαλύτερη από ένα προκαθορισμένο κατώφλι, το οποίο ελέγχει την ευαισθησία $\bar{c}_{obj}(\mathbf{x}) > \theta$.

Ενδεικτικό παράδειγμα. Το Σχήμα 16 απεικονίζει τη ροή αντικείμενου για τις περιοχές $\Omega^{(1)}$ και $\Omega^{(2)}$. Ο εκπαιδευμένος ταξινομητής αξιολογείται σε ζεύγη υπό-εικόνων, χρησιμοποιώντας ως σημείο αναφοράς μια υπό-εικόνα (ref^1 , ref^2) τη χρονική στιγμή t σε συνδυασμό με υπό-εικόνες από τις αντίστοιχες τοπικές περιφέρειες, $(\Omega^{(1)}, \Omega^{(2)})$, αντίστοιχα τη χρονική στιγμή $t+1$. Συγκεκριμένα χρησιμοποιείται ένα πλέγμα με επικαλυπτόμενες υπό-εικόνες του ίδιου μεγέθους, οι οποίες έχουν ως κέντρο την υπό-εικόνα (ref^1 , ref^2) που έχει οριστεί ως σημείο αναφοράς. Στο τρισδιάστατο γράφημα για την περιοχή $\Omega^{(2)}$, υψηλές τιμές απόκρισης παρατηρούνται σε σημεία όπου το αντικείμενο υπό μελέτη είναι πιθανό να εμφανιστεί την χρονική στιγμή $t+1$. Από την άλλη πλευρά, οι τιμές απόκρισης του ταξινομητή είναι πολύ χαμηλές για την περιοχή $\Omega^{(1)}$, δεδομένου ότι δεν υπάρχουν αντικείμενα υπό μελέτη στο εσωτερικό της. Για την απεικόνιση της γωνίας $\varphi_{obj}(ref^1)$, $\varphi_{obj}(ref^2)$ και του πλάτους μετατόπισης $D_{obj}(ref^1)$, $D_{obj}(ref^2)$ αντίστοιχα (βλ.

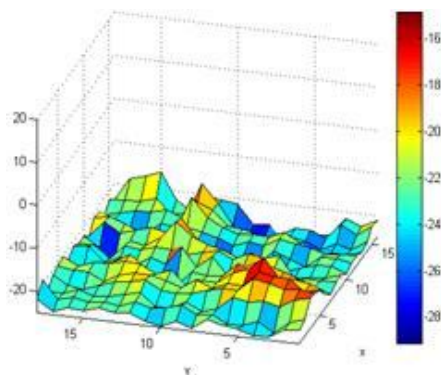
Εξίσωση (5.5)), χρησιμοποιήθηκε το κανάλι απόχρωσης (*hue*) και κορεσμού (*saturation*) από το *HSV* χρωματικό πεδίο.



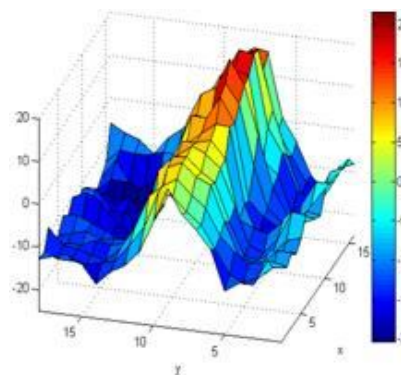
(α) Εικόνες



(β) Ροή Αντικείμενου



(γ) Απόκριση για την περιοχή $\Omega^{(1)}$



(δ) Απόκριση για την περιοχή $\Omega^{(2)}$

Σχήμα 16. Αποκρίσεις ταξινομητή για τις περιοχές $\Omega^{(1)}$ και $\Omega^{(2)}$, (α). Χαμηλές τιμές απόκρισης λαμβάνονται στην περίπτωση όπου δεν βρίσκεται το αντικείμενο στην περιοχή (γ). Αντίθετα, οι κορυφές δείχνουν την μετατόπιση του αντικείμενου (δ). Το τελικό πεδίο ροής αντικειμένου (β) βασίζεται σε αυτές τις τοπικές αποκρίσεις.

5.3 Παραδείγματα ροής αντικειμένου

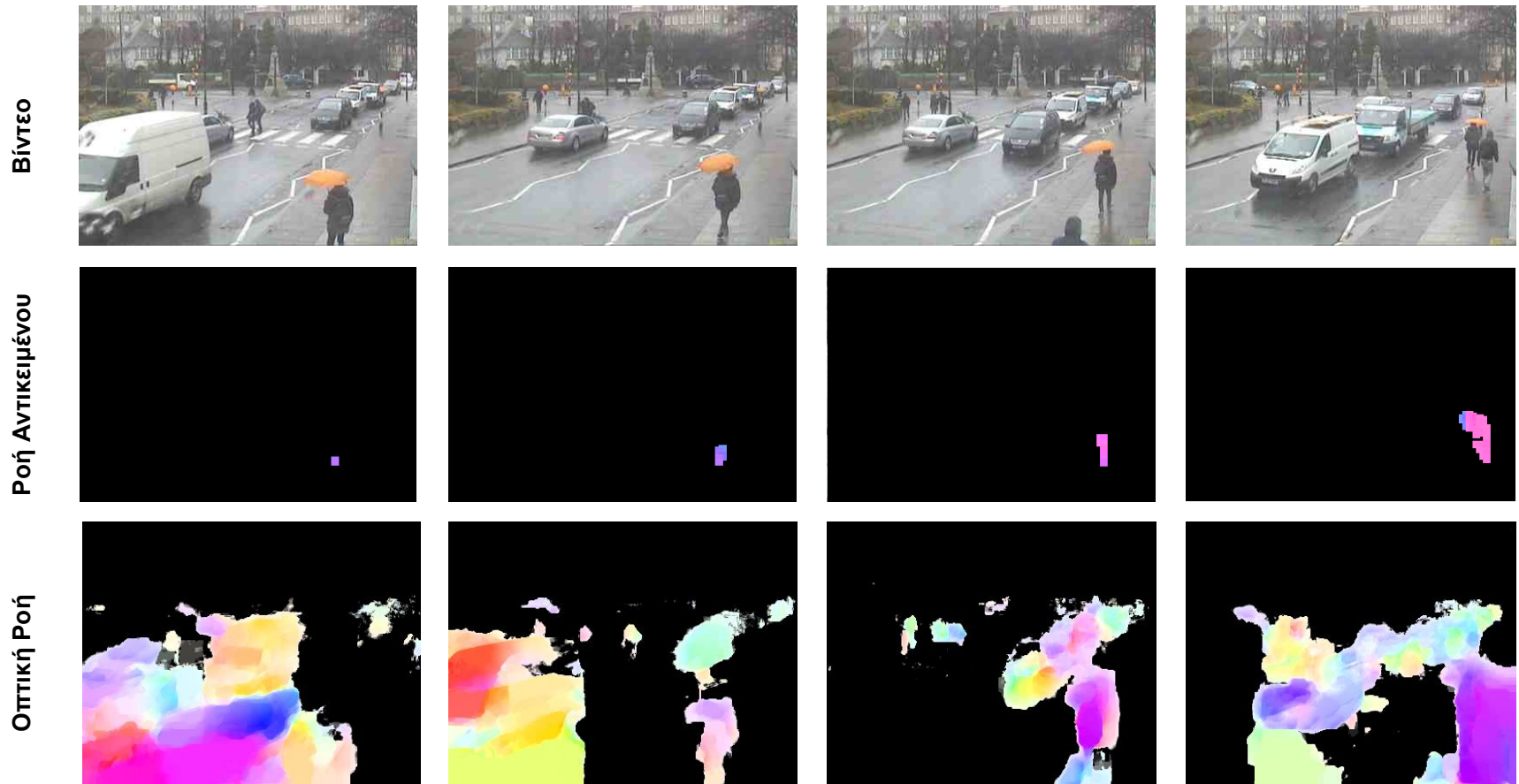
Στην ενότητα αυτή παρουσιάζονται μερικά παραδείγματα ροής αντικειμένου. Η αποτελεσματικότητα της προσέγγισής παρουσιάζεται, χρησιμοποιώντας σενάρια τα οποία περιλαμβάνουν χαμηλό ρυθμό καρέ (*low frame rate*) και θόλωμα εικόνας λόγω της κίνησης της κάμερας. Επίσης, έγιναν συγκρίσεις της αποτελεσματικότητας της ροής αντικειμένου με άλλες μεθόδους, όπως με ανιχνευτές αντικείμενου (*object detectors*), ιχνηλάτη (*tracker*) και με την τεχνική της οπτική ροής (*optical flow*). Θα πρέπει να σημειωθεί ότι η προτεινόμενη αναπαράσταση κίνησης μπορεί να χρησιμοποιηθεί σε πληθώρα εφαρμογών που περιλαμβάνουν είτε στατικές είτε κινούμενες κάμερες. Σε όλα τα πειράματα χρησιμοποιήθηκε ένα πυκνό πλέγμα, το οποίο αποτελείται από 81×81 επικαλύψεις ίσου μεγέθους κελίων (υπό - εικόνες) θέτοντας ως κατώφλι $\theta = 0,35$ (βλ. Ενότητα 5.2). Όλα τα πειράματα εκτελέστηκαν σε $2,67 \text{ GHz PC}$ με 4 GB μνήμης *RAM*.

5.3.1 Ροή αντικειμένου για πεζούς

Αρχικά, μαγνητοσκοπήθηκε ένα βίντεο που αποτελείται από $49,000$ εικόνες, χρησιμοποιώντας μια κάμερα η οποία βρίσκεται στο *Abbey Road* του Λονδίνου. Σε αυτό το βίντεο κάθε εικόνα έχει ανάλυση 384×284 και χαρακτηρίζεται από χαμηλό ρυθμό καρέ ($\sim 6 \text{ fps}$). Επίσης περιέχει ένα σύνθετο φόντο το οποίο αποτελείται από διάφορα κινούμενα αντικείμενα (π.χ. αυτοκίνητα). Ως εκ τούτου, χρησιμοποιήθηκε μια περιοχή που αποτελείται από 12×12 κελία, δεδομένου ότι η κίνηση του αντικειμένου (πεζοί) είναι αρκετά απότομη λόγω του χαμηλού ρυθμού καρέ. Οι

πρώτες **40,000** εικόνες χρησιμοποιήθηκαν με σκοπό την συλλογή δειγμάτων εκπαίδευσης (βλ. Κεφάλαιο 5.1) Οι υπόλοιπες εικόνες χρησιμοποιήθηκαν με σκοπό την αξιολόγηση της προτεινόμενης μεθόδου. Πιο συγκεκριμένα, τα αποτελέσματα που περιγράφονται σε αυτό το κομμάτι της εργασίας μπορούν να παραχθούν από έναν ταξινομητή (*classifier*) που έχει εκπαιδευτεί σε μη πραγματικό χρόνο, χρησιμοποιώντας μια δεξαμενή $|X^+| \approx 2,000$ θετικών και $|X_{obj}^-| \approx 15,000$ αρνητικών δειγμάτων και πολυάριθμων αρνητικών δειγμάτων X_{back}^- από τον περιβάλλοντα χώρο.

Πραγματοποιήθηκαν ποιοτικές συγκρίσεις με μεθόδους οπτικής ροής, ανιχνευτές ανθρώπων (*human detectors*) και ιχνηλάτες (*trackers*). Τα πειράματα βασίστηκαν στην προσέγγιση που περιγράφεται στην δημοσίευση [6] για τον υπολογισμό της οπτικής ροής, προκειμένου να συγκριθεί με την προτεινόμενη τεχνική. Για την ανίχνευση και ιχνηλάτιση πεζών υιοθετήθηκαν οι προσεγγίσεις που περιγράφονται στις δημοσιεύσεις [7] και [6], αντίστοιχα. Όλες οι ανταγωνιστικές τεχνικές χρησιμοποιήθηκαν χωρίς καμία τροποποίηση των αρχικών παραμέτρων τους (δηλαδή, παραμείναν οι ίδιες με την αρχική εφαρμογή τους).





Σχήμα 17. Σε αυτό το πείραμα παρουσιάζονται τα οφέλη της ροής αντικειμένου. Οι προσεγγίσεις οπτικής ροής αποπροσανατολίζονται, όταν παρόμοια αντικείμενα κινούνται προς την ίδια ή διαφορετική κατεύθυνση με το αντικείμενο-του-ενδιαφέροντος (τρίτη γραμμή). Επιπλέον, οι προσεγγίσεις ανθρώπινης ανίχνευσης δεν έχουν σταθερό ρυθμό ανίχνευσης (τέταρτη γραμμή). Επίσης, ο εντοπισμός αντικειμένων δυσκολεύει σε σύνθετα περιβάλλοντα (πέμπτη γραμμή). Αντιθέτως η ροή αντικειμένου (δεύτερη σειρά) μπορεί να προσομοιώσει το πεδίο κίνησης ενός αντικειμένου με επιτυχία, ενώ παράλληλα είναι σε θέση να επικεντρώνετε μόνο στο αντικείμενο υπό μελέτη.

Η σελίδα αυτή είναι σκόπιμα λευκή.

Όπως μπορεί να παρατηρηθεί στο Σχήμα 17, η προτεινόμενη προσέγγισή έχει καλή απόδοση στον εντοπισμό ανθρώπων. Επιπλέον, η εκτίμηση που γίνεται για την κατεύθυνση των πεζών (βλ. δεύτερη σειρά) είναι η ίδια με εκείνη που παρουσιάζεται από την μέθοδο της οπτικής ροής, η οποία επικεντρώνεται σε όλα τα κινούμενα αντικείμενα μέσα στην σκηνή (π.χ. αυτοκίνητα, βλ. τρίτη σειρά). Από την άλλη πλευρά, ο συνδυασμός μεθόδων οπτικής ροής με προσεγγίσεις ανίχνευσης αντικειμένων (τέταρτη γραμμή) μπορεί να οδηγήσει σε σφάλματα, αφού τέτοιες προσεγγίσεις δεν έχουν σταθερό ρυθμό ανίχνευσης και συνεπώς έχουν περιορισμένη αποτελεσματικότητα σε δύσκολες συνθήκες. Ομοίως, οι μέθοδοι ιχνηλάτισης (*tracking*, βλ. πέμπτη γραμμή) αποπροσανατολίζονται σε σύνθετα περιβάλλοντα, δεδομένου ότι αντικείμενα του ίδιου χρώματος ή της ίδιας δομής μπορεί να εμφανιστούν μέσα στη σκηνή.

5.3.2 Ροή αντικειμένου για διαφορετικά αντικείμενα

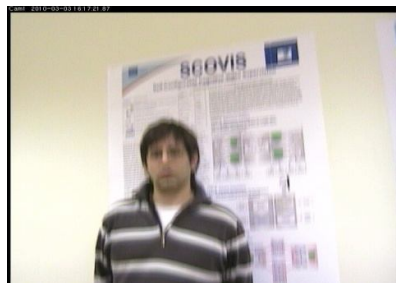
Στην ενότητα αυτή παρουσιάζεται η απόδοση της μεθόδου ροής αντικείμενου σε διαφορετικές κλάσεις αντικειμένων, όπως για διαφορετικά πρόσωπα, κούπα του καφέ και για ανθρώπους που κινούνται μέσα σε βιομηχανικό περιβάλλον. Ο αλγόριθμος αξιολογήθηκε σε δύσκολα σενάρια τα οποία περιέχουν συνθήκες όπως απότομης κίνησης και περιβάλλον κινούμενης κάμερας. Πιο αναλυτικά, χρησιμοποιήθηκαν τέσσερις διαφορετικές ακολουθίες βίντεο που αποτελούνται από **1200** καρέ εικόνων, όπου τα **1000** πρώτα καρέ χρησιμοποιούνται για την εκπαίδευση και τα υπόλοιπα **200** καρέ για την αξιολόγηση του ταξινομητή αντίστοιχα. Αρχικά, μαγνητοσκοπήθηκαν δύο διαφορετικές ακολουθίες από μια κινούμενη κάμερα εσωτερικού χώρου και περιλαμβάνουν ένα κινούμενο πρόσωπο και μια κούπα του καφέ αντίστοιχα. Μια άλλη ακολουθία μαγνητοσκοπήθηκε από μια στατική εσωτερική κάμερα και περιλαμβάνει δύο κινούμενα πρόσωπα. Συγκεκριμένα, ανάλυση **704 × 576** χρησιμοποιήθηκε στις πρώτες δύο ακολουθίες. Τέλος, όλες οι ακολουθίες ελήφθησαν στα **25 fps** χρησιμοποιώντας μια *AXIS 213 PTZ* κάμερα.

Η προτεινόμενη μέθοδος έχει αξιολογηθεί για κάθε \mathbf{x} εσωτερική υπό-εικόνα χρησιμοποιώντας μια περιοχή που αποτελείται από **6 × 6** κελία. Για όλες τις ακολουθίες ο ταξινομητής εκπαιδεύτηκε με μια ομάδα $|\mathbf{X}^+| \approx 1,000$ θετικών, $|\mathbf{X}_{obj}^-| \approx 4,000$ αρνητικών δειγμάτων και πολυάριθμων αρνητικών δειγμάτων \mathbf{X}_{back}^- από τον περιβάλλοντα χώρο. Ειδικότερα, για τη δημιουργία δειγμάτων προσώπου χρησιμοποιήθηκε μια γνωστή προσέγγιση ανίχνευσης προσώπων [3].

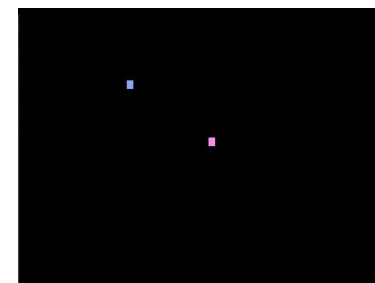
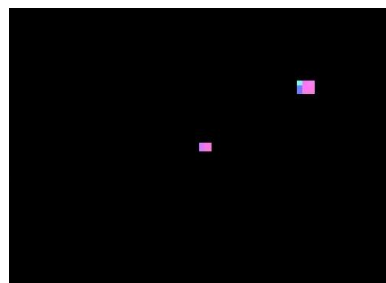
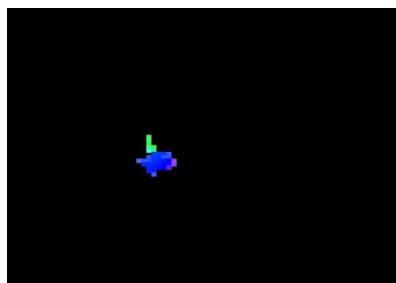
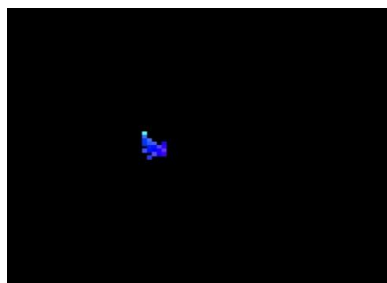
Τα ποιοτικά αποτελέσματα απεικονίζονται στο Σχήμα 18. Όπως μπορεί να παρατηρηθεί, η μέθοδος της ροής αντικειμένου επικεντρώνεται στα αντικείμενα

ενδιαφέροντος, ακόμη και σε περιπτώσεις όπου η κίνηση της κάμερας είναι απότομη (δεύτερη και τέταρτη γραμμή). Επιπλέον, η προτεινόμενη μέθοδος μπορεί να υπολογίσει την κίνηση περισσότερων από ένα αντικείμενα τα οποία παρευρίσκονται στον ίδιο χώρο (βλ. Σχήμα 18, πρώτη και δεύτερη σειρά, στην τρίτη και τέταρτη στήλη).

Βίντεο



Ροή Αντικειμένων



Βίντεο



Ροή Αντικειμένου



Σχήμα 18. Σε αυτό το πείραμα εκπαιδεύσαμε τον ταξινομητή σε διαφορετικές κατηγορίες αντικειμένων. Η πρώτη και η τρίτη σειρά απεικονίζουν καρέ από διαφορετικές σειρές δοκιμών (κίνηση της κάμερας, θόλωση λόγω κίνησης και πολλαπλά αντικείμενα).

Η σελίδα αυτή είναι σκόπιμα λευκή.

5.4 Ποσοτική αξιολόγηση

Σε αυτό το κεφάλαιο παρουσιάζονται δύο διαφορετικά πειράματα με στόχο την ποσοτική μελέτη της αξιοπιστίας της μεθόδου ροής αντικειμένου. Αρχικά χρησιμοποιείται μια εικονοσειρά με μία κινούμενη κούπα, το οποίο είναι ένα σενάριο με κινούμενη κάμερα και αποτελείται από **200** καρέ εικόνων. Σε αυτή την εικονοσειρά, η σήμανση των περιοχών που βρίσκεται το πραγματικό αντικείμενο έγινε μέσω αντικειμενικής επιλογής (*ground truth*) με σκοπό να υπολογιστεί η ακρίβεια του πλάτους μετατόπισης. Συγκεκριμένα για κάθε εικόνα υπολογίστηκε το απόλυτο σφάλμα μεταξύ των αντικειμενικών τιμών και των τιμών που παρέχονται από την προτεινόμενη προσέγγιση. Δεδομένου ότι υπάρχει μόνο ένα αντικείμενο ενδιαφέροντος για την σκηνή, χρησιμοποιήθηκε το πλάτος μετατόπισης $D_{obj}(\mathbf{x})$ για το οποίο ο ταξινομητής έχει τη μέγιστη ανταπόκριση $\bar{C}_{obj}(\mathbf{x})$, σύμφωνα με την Εξίσωση (5.6).

Επίσης, για λόγους σύγκρισης και αναφοράς, χρησιμοποιείται μια απλή προσέγγιση που συνδυάζει την ανίχνευση αντικειμένων και την οπτική ροή. Στην περίπτωση αυτή, η μετατόπιση υπολογίζεται από την εύρεση της μέσης οπτικής ροής εντός της περιοχής της ανίχνευσης (π.χ. στην περίπτωση που υπάρχει ανίχνευση αντικειμένου). Ως εκ τούτου, ο ταξινομητής [3] εκπαιδεύτηκε χρησιμοποιώντας $|\mathbf{X}^+| \approx 1,000$ θετικά δείγματα για το αντικείμενο και ένα αρνητικό σύνολο που περιέχει δείγματα χωρίς το αντικείμενο από τον περιβάλλοντα χώρο. Τα αποτελέσματα ανίχνευσης συνδυάζονται μαζί με την εφαρμογή μη-μέγιστη καταστολής (*min-max suppression*). Τέλος, υιοθετήθηκε η μέθοδος του *Lucas-Kanade* [8] με σκοπό την εκτίμηση της οπτικής ροής.

Ο Πίνακας 3 παρουσιάζει το μέσο απόλυτο σφάλμα και την διακύμανση για τη μετατόπιση αντικειμένου για το σύνολο της ακολουθίας.

Πίνακας 3. Η μέση τιμή και η διακύμανση σφάλματος μετατόπισης για δύο διαφορετικές μεθόδους εκτίμησης για σενάριο με κινούμενη κάμερα.

	Οπτική Ροή + Ανίχνευση αντικειμένων	Ροή αντικειμένου
Μέση τιμή σφάλματος μετατόπισης	12,72	3,1
Διακύμανση σφάλματος μετατόπισης	15,76	10,25

Με παρόμοιο τρόπο εξετάστηκε η απόδοση της ροής αντικειμένου σε ένα σύνολο δεδομένων από μια κάμερα που βρίσκεται σε ένα βιομηχανικό περιβάλλον (βλ. Σχήμα 19). Αυτό είναι ένα δύσκολο σενάριο που περιέχει ένα σύνθετο περιβάλλον με διάφορα άσχετα κινούμενα αντικείμενα (π.χ., συσκευές συγκόλλησης). Η σήμανση των περιοχών που βρίσκεται το πραγματικό αντικείμενο γίνεται μέσω αντικειμενικής επιλογής (*ground truth*). Για λόγους σύγκρισης, χρησιμοποιήθηκε η μέθοδος οπτικής ροής που περιγράφεται στο [8]. Ειδικότερα, η αξιολόγηση της προτεινόμενης προσέγγισης γίνεται με την εκπαίδευση ενός ταξινομητή με $|X^+| \approx 1,130$ θετικά, $|X_{obj}^-| \approx 4,000$ αρνητικά δείγματα και πολυάριθμα αρνητικά δείγματα X_{back}^- από τον περιβάλλοντα χώρο.

Ο Πίνακας 4 παρουσιάζει το μέσο σφάλμα σε *pixels* και την διακύμανση για τη μετατόπιση αντικειμένου για το σύνολο της ακολουθίας. Μπορεί να παρατηρηθεί ότι η ροή του αντικειμένου έχει καλύτερη απόδοση σε σύγκριση με την άλλη προσέγγιση. Επομένως, η μέθοδος ροής αντικειμένου μπορεί να χρησιμοποιηθεί

αποτελεσματικά σε μια εφαρμογή ιχνιλάτισης (*tracking*) (βλ. Κεφάλαιο 6.1.2), δεδομένου ότι υπάρχει μια ακριβή εκτίμηση της μετατόπισης του αντικειμένου.

Πίνακας 4. Η μέση τιμή και η διακύμανση σφάλματος μετατόπισης για δύο διαφορετικές μεθόδους εκτίμησης. Ο υπολογισμός έγινε με ένα σύνολο δεδομένων από μια κάμερα που βρίσκεται σε ένα βιομηχανικό περιβάλλον.

	Οπτική Ροή	Ροή αντικειμένου
<i>Μέση τιμή σφάλματος μετατόπισης</i>	3,07	2,9
<i>Διακύμανση σφάλματος μετατόπισης</i>	13,42	8,15

Η σελίδα αυτή είναι σκόπιμα λευκή.

6

Προτεινόμενο

Σύστημα

Παρακολούθησης

Η παρακολούθηση αντικειμένων μέσω καμερών είναι ένα απαραίτητο κομμάτι των συστημάτων όρασης υπολογιστών, έτσι ώστε να εξαχθούν συμπεράσματα για την κατάσταση σε μία σκηνή. Ωστόσο, ο τομέας της παρακολούθησης αντικειμένων εξακολουθεί να παραμένει μία από τις μεγαλύτερες προκλήσεις στο ερευνητικό πεδίο της όρασης υπολογιστών, παρόλο που έχει μελετηθεί εκτενώς τα προηγούμενα χρόνια. Σε αυτήν την ενότητα παρουσιάζεται μία αξιόπιστη προσέγγιση παρακολούθησης αντικειμένων που βασίζεται στην τεχνική ροής αντικειμένου (βλ. Κεφάλαιο 5). Αυτή η τεχνική μπορεί να εκτιμήσει τόσο την μετατόπιση όσο και την κατεύθυνση ενός αντικειμένου συγκεκριμένου ενδιαφέροντος. Επιπλέον, ένα νέο μοντέλο περιγραφής αντικειμένου ενσωματώνεται στο σύστημα παρακολούθησης. Αυτό το μοντέλο βασίζεται σε μία παραγωγική μέθοδο μάθησης (βλ. Κεφάλαιο 4.1.2) και βοηθάει στον ευκολότερο εντοπισμό των αντικειμένων που έχουν πολλαπλές αλλαγές στην εμφάνιση τους. Η αποτελεσματικότητα του προτεινόμενου συστήματος

παρακολούθησης αποδεικνύεται με πειραματικές ακολουθίες που περιέχουν ένα ιδιαίτερα σύνθετο βιομηχανικό περιβάλλον.

Θα πρέπει να σημειωθεί ότι η εξαγωγή οπτικών πληροφοριών από εικονοσειρές είναι μία δύσκολη διαδικασία, δεδομένου ότι διάφορα μαθηματικά μοντέλα δεν μπορούν με ευκολία να ερμηνεύσουν πολύπλοκα δεδομένα. Για αυτό το λόγο οι αλγόριθμοι εντοπισμού συχνά κάνουν σφάλματα λόγω της ύπαρξης θορύβου. Έτσι ο αυτόματος εντοπισμός διαφόρων αντικειμένων εξακολουθεί να είναι ένα δύσκολο πρόβλημα προς επίλυση. Στην παρούσα εργασία χρησιμοποιήθηκε ένα βιομηχανικό περιβάλλον, όπου τόσο η εμφάνιση του αντικειμένου όσο και η κίνηση του περιβάλλοντος αλλάζει σημαντικά. Το Σχήμα 19 απεικονίζει μερικές εικόνες από μια πραγματική αυτοκινητοβιομηχανία [69], όπου η πολυπλοκότητα της σκηνής είναι εμφανής με μία απλή παρατήρηση.



Σχήμα 19. Παραδείγματα εικόνων από βίντεο που μαγνητοσκοπήθηκε σε μία βιομηχανική εγκατάσταση [69]. Οι απλές προσεγγίσεις παρακολούθησης αντικειμένων δεν μπορούν να εφαρμοστούν με ευκολία, λόγω του δύσκολου περιβάλλοντος.

Οι παραδοσιακοί αλγόριθμοι παρακολούθησης αντικειμένων δεν μπορούν να εφαρμοστούν άμεσα σε αυτήν την ακολουθία διότι, (α) η μετατόπιση των ατόμων έχει

απρόβλεπτη τροχία και τα άτομα έχουν παρόμοιο χρώμα με το φόντο (δηλαδή, υπάρχει χαμηλή διακριτική πληροφορία μεταξύ του προσκηνίου και του παρασκηνίου), (β) ο χώρος που κινούνται τα άτομα αλλάζει απότομα (π.χ. υπάρχουν σπινθήρες από μηχανές συγκόλλησης, κινούμενα ρομπότ, κλπ.), (γ) υπάρχουν ασυνέχειες στο μοντέλο που περιγράφει το αντικείμενο λόγω ολικών ή μερικών επικαλύψεων. Ως εκ τούτου, διάφορες τεχνικές που υπάρχουν στην τρέχουσα βιβλιογραφία αποτυγχάνουν στο να παρακολουθήσουν με επιτυχία το αντικείμενο ενδιαφέροντος.

Στην παρούσα εργασία παρουσιάζεται μια προσέγγιση παρακολούθησης αντικειμένου, η οποία είναι ικανή να αντιμετωπίσει τις δυσκολίες που προκύπτουν σε ακολουθίες εικόνων από σύνθετα βιομηχανικά περιβάλλοντα (βλ. Σχήμα 19). Πιο αναλυτικά, η τεχνική της ροής αντικειμένου (βλ. Κεφάλαιο 5) ενσωματώνεται σε ένα φίλτρο σωματιδίων. Η ροή αντικειμένου είναι μία τεχνική μάθησης, η οποία παρέχει πληροφορίες για την κατεύθυνση και την μετατόπιση ενός αντικειμένου, ενώ παράλληλα μπορεί να αγνοεί άλλες άσχετες κινήσεις μέσα σε μία σκηνή. Η προτεινόμενη μέθοδος παρακολούθησης χρησιμοποιεί την ροή αντικειμένου τόσο ως μοντέλο κίνησης όσο ως μοντέλο περιγραφής αντικειμένου. Ωστόσο, εισάγεται ένα επιπλέον μοντέλο περιγραφής αντικειμένου, με σκοπό να αντιμετωπίσει προβλήματα που σχετίζονται με αναξιόπιστες ανιχνεύσεις αντικειμένου λόγω της πολυπλοκότητας του βιομηχανικού περιβάλλοντος. Αυτό το μοντέλο περιγραφής χρησιμοποιεί μια παραγωγική *prior* μέθοδο, η οποία βασίζεται στο *GPLVM* (βλ. Κεφάλαιο 6.1.3.1). Σε αυτή την προσέγγιση το αντικείμενο υπό μελέτη (δηλαδή ο άνθρωπος που μετακινείται) υποβάλλεται σε διάφορες αλλαγές στην εμφάνιση του. Έτσι η *GPLVM* μπορεί να παρέχει μία αποτελεσματική περιγραφή της κατάστασης της εμφάνισης του.

Επίσης παρουσιάζεται ένας μηχανισμός ελέγχου δειγμάτων εκπαίδευσης πραγματικού χρόνου (βλ. Κεφάλαιο 6.1.5). Αυτός ο μηχανισμός χρησιμοποιεί την κύρτωση ενός διαγράμματος απόκρισης για να ελέγξει αν το αποτέλεσμα του συστήματος παρακολούθησης είναι κατάλληλο για την εκπαίδευση του ταξινομητή σε πραγματικό χρόνο. Τα αποτελέσματα αυτής της μεθόδου εξετάστηκαν παράλληλα με το προτεινόμενο σύστημα παρακολούθησης αλλά για διαφορετικές εικονοσειρές. Αυτό συμβαίνει γιατί οι εικονοσειρές που μαγνητοσκοπήθηκαν σε βιομηχανικό περιβάλλον είναι από πανοραμική λήψη οπότε δεν υπάρχουν επικαλύψεις (*occlusions*) από άλλα αντικείμενα.

Θα πρέπει να σημειωθεί ότι τα συστήματα παρακολούθησης είναι από την φύση τους μία ήμι-αυτόματη (*semi-automatic*) διαδικασία, δεδομένου ότι ο χρήστης καλείται συχνά να καθορίσει την αρχική περιοχή, η οποία περιέχει το αντικείμενο που επιθυμεί να παρακολουθηθεί. Σε αυτήν την εργασία το σύστημα παρακολούθησης είναι συνυφασμένο με αλγόριθμους ανίχνευσης οι οποίοι χρησιμοποιούνται ως βάση για τον υπολογισμό των ροών αντικείμενο. Αυτό αποτελεί μια σημαντική καινοτομία αυτής της εργασίας, δεδομένου ότι οι ροές που προκύπτουν είναι ειδικά προσαρμοσμένες στις συγκεκριμένες ιδιότητες ενός αντικειμένου (π.χ. εργάτης σε μία αυτοκινητοβιομηχανία). Η ενσωμάτωση της αυτό-αρχικοποίησης [70] μπορεί να βελτιώσει σημαντικά την αυτοματοποίηση του αλγόριθμου, καθιστώντας την μέθοδο ισχυρή σε δυναμικές αλλαγές του περιβάλλοντος.

6.1 Προτεινόμενο Σύστημα Παρακολούθησης

Στην ενότητα αυτή περιγράφεται μια αξιόπιστη προσέγγιση παρακολούθησης, η οποία χρησιμοποιεί την ροή αντικείμενου ως μοντέλο περιγραφής και κίνησης και ένα επιπλέον μοντέλο περιγραφής που βασίζεται σε μία παραγωγική μέθοδο.

6.1.1 Φίλτρο Σωματιδίων

Τα φίλτρα σωματιδίων έχουν αποδειχθεί ότι είναι μια καλή λύση για τα συστήματα παρακολούθησης. Το πρόβλημα παρακολούθησης αντικειμένων μπορεί να οριστεί από μία Μπεϋζιανή προοπτική, όπου γίνεται ο αναδρομικός υπολογισμός της κρυφής κατάστασης, η οποία σχετίζεται με μία σειρά από θορυβώδη παρατηρήσεις. Στο πλαίσιο των φίλτρων σωματιδίων, η κατανομή *posterior* $p(X_{0:t} | Y_{1:t})$ μίας ακολουθίας παρατηρήσεων $Y_{1:t}$ που εξαρτάται από μία σειρά κρυμμένων καταστάσεων $X_{0:t}$ (*hidden states*) μπορεί να υπολογιστεί με την βοήθεια προσομοιώσεων *Monte Carlo*, με τα κατάλληλα βήματα πρόβλεψης και ενημέρωσης. Τα τυχαία δείγματα που προέρχονται από μία κατανομή $X_t^i \sim q(\cdot)$ παράγονται μέσω της τεχνικής *importance sampling*. Για κάθε δείγμα-σωματίδιο, η εξίσωση βάρους εκφράζεται ως,

$$w_t^i = w_{t-1}^i \frac{p(Y_t | X_{0:t}^i, Y_{1:t-1}) p(X_t^i | X_{0:t-1}^i, Y_{1:t-1})}{q(X_t^i | X_{0:t-1}^i, Y_{1:t})} \quad (6.1)$$

Ωστόσο, αρκετές φορές θεωρείται σκόπιμο ότι η σειρά παρατηρήσεων $Y_{1:t}$ είναι ανεξάρτητη και η σειρά των κρυμμένων καταστάσεων ακολουθεί μία σειρά *Markov* πρώτου βαθμού (δηλαδή, $p(Y_t | X_{0:t}^i, Y_{1:t-1}) = p(Y_t | X_t^i)$ και $p(X_t^i | X_{0:t-1}^i, Y_{1:t-1}) = p(X_t^i | X_{t-1}^i)$).

Επίσης, ένα από τα προβλήματα που εμφανίζονται στην λειτουργία του φίλτρου σωματιδίων είναι η διακύμανση των βαρών. Αυτό το πρόβλημα μπορεί να λυθεί με την χρήση ενός επιπλέον βήματος επαναδειγματοληψίας (*re-sampling*).

6.1.2 Ροή Αντικειμένου για Αντικείμενα σε Βιομηχανικό Περιβάλλον

Στην ενότητα αυτή, παρουσιάζεται η ροή αντικειμένου (βλ. Κεφάλαιο 5) για αντικείμενα σε βιομηχανικό περιβάλλον. Η προσέγγιση αυτή μπορεί να χρησιμοποιηθεί για την παρακολούθηση και για την εκτίμηση της μετατόπισης ενός κινούμενου αντικειμένου ενδιαφέροντος. Επιπλέον, μπορεί να προσφέρει βελτιωμένες επιδόσεις εκτίμησης, δεδομένου ότι δεν επηρεάζεται από άλλες άσχετες κινήσεις μέσα στη σκηνή. Σε αυτή την εργασία, η ροή αντικειμένου ενσωματώνεται σε ένα φίλτρο σωματιδίων (βλ. Κεφάλαιο 2.4) και μπορεί να χρησιμοποιηθεί τόσο για το μοντέλο κίνησης όσο και για τη λήψη μετρήσεων (π.χ. περιγραφή αντικειμένου).

Ένας δυαδικός ταξινομητής (*margin-based binary classifier*) μπορεί να εκπαιδευτεί στην μετατόπιση αντικειμένου χρησιμοποιώντας ένα σύνολο θετικών και αρνητικών δειγμάτων. Τα θετικά δείγματα X^+ περιέχουν πληροφορίες για τον τρόπο που ένα αντικείμενο μεταλλάσσεται μέσα στο χρόνο. Συγκεκριμένα δημιουργούνται από την συνένωση δύο δειγμάτων που περιέχουν το αντικείμενο ενδιαφέροντος σε δύο διαφορετικές χρονικές στιγμές. Αυτά τα δείγματα μπορεί να παρέχονται είτε χειροκίνητα είτε από ένα επαληθευτή αντικείμενου πραγματικού χρόνου. Από την άλλη πλευρά, τα αρνητικά δείγματα εκπαίδευσης χωρίζονται σε δύο υποσύνολα, $X^- = X_{\text{background}}^- \cup X_{\text{object}}^-$. Το πρώτο υποσύνολο $X_{\text{background}}^-$ περιέχει ζεύγη εικόνας από το περιβάλλον. Το δεύτερο υποσύνολο X_{object}^- περιέχει δείγματα με το πραγματικό αντικείμενο και ένα κομμάτι του αντικειμένου σε ένα διαφορετικό πλαίσιο. Παραδείγματα των θετικών και αρνητικών δειγμάτων απεικονίζεται στο Σχήμα 20.



Σχήμα 20. Παραδείγματα εικόνων που χρησιμοποιούνται για την εκπαίδευση του ταξινομητή.

Έστω ότι, $C(p_t, p_{t-1})$ είναι η απόκριση του εκπαιδευμένου ταξινομητή για ένα ζευγάρι εικόνων, όπου p_t είναι ένα κομμάτι της παρούσας εικόνας και p_{t-1} είναι ένα κομμάτι που ανήκει στην κοντινή περιοχή της προηγούμενης εικόνας. Ορίζουμε την μετατόπιση ενός αντικειμένου ενδιαφέροντος στον x και y άξονα ως Δx και Δy αντίστοιχα. Αυτή η μετατόπιση υπολογίζεται ως το συνολικό άθροισμα των αποστάσεων εντός μίας τοπικής περιοχής Ω , δηλαδή,

$$\begin{pmatrix} \Delta x_{obj}(p_t) \\ \Delta y_{obj}(p_t) \end{pmatrix} = \frac{1}{\sum_{p_{t-1} \in \Omega} C(p_t, p_{t-1})} \sum_{p_{t-1} \in \Omega} C(p_t, p_{t-1}) \begin{pmatrix} dx \\ dy \end{pmatrix} \quad (6.2)$$

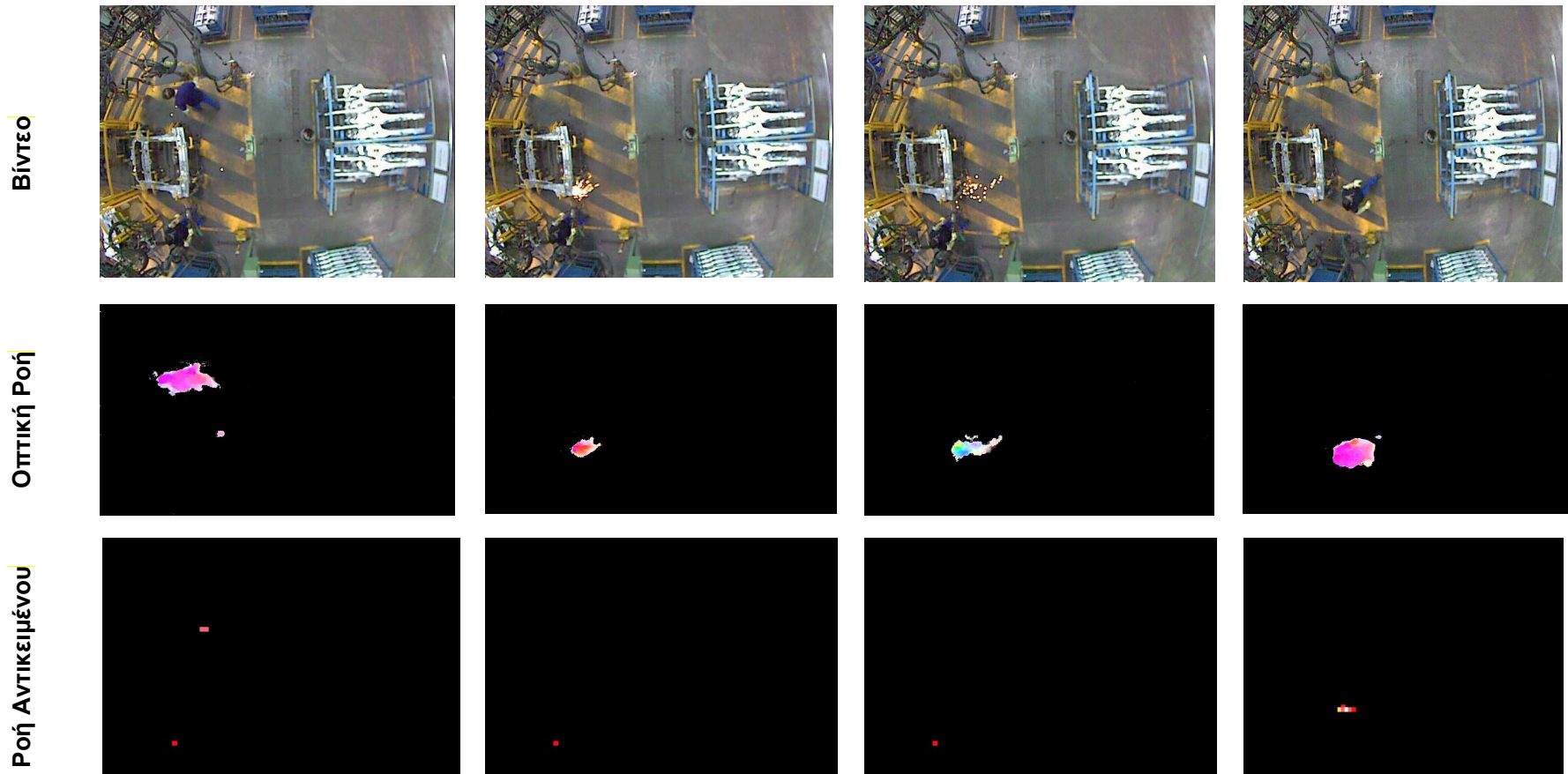
Άρα, η εκτίμηση μετατόπισης για ένα κινούμενο αντικείμενο μπορεί να υπολογιστεί ως,

$$(\hat{x}, \hat{y})_t = \left(\Delta x_{obj}(p_t), \Delta y_{obj}(p_t) \right) + (\hat{x}, \hat{y})_{t-1} \quad (6.3)$$

Με στόχο να μειωθούν οι ακραίες τιμές, οι αποκρίσεις του ταξινομητή εντός της περιοχής Ω πρέπει να ξεπερνάνε ένα συγκεκριμένο κατώφλι δηλαδή,

$$c_{\sigma\sigma_j}^-(p_t) = \frac{1}{|\Omega|} \sum_{p_{t-1} \in \Omega} c^{\sigma}(p_t, p_{t-1})^2 \quad (6.4)$$

όπου $c^{\sigma}(p_t, p_{t-1})^2 = \max(0, c(p_t, p_{t-1}))$



Σχήμα 21. Παρουσίαση της ροής αντικειμένου για το βιομηχανικό περιβάλλον. Οι προσεγγίσεις της οπτικής ροής αποπροσανατολίζονται, από εξωγενείς παράγοντες (π.χ. σπίθες από μηχανή συγκόλλησης), ειδικά όταν κινούνται προς την ίδια κατεύθυνση με το αντικείμενο-του-ενδιαφέροντος (δεύτερη γραμμή). Η τρίτη σειρά παρουσιάζει την εκτιμώμενη ροή αντικειμένου, η οποία επικεντρώνεται μόνο στο αντικείμενο-του-ενδιαφέροντος.

Η σελίδα αυτή είναι σκόπιμα λευκή.

6.1.3 Μοντέλο περιγραφής αντικειμένων βάσει Παραγωγικού prior

Στο προτεινόμενο σύστημα παρακολούθησης εισάγεται ένα νέο μοντέλο περιγραφής αντικειμένου. Αυτό το μοντέλο χρησιμοποιείται για να υπολογίσει τον συσχετισμό μεταξύ των δειγμάτων εισροής και ενός μοντέλου, το οποίο δημιουργείται με μια παραγωγική μέθοδο. Συγκεκριμένα, η αναπαράσταση αυτή μπορεί να επιτευχθεί χρησιμοποιώντας την μέθοδο *GPLVM* (βλ. Κεφάλαιο 6.1.3.1). Το θετικό αυτής της προσέγγισης είναι ότι μπορεί να χειριστεί δραματικές αλλαγές στην εμφάνιση του αντικειμένου, χρησιμοποιώντας ελάχιστα δείγματα για την εκπαίδευση αυτού του μοντέλου.

6.1.3.1 Gaussian Process Latent Variable Model

Στο κεφάλαιο αυτό περιγράφεται η τεχνική της *GPLVM*. Η *GPLVM* ταιριάζει ένα υψηλό επίπεδο δεδομένων σε ένα χαμηλότερο υπό-επίπεδο, χρησιμοποιώντας γκαουσιανές διαδικασίες (*Gaussian Processes*). Η τεχνική αυτή χρησιμοποιείται συχνά για εφαρμογές μείωσης διαστάσεων [60], αλλά μπορεί επίσης να παρέχει μια ισχυρή παραγωγική πιθανολογική προσέγγιση σε διάφορες εφαρμογές όρασης υπολογιστών [61],[25]. Σε αντίθεση με την προσέγγιση *PCA* [28], η *GPLVM* μπορεί επίσης να αντιμετωπίσει περιπτώσεις όπου η υπόθεση για ισόνομα σημεία δεδομένων (*identically distributed data*) έχει παραβιαστεί. Αυτό σημαίνει ότι η μέθοδος *GPLVM* επιτρέπει μη-γραμμικές χαρτογραφήσεις υπέρ-επιπέδων. Αυτή η τεχνική παρουσιάζεται παρακάτω.

Η μη-γραμμική *GPLVM* μέθοδος κατασκευάζει μέσω μίας $N \times N$ *radial basis* συνάρτησης (*RBF*) ένα πίνακα *kernel*, με καταχωρήσεις $k_{ij} = k(x_i, x_j)$. Αυτός ο πίνακας *kernel* μπορεί να χαρτογραφήσει τα σημεία ενός χαμηλού υπέρ-επιπέδου $X \subset \mathbb{R}^{N \times d}$ (*latent space*) στο υπέρ-επίπεδο των δεδομένων $Y \in \mathbb{R}^{N \times d}$ (*data space*),

όπου $q \ll d$. Συγκεκριμένα, μία καταχώρηση $k_{ij}(\cdot, \cdot)$ του πίνακα *kernel* μπορεί να οριστεί ως,

$$k_{ij}(x_i, x_j) = \beta_1 \exp\left(-\frac{\beta_2}{2} \|x_i - x_j\|^2\right) + \frac{\delta_{x_i x_j}}{\beta_3} + \beta_4 \quad (6.5)$$

Όπου x_i και x_j είναι οι i και j κολώνες του πίνακα X^T και $\beta_{1, \dots, 4}$ είναι οι υπέρ-παράμετροι του *kernel*. Η μέθοδος *GPLVM* ελαχιστοποιεί την επόμενη συνάρτηση *log-likelihood* σε σχέση με $X \in \mathbb{R}^{N \times q}$ και $\beta_{1, \dots, 4}$.

$$L = -\left(\frac{d}{2} \ln |K_Y| + \frac{1}{2} \text{tr}(K_Y^{-1} Y Y^T) + \frac{1}{2} \sum_{i=1}^N \|x_i\|^2 + \sum_{i=1}^4 \ln \beta_i\right) \quad (6.6)$$

Η αναπαράσταση ενός στοιχείου εισόδου (*test point*) $y^i \in \mathbb{R}^{1 \times d}$ σε ένα χαμηλό υπέρ-επίπεδο μπορεί να επιτευχθεί μέσω ενός επιπροσθέτου σταδίου βελτιστοποίησης. Ωστόσο αυτό μπορεί να αποφευχθεί χρησιμοποιώντας την μέθοδο *GPLVM* με *back constraints* [62]. Αυτό μπορεί να γίνει ελαχιστοποιώντας την Εξίσωση (6.5) σε σχέση με την επόμενη συνάρτηση,

$$x_{ij} = g_j(y_i; \mathbf{a}) = \sum_{m=1}^m a_{jm} k_{bc}(y_i - y_m) \quad (6.7)$$

Όπου x_{ij} είναι το j στοιχείο του x_i και $k_{bc}(\cdot, \cdot)$ είναι ένα *RBF kernel*, το οποίο μπορεί να οριστεί ως,

$$k_{bc}(y_i - y_m) = \exp\left(-\frac{\gamma}{2} \|y_i - y_m\|^2\right) \quad (6.8)$$

Άρα η αναπαράσταση χαμηλού υπέρ-επιπέδου \mathbf{y}^f μπορεί να υπολογιστεί ως, $\mathbf{x}^f = \mathbf{g}_f(\mathbf{y}^f; \mathbf{a})$.

6.1.3.2 Συσχέτιση κινούμενου αντικειμένου με GPLVM

Όπως αναφέρθηκε, η μέθοδος *GPLVM* μπορεί να αξιοποιηθεί για την ανάπτυξη ενός μοντέλου περιγραφής αντικειμένων. Για την εκπαίδευση ενός απλού μοντέλου *GPLVM* (βλ. Εξίσωση (6.6)) μπορεί να χρησιμοποιηθεί μία συλλογή από εικόνες $\mathbf{Y} \in \mathbb{R}^{N \times d}$, η οποία αποτελείται από ένα αριθμό αντιπροσωπευτικών δειγμάτων που περιέχουν το αντικείμενο υπό μελέτη σε διαφορετικές πόζες. Επίσης, αυτά τα δείγματα μπορούν να ληφθούν χρησιμοποιώντας την τεχνική *key frame extraction* [71].

Η αναπαράσταση \mathbf{x}^f της εικόνας \mathbf{p}_z , που ανήκει στο σωματίδιο \mathbf{X}_z^f μπορεί να υπολογιστεί μέσω αντίστροφης χαρτογράφησης (βλ. Εξίσωση (6.7)). Στη συνέχεια, υπολογίζεται η απόσταση *Mahalanobis* για να συγκριθεί το διάνυσμα \mathbf{x}^f με το διάνυσμα μ_x , το οποίο περιέχει το μέσο όρο της κάθε στήλης του πίνακα $\mathbf{X} \in \mathbb{R}^{N \times q}$.

Επίσης, μία παρόμοια προσέγγιση χρησιμοποιήθηκε στις δημοσιεύσεις [72] και [29].

Αυτό το μέτρο σύγκρισης μπορεί να εκφραστεί ως,

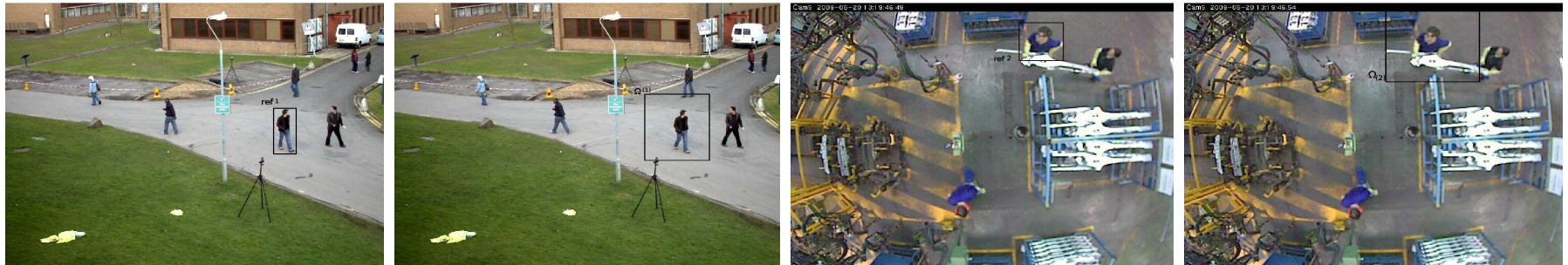
$$d(\mathbf{x}^f, \mu_x) = \sqrt{(\mathbf{x}^f - \mu_x) S_x^{-1} (\mathbf{x}^f - \mu_x)} \quad (6.9)$$

Ωστόσο, δεδομένου ότι είναι μια πιθανοθεωρητική προσέγγιση (δηλαδή, φίλτρα σωματιδίων), άρα η απόσταση (βλ. Εξίσωση (6.9)) πρέπει να εκφραστεί με όρους κατανομής πιθανοτήτων. Προς αυτή την κατεύθυνση, η απόσταση *Mahalanobis* θεωρείται ότι ακολουθεί μία εκθετική κατανομή,

$$p(Y_t^a | X_t^a) = \exp(-d^2(x', \mu_x)) \quad (6.10)$$

6.1.4 Ενσωμάτωση Ροής Αντικειμένου σε Φίλτρο Σωματιδίων

Στην ενότητα αυτή περιγράφεται το πλαίσιο για την ενσωμάτωση ενός εκτιμητή κίνησης σε ένα φίλτρο σωματιδίων. Όπως αναφέρθηκε σε προηγούμενη ενότητα, είναι συνήθης πρακτική να επιλέγεται ως μοντέλο κίνησης η δυναμική κατάσταση (βλ. Κεφάλαιο 2.4). Ωστόσο, όταν το αντικείμενο που παρακολουθείται κάνει μία απότομη κίνηση, η μοντελοποίηση μέσω της δυναμικής κατάστασης μπορεί να είναι αναποτελεσματική με αποτέλεσμα πολλά σωματίδια να χαθούν. Ως εκ τούτου, η επιλογή ενός αποτελεσματικού μοντέλου κίνησης είναι ιδιαίτερος σημαντική για την σωστή λειτουργία του φίλτρου σωματιδίων. Σε αυτή την προσέγγιση χρησιμοποιούνται οι εκτιμήσεις κίνησης που παρέχονται από την προτεινόμενη μέθοδο ροής αντικειμένου (βλ. Κεφάλαιο 6.1.2).

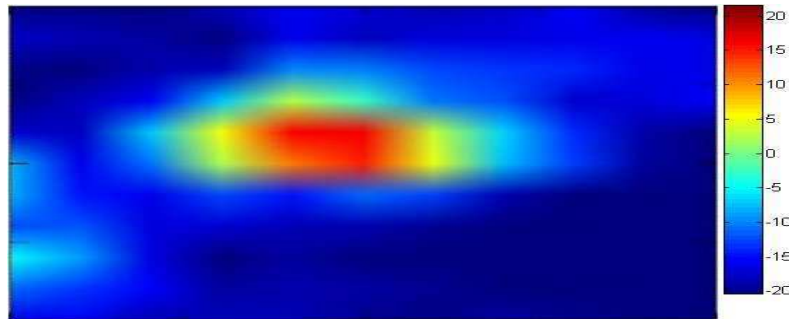


(α) time $t - 1$

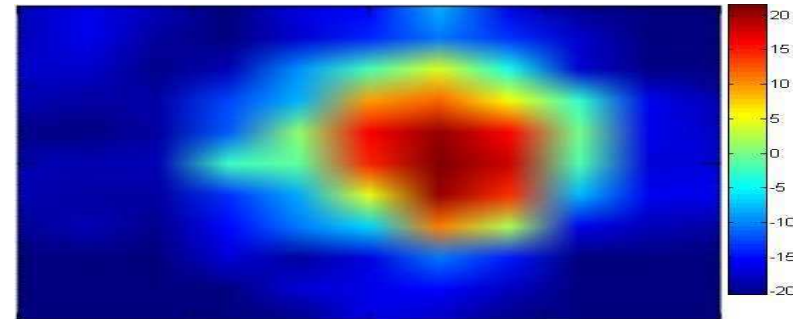
(β) time t

(γ) time $t - 1$

(δ) time t



(ε) απόκριση στην περιοχή $\Omega(1)$



(ζ) απόκριση στην περιοχή $\Omega(2)$

Σχήμα 22. Παρουσίαση της διαδικασίας εκτίμησης για ένα συγκεκριμένο σωματίδιο (*particle*). Η δειγματοληψία γίνεται με την βάση των περιοχών απόκρισης(ε) και (ζ).

Η σελίδα αυτή είναι σκόπιμα λευκή.

Πιο συγκεκριμένα, για κάθε σωματίδιο (*particle*) η διαδικασία της δειγματοληψίας εκτελείται βάσει των εκτιμήσεων $\mathbf{x}_t^s, \mathbf{y}_t^s$ που παρέχονται από την Εξίσωση (6.3).

Αυτές οι εκτιμήσεις υπολογίζονται χρησιμοποιώντας τις αποκρίσεις του ταξινομητή για συγκεκριμένες περιοχές της εικόνας. Το Σχήμα 22(ε),(ζ) παρουσιάζει τις αποκρίσεις για δύο διαφορετικές περιοχές $\Omega(1)$ και $\Omega(2)$ ενός εκπαιδευμένου ταξινομητή σε δύο διαφορετικές εικονοσειρές, με τις οι κόκκινες περιοχές να είναι εκείνες, όπου το αντικείμενο είναι πιθανό να εμφανιστεί. Στη συνέχεια, τα νέα σωματίδια (*particles*) διαδίδονται χρησιμοποιώντας μία γκαουσιανή κατανομή $N(\mathbf{x}_t^s; \mathbf{x}_t, \sigma_x^2)$ και $N(\mathbf{y}_t^s; \mathbf{y}_t, \sigma_y^2)$ με προκαθορισμένη διακύμανση σ_x^2 και σ_y^2 και με μέση τιμή ίση με την μετατόπιση \mathbf{x}_t και \mathbf{y}_t που υπολογίστηκε από την ροή αντικειμένου (βλ. Σχήμα 22).

Σε αυτό το σύστημα παρακολούθησης, η κατάσταση του φίλτρου σωματιδίων ορίζεται ως $X_t = \{\mathbf{x}_t, \mathbf{y}_t\}$, το οποίο περιγράφει μία θέση στην εικόνα την χρονική στιγμή t . Το βάρος του κάθε σωματιδίου (*particle weight*) μπορεί να υπολογιστεί από την παρακάτω εξίσωση,

$$w_t^i \propto w_{t-1}^i p(Y_t^{o_b j} | X_t^i) p(Y_t^a | X_t^i) \quad (6.11)$$

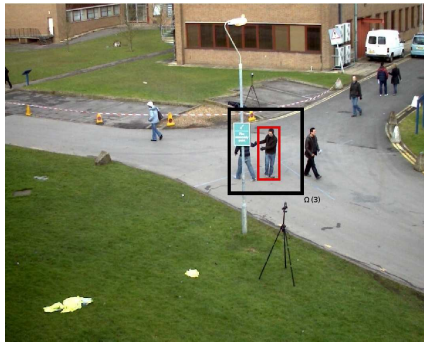
όπου, $p(Y_t^{o_b j} | X_t^i) = C_{o_b j}^-(p_t)$ είναι απόκριση του ταξινομητή για ένα ζεύγος εικόνων και $p(Y_t^a | X_t^i)$ η συσχέτιση του κινούμενου αντικειμένου με την μέθοδο *GPLVM*. Θα πρέπει να σημειωθεί ότι η διαδικασία της δειγματοληψίας επαναλαμβάνεται σε κάθε βήμα του αλγορίθμου. Επίσης σωματίδια (*particles*) με ασήμαντο βάρος αντικαθιστώνται από εκείνα τα οποία έχουν συνεισφορά στην τελική *posterior* κατανομή. Συγκεκριμένα υιοθετείται η τεχνική της συστηματικής

επαναδειγματοληψίας (*systematic re-sampling*), η οποία είναι αρκετά αποτελεσματική για τέτοιες εφαρμογές.

6.1.5 Μηχανισμός Έλεγχου Δειγμάτων Εκπαίδευσης

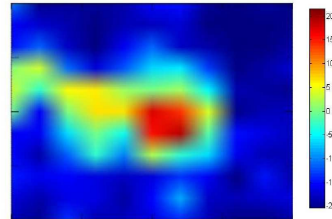
Σε αυτή την ενότητα προτείνεται ένας αποτελεσματικός μηχανισμός ελέγχου δειγμάτων εκπαίδευσης πραγματικού χρόνου. Πιο συγκεκριμένα, ένα προ-εκπαιδευμένο μοντέλο περιγραφής αντικειμένου μπορεί να ενημερώνεται σε πραγματικό χρόνο χρησιμοποιώντας το αποτέλεσμα από το σύστημα παρακολούθησης. Ωστόσο, αυτό το αποτέλεσμα μπορεί να περιέχει θορυβώδη χαρακτηριστικά (π.χ. το πραγματικό αντικείμενο να επικαλύπτεται από άλλο κινούμενο αντικείμενο ή από κάποιο εμπόδιο). Αυτό το γεγονός μπορεί να οδηγήσει σε εκτροπή (*target drift*) του συστήματος παρακολούθησης. Ως εκ τούτου, το κλειδί για την καλή απόδοση ενός συστήματος παρακολούθησης είναι η ανάπτυξη ενός μηχανισμού επαλήθευσης του δείγματος εκπαίδευσης, ο οποίος έχει τη δυνατότητα να απορρίπτει τα δείγματα εκείνα που περιέχουν θορυβώδη χαρακτηριστικά.

Θετικά δείγματα

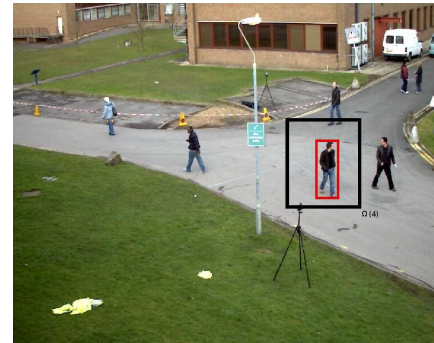


(α) Αρχική εικόνα

kurtosis: 0.779008

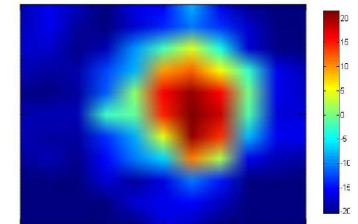


(β) Περιοχή Απόκρισης



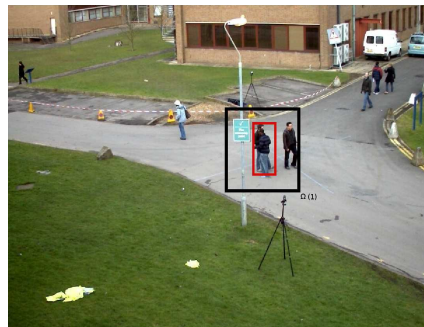
(γ) Αρχική εικόνα

kurtosis: 1.403466



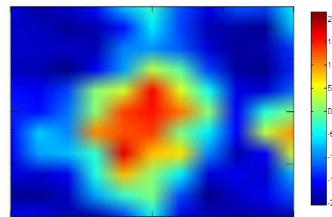
(δ) Περιοχή Απόκρισης

Αρνητικά δείγματα



(ε) Αρχική εικόνα

kurtosis: -0.213594

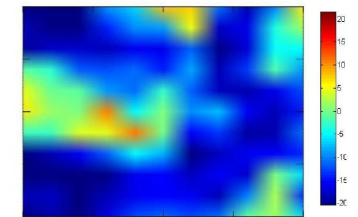


(ζ) Περιοχή Απόκρισης



(η) Αρχική εικόνα

kurtosis: -0.761835



(θ) Περιοχή Απόκρισης

Σχήμα 23. Ενδεικτικά παραδείγματα του μηχανισμού επαλήθευσης του δείγματος εκπαίδευσης. Χαμηλές τιμές κύρτωσης λαμβάνονται όταν ο άνθρωπος που παρακολουθείται επικαλύπτεται από άλλον άνθρωπο (ε) ή από ένα εμπόδιο (η). Αντίθετα, υψηλότερες τιμές κύρτωσης (β), (δ) παρατηρούνται στις περιοχές όπου το αντικείμενο δεν περιβάλλεται από θορυβώδη χαρακτηριστικά. Μάθηση πραγματικού χρόνου γίνεται με τα δείγματα που έχουν θετικές τιμές κύρτωσης.

Η σελίδα αυτή είναι σκόπιμα λευκή.

Η προσέγγιση για τον έλεγχο του δείγματος εκπαίδευσης εκμεταλλεύεται τις πληροφορίες από τα διαγράμματα απόκρισης (δηλαδή, τις αποκρίσεις ταξινομητή, βλ. Σχήμα 24 (β), (δ), (ζ) και (θ)) για την περιοχή που περικλείει ένα δείγμα εκπαίδευσης. Αυτά τα διαγράμματα παρέχονται από ένα προ-εκπαιδευμένο ταξινομητή επάνω στην μετατόπιση αντικειμένων. Στα παραπάνω διαγράμματα, οι τιμές απόκρισης δείχνουν την πιθανότητα για το αντικείμενο που παρακολουθείται να βρίσκεται εντός μίας συγκεκριμένης περιοχής. Ωστόσο, χαμηλότερες τιμές απόκρισης παρατηρούνται, όταν το αντικείμενο που παρακολουθείται επικαλύπτεται από άλλο (βλ. Σχήμα 25 (ε)) ή από ένα εμπόδιο (βλ. Σχήμα 26 (η)). Ως εκ τούτου, η ύπαρξη περιοχών που συσσωρεύονται υψηλές τιμές σημαίνει την ύπαρξη του αντικειμένου εντός αυτής, χωρίς όμως αυτό να περιβάλλεται από θορυβώδη χαρακτηριστικά.

Ο προτεινόμενος μηχανισμός έλεγχου δείγματος εκπαίδευσης εκμεταλλεύεται την κύρτωση. Συγκεκριμένα, η κύρτωση μπορεί να χρησιμοποιηθεί για την αξιολόγηση της κυρτότητας-*peakedness* ενός διαγράμματος απόκρισης. Ένα δείγμα εκπαίδευσης επιλέγεται όταν η κύρτωση ενός διαγράμματος απόκρισης είναι θετική. Αντίθετα, αν η κύρτωση είναι αρνητική, το δείγμα εκπαίδευσης απορρίπτεται από τον μηχανισμό. Έστω ότι $\mathbf{R} = \{r_1, \dots, r_n\}$ είναι ένα σύνολο που περιέχει τις αποκρίσεις του ταξινομητή για μία συγκεκριμένη περιοχή. Η κύρτωση δείγματος g για το σύνολο \mathbf{R} με μέση τιμή \bar{r} ορίζεται ως,

$$g = \frac{1/n \sum_{i=1}^n (r_i - \bar{r})^4}{(1/n \sum_{i=1}^n (r_i - \bar{r})^2)^2} - 3 \quad (6.12)$$

Ο παραπάνω μηχανισμός ενσωματώθηκε στο μηχανισμό παρακολούθησης που περιγράφηκε στο κεφάλαιο 6.1.4.

6.2 Πειραματικά Αποτελέσματα

6.2.1 Πειραματικά Αποτελέσματα για Σύνθετο Βιομηχανικό Περιβάλλον

Σε αυτή την ενότητα παρουσιάζονται τα ποιοτικά και ποσοτικά αποτελέσματα του προτεινόμενου συστήματος παρακολούθησης. Με σκοπό να διασφαλιστεί η αποτελεσματικότητα της προτεινόμενης μεθόδου, χρησιμοποιήθηκαν αρκετά περίπλοκες εικονοσειρές με διαφορετικές συνθήκες και σενάρια, οι οποίες μαγνητοσκοπήθηκαν σε μία αυτοκινητοβιομηχανία (βλ. Σχήμα 19). Θα πρέπει να σημειωθεί ότι σε περίπτωση που το σύστημα παρακολούθησης παρεκκλίνει του αρχικού στόχου θα πρέπει να γίνεται επαναρχικοποίηση (*re-initialization*) του αλγορίθμου, είτε χρησιμοποιώντας την βοήθεια των κατάλληλων χρηστών ή χρησιμοποιώντας τους κατάλληλους αλγορίθμους επαναρχικοποίησης [70]. Κατά την διάρκεια της λειτουργίας του συγκεκριμένου συστήματος παρακολούθησης καμία τεχνική μάθησης πραγματικού χρόνου δεν χρησιμοποιείται, άρα δεν υπάρχει ο κίνδυνος αποπροσανατολισμού (δηλαδή να προσαρμοστεί ο αλγόριθμος σε άλλο διαφορετικό αντικείμενο, εκτός από εκείνο που αρχικά παρακολουθείται). Επιπλέον γίνεται σύγκριση της προτεινόμενης μεθόδου με άλλες, όπως με ένα φίλτρο σωματιδίων που χρησιμοποιεί ένα απλό μοντέλο χρώματος (*Color Tracker*), με δύο διαφορετικές τεχνικές *Adaboost* (*OAB1* και *OAB5*) [4] και με τον ιχνηλάτη *MIL* [73]. Όλα τα πειράματα έγιναν σε ένα *i7 @ 2,67 GHz* υπολογιστή με *4 Gb* μνήμης.

Το προτεινόμενο σύστημα παρακολούθησης είναι ικανό να παρακολουθεί παραπάνω από ένα αντικείμενο στην σκηνή. Συγκεκριμένα, η ροή αντικειμένου μπορεί να παρέχει πληροφορίες κίνησης για πολλαπλά αντικείμενα ενδιαφέροντος μέσα στην σκηνή. Αυτό είναι εμφανές από το Σχήμα 21 (βλ. τρίτη σειρά). Ωστόσο προβλήματα μπορεί να προκύψουν σε σενάρια που περιλαμβάνουν πολυσύχναστες σκηνές, ειδικά

όταν πολλαπλά αντικείμενα κινούνται προς μία συγκεκριμένη κατεύθυνση. Θα πρέπει όμως να σημειωθεί ότι τέτοια περίπτωση δεν υπάρχει στην εικονοσειρά που μαγνητοσκοπήθηκε στην αυτοκινητοβιομηχανία (βλ. Σχήμα 19).

Η προτεινόμενη μέθοδος μπορεί να χρησιμοποιηθεί είτε σε εφαρμογές που περιλαμβάνουν στατικές κάμερες ή σε εφαρμογές που περιλαμβάνουν κινούμενες κάμερες. Η τεχνική *boosting for feature selection* [3] και τα χαρακτηριστικά εικόνας *Haar* επιλέχθηκαν για την μάθηση της ροής αντικειμένου. Για τα παρακάτω πειράματα η εκτίμηση της ροής αντικειμένου έγινε μέσω ενός πυκνού πλέγματος που αποτελείται από 81×81 επικαλυπτόμενα κελία των ίδιων διαστάσεων. Επίσης, χρησιμοποιείται μία περιοχή Ω η οποία αποτελείται από 5×5 κελία. Για το φίλτρο σωματιδίων χρησιμοποιούνται 30 σωματίδια συγκεκριμένου πλάτους και ύψους και διακύμανσης $\sigma_x^2, \sigma_y^2 = 2,0$. Για το μοντέλο παρατήρησης *GPLVM* χρησιμοποιήθηκε ο αλγόριθμος *scaled conjugate gradients*, ο οποίος τέθηκε σε λειτουργία για 400 κύκλους και η παράμετρος γ τέθηκε ίση με 0,0001 (βλ. Εξίσωση (6.7)). Τέλος, ο πίνακας X^T (βλ. Κεφάλαιο 6.1.3.1) αρχικοποιείται με την βοήθεια της μεθόδου *PCA*.

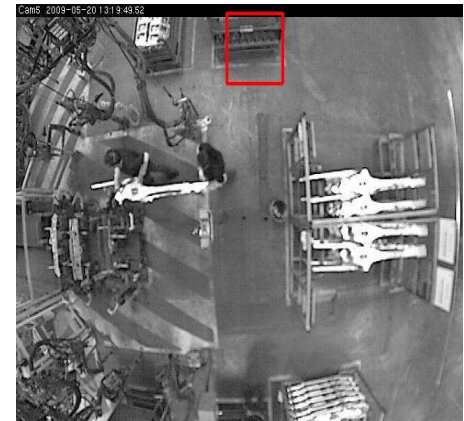
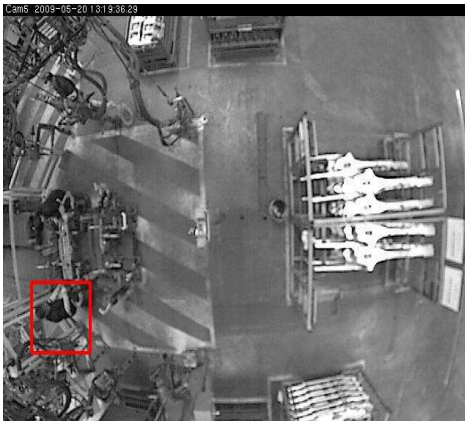
Έγινε η συλλογή πέντε διαφορετικών εικονοσειρών μέσω μίας πανοραμικής κάμερας, η οποία τοποθετήθηκε στο ταβάνι μίας αυτοκινητοβιομηχανίας. Αυτές οι εικονοσειρές έχουν ανάλυση 680×480 pixel και περιέχουν δύσκολα σενάρια με αρκετά διαφορετικά κινούμενα αντικείμενα στο περιβάλλον. Για όλες τις εικονοσειρές, η εκπαίδευση της ροής αντικειμένου γίνεται με $|X^+| \approx 1,000$ θετικά, $|X^-| \approx 4,000$ αρνητικά και πολυάριθμα δείγματα από το περιβάλλον. Επίσης, η

μοντελοποίηση της εμφάνισης για το αντικείμενο ενδιαφέροντος γίνεται με **400** δείγματα με ανάλυση **70 x 70 pixel**.

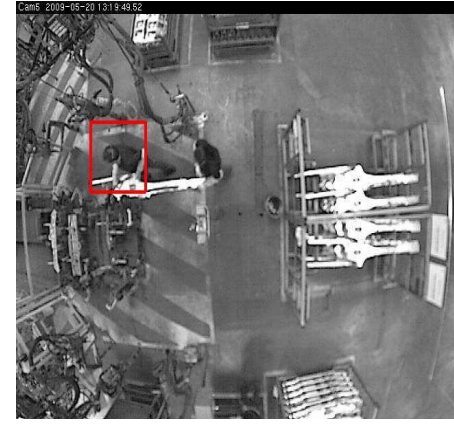
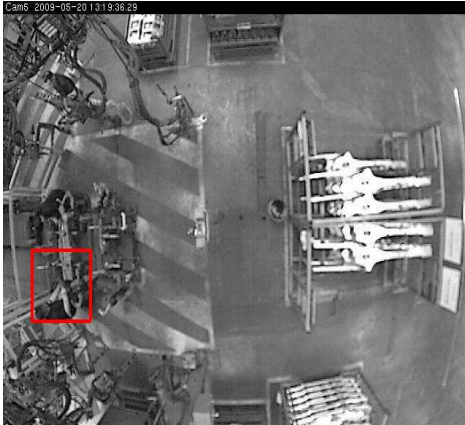
Βίντεο



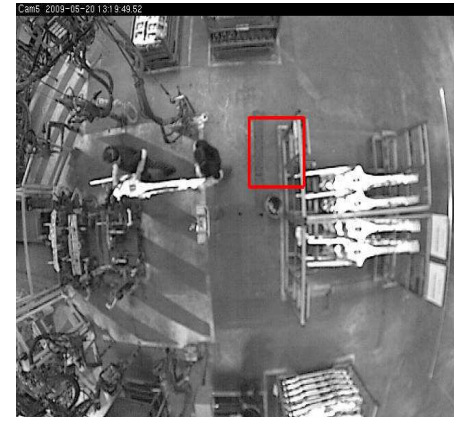
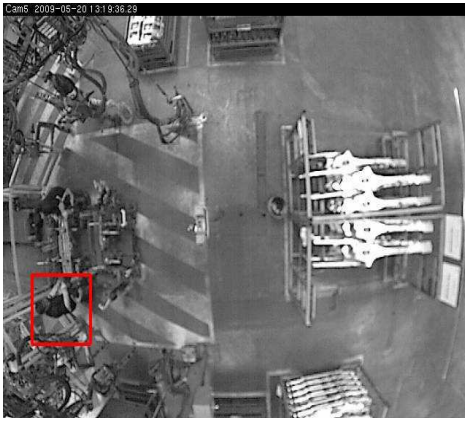
ΟΑΒ 1



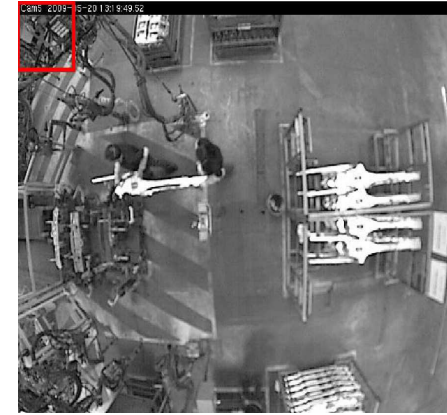
OAB 5



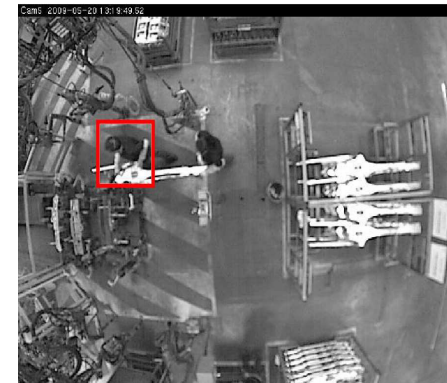
MIL



Color Tracker



Προτεινόμενη
Μέθοδος

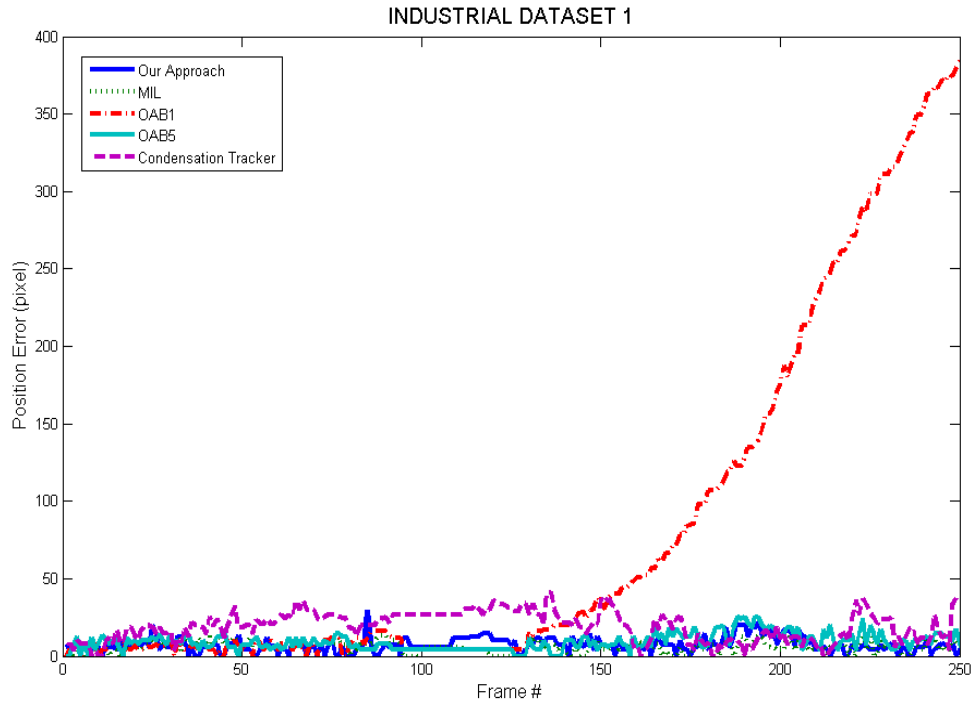


Σχήμα 27. Ενδεικτικά αποτελέσματα. Τα συστήματα *OAB1* και *MIL* (δεύτερη και τέταρτη γραμμή) παρασύρονται μακριά από το στόχο, δεδομένου ότι επηρεάζονται από το βιομηχανικό περιβάλλον. Το *OAB5* (τρίτη γραμμή) δεν παραμένει απολύτως εστιασμένο στο στόχο για όλη την ακολουθία εικόνων. Ο *Color tracker* (πέμπτη γραμμή) έχει εντελώς ασταθή απόδοση, δεδομένου ότι επηρεάζεται από άλλα χρώματα (π.χ. σπινθήρες από την μηχανή συγκόλλησης). Η προτεινόμενη προσέγγισή (έκτη γραμμή) παραμένει επικεντρωμένη στον κινούμενο στόχο για όλη την ακολουθία.

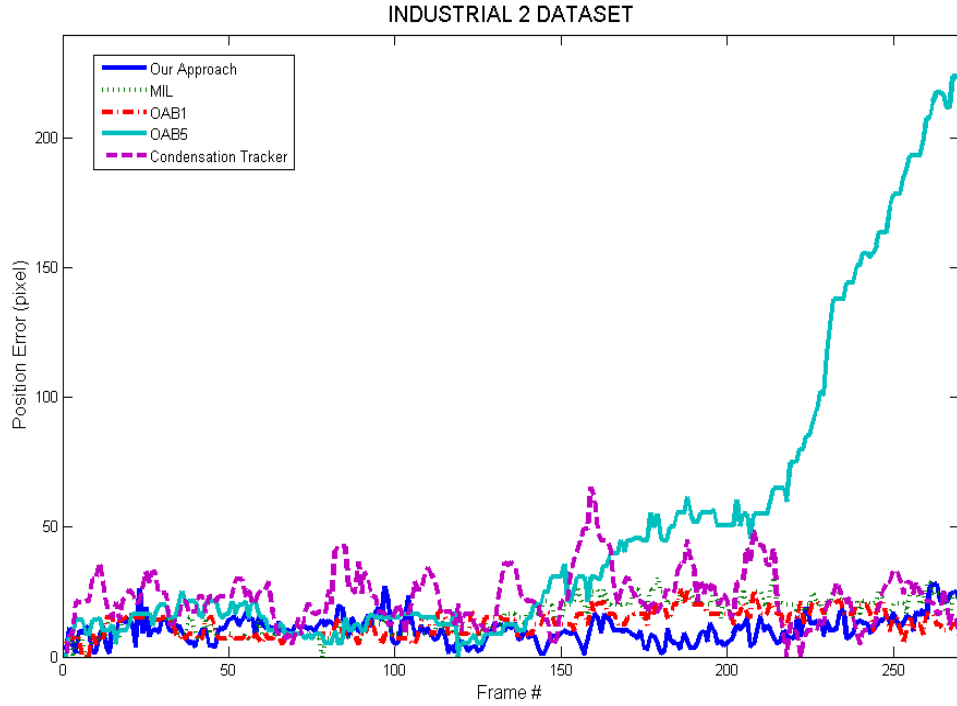
Η σελίδα αυτή είναι σκόπιμα λευκή.

Τα αποτελέσματα από τα διάφορα συστήματα παρακολούθησης απεικονίζονται επεξηγηματικά στο Σχήμα 27. Όπως μπορεί να διαπιστωθεί, η προτεινόμενη προσέγγισή εντοπισμού (έκτη γραμμή) έχει τη δυνατότητα να παραμείνει συγκεντρωμένη στο κινούμενο στόχο, ενώ η σταθερότητα των άλλων προσεγγίσεων μπορεί να αλλοιωθεί από το σύνθετο βιομηχανικό περιβάλλον και από άλλα κινούμενα αντικείμενα.

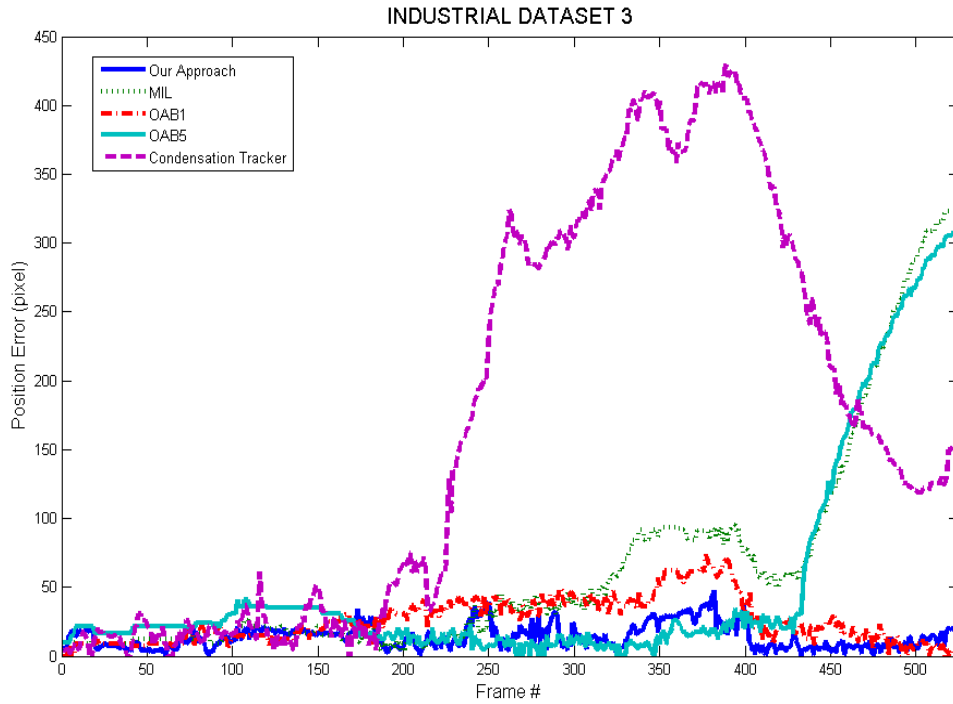
Ο *Color Tracker* (πέμπτη γραμμή) παρασύρεται μακριά από το στόχο, δεδομένου ότι επηρεάζεται από αντικείμενα του ίδιου χρώματος μέσα στη σκηνή. Ομοίως, οι τεχνικές παρακολούθησης *OABI* [4] και *MIL* [73] (δεύτερη και τέταρτη γραμμή) δεν μπορούν να προσαρμοστούν σε σημαντικές αλλαγές της εμφάνισης του αντικειμένου και αποπροσανατολίζονται σε σύνθετα περιβάλλοντα. Τέλος, ενώ η τεχνική *OAB5* [4] (τρίτη γραμμή) έχει καλύτερες επιδόσεις από τις δύο προηγούμενες, δεν παραμένει απολύτως αφοσιωμένη στο στόχο κατά την διάρκεια της ακολουθίας.



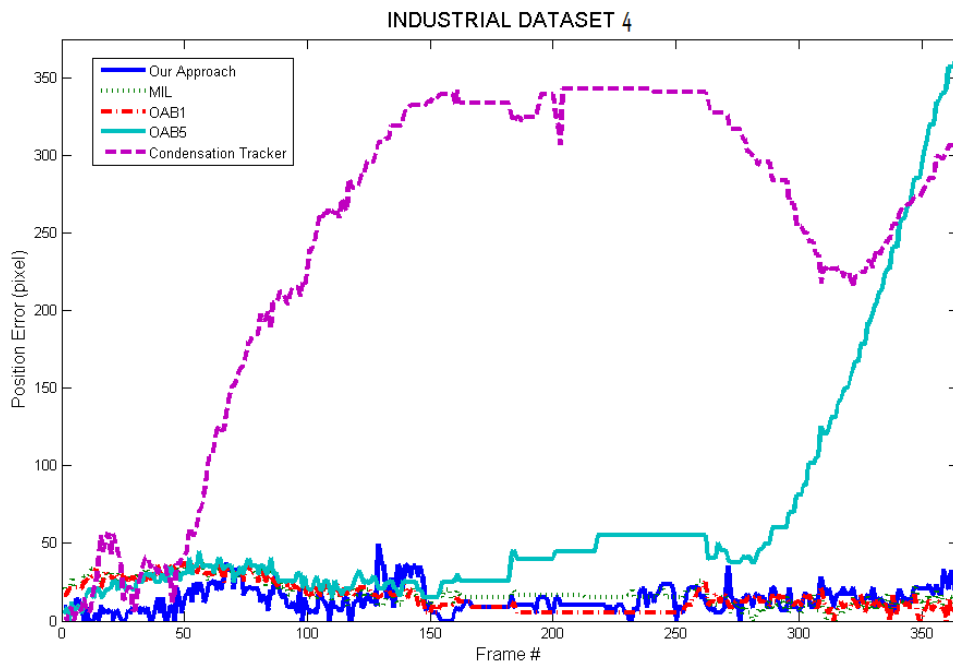
(α) Εικονοσειρά 1



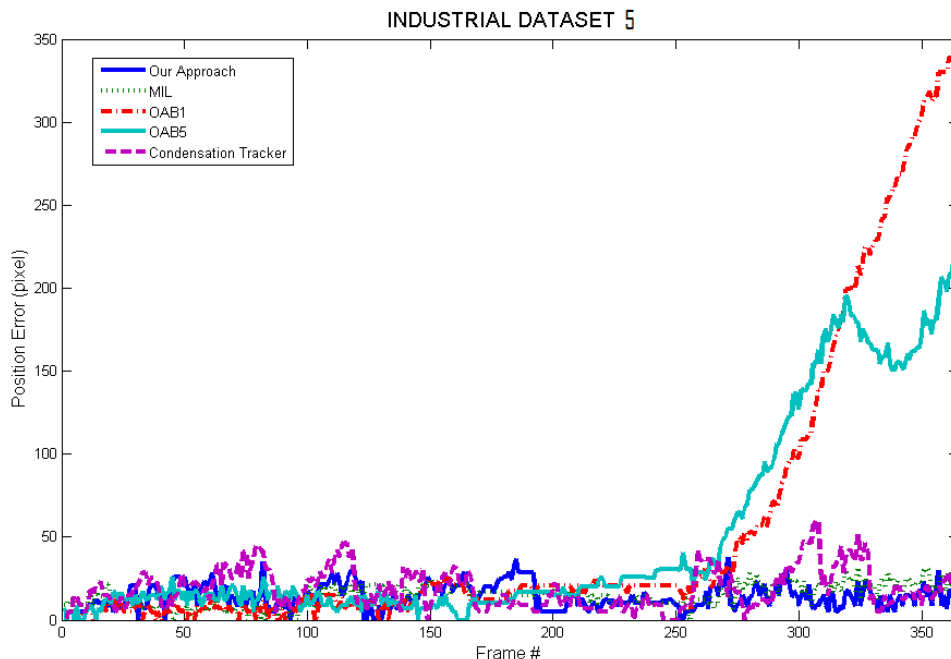
(β) Εικονοσειρά 2



(γ) Εικονοσειρά 3



(δ) Εικονοσειρά 4



(ε) Εικονοσειρά 4

Σχήμα 28. Ποσοτικά αποτελέσματα. Η προτεινόμενη προσέγγιση παρακολούθησης έχει σταθερή απόδοση σε όλες τις ακολουθίες. Οι άλλες προσεγγίσεις παρασύρονται μακριά από τον στόχο, διότι επηρεάζονται από το σύνθετο βιομηχανικό περιβάλλον.

Επιπλέον, έγινε μία ποσοτική σύγκριση της προτεινόμενης προσέγγισης με τις προαναφερθείσες τεχνικές παρακολούθησης. Για όλες τις ακολουθίες δημιουργήθηκε *ground truth* για το αντικείμενο υπό μελέτη. Στο Σχήμα 28 παρουσιάζεται το απόλυτο σφάλμα (σε *pixels*). Όπως μπορεί να παρατηρηθεί, η προτεινόμενη μέθοδος έχει πιο σταθερή απόδοση από τις άλλες μεθόδους εντοπισμού. Αυτό μπορεί επίσης να παρατηρηθεί στον Πίνακα 5, όπου εμφανίζεται το μέσο σφάλμα θέσης για όλα τα εμπλεκόμενα συστήματα παρακολούθησης.

Πίνακας 5. Αριθμητική παρουσίαση του μέσου σφάλματος (σε pixel) για τις 5 διαφορετικές εικονοσειρές. Το προτεινόμενο σύστημα παρακολούθησης συγκρίνεται με τις υπόλοιπες μεθόδους.

Μέσο σφάλμα (σε pixel)	Εικονοσειρά 1	Εικονοσειρά 2	Εικονοσειρά 3	Εικονοσειρά 4	Εικονοσειρά 5
<i>OABI</i>	81,762	13,642	51,154	15,542	56,711
<i>OAB5</i>	8,9843	117,27	25,758	67,504	49,297
<i>MIL</i>	7,1575	16,509	64,905	17,959	14,828
<i>Color Tracker</i>	19,152	20,082	168,84	245,798	18,824
Προτεινόμενη Μέθοδος	5,9877	11,233	13,862	13,338	13,921

Υπολογιστικό κόστος. Ο μέσος όρος χρόνος εκτέλεσης της μη-βελτιστοποιημένης εφαρμογής είναι ~ 10fps. Πιο συγκεκριμένα, η πολυπλοκότητα του φίλτρου σωματιδίων εξαρτάται από τον αριθμό των σωματιδίων που χρησιμοποιούνται. Ωστόσο, έχει επιλεγθεί ένας μικρός αριθμός σωματιδίων (δηλαδή 30), έτσι η πολυπλοκότητα παραμένει πολύ χαμηλή και η λειτουργία γίνεται σε πραγματικό χρόνο (ή σχεδόν πραγματικό χρόνο). Θα πρέπει να σημειωθεί ότι υπάρχει αρκετή υπολογιστική πολυπλοκότητα για τον υπολογισμό της *GPLVM* αναπαράστασης (βλ. Κεφάλαιο 6.1.3.2). Ωστόσο, αυτό το ζήτημα μπορεί να επιλυθεί με την κατάλληλη βελτιστοποίηση κώδικα.

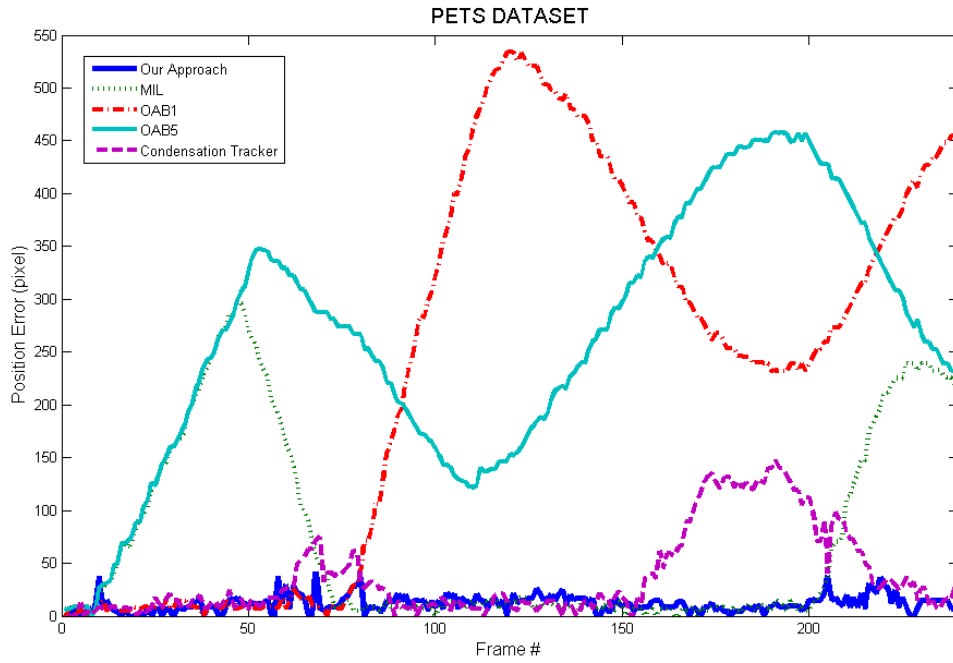
6.2.2 Πειραματικά Αποτελέσματα για τον Μηχανισμό Έλεγχου

Δειγμάτων Εκπαίδευσης

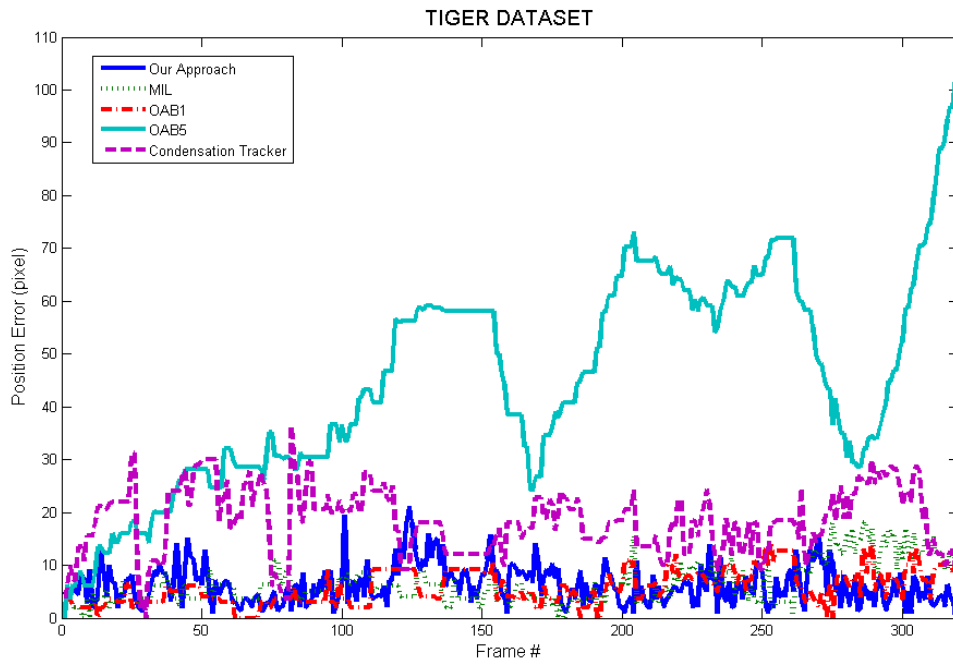
Ο μηχανισμός έλεγχου δειγμάτων εκπαίδευσης πραγματικού χρόνου ενσωματώθηκε στο προτεινόμενο σύστημα παρακολούθησης (βλ. Κεφάλαιο 6.1). Θα πρέπει να σημειωθεί ότι στα συγκεκριμένα πειράματα δεν χρησιμοποιήθηκε το μοντέλο περιγραφής αντικειμένου βάσει παραγωγικού *prior* (βλ. Κεφάλαιο 6.1.3). Συγκεκριμένα, ο στόχος των παρακάτω πειραμάτων είναι να εξεταστεί η αξιοπιστία του προτεινόμενου μηχανισμού εκπαίδευσης, για εικονοσειρές όπου το αντικείμενο που παρακολουθείται επικαλύπτεται από άλλα στατικά ή κινούμενα αντικείμενα. Σε αυτή την προσέγγιση χρησιμοποιήθηκε ο ταξινομητής που προτάθηκε από τον *Grabner* [4], ο οποίος εκπαιδευόταν σε πραγματικό χρόνο με δείγματα που κρίθηκαν ως καταλληλά από τον μηχανισμό.

Για τα πειράματα χρησιμοποιήθηκαν οι εικονοσειρές *PETS*⁵, *TIGER* και *MUG*. Ειδικότερα, η εικονοσειρά *PETS* είναι αρκετά γνωστή στον τομέα της όρασης υπολογιστών και περιέχει ένα πολυσύχναστο χώρο, όπου το αντικείμενο που παρακολουθείται, επικαλύπτεται από άλλα κινούμενα άτομα και εμπόδια. Αυτή η εικονοσειρά έχει ανάλυση **768 × 578 pixel**. Η εικονοσειρά *TIGER* έχει ανάλυση **320 × 240 pixel** και περιέχει ένα κινούμενο αντικείμενο, του οποίου η εμφάνιση συχνά μεταλλάσσεται. Τέλος, η εικονοσειρά *MUG* έχει ανάλυση **704 × 576 pixel** και περιέχει ένα σενάριο με κινούμενη κάμερα. Αυτή η εικονοσειρά μαγνητοσκοπήθηκε με την βοήθεια μίας *AXIS 213 PTZ* κάμερας. Σε όλες τις παρακάτω εικονοσειρές, η αρχική εκπαίδευση της ροής αντικειμένου (δηλαδή, σε μη - πραγματικό χρόνο) γίνεται με ελάχιστα δείγματα, $|X^+| \approx 10$ θετικά, $|X^-| \approx 40$ αρνητικά.

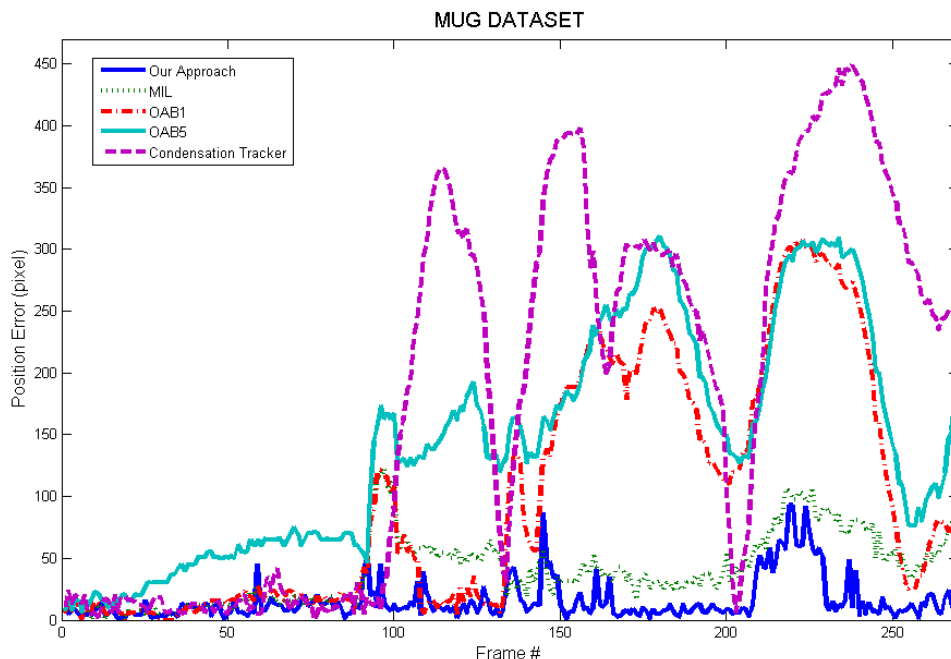
⁵ <http://www.cvg.rdg.ac.uk/slides/pets.html>



(α) Εικονοσειρά 1



(β) Εικονοσειρά 2



(γ) Εικονοσειρά 3

Σχήμα 29. Ποσοτικά αποτελέσματα της προτεινόμενης προσέγγισης παρακολούθησης σε σχέση με άλλες μεθόδους. Η προτεινόμενη μέθοδος λειτουργεί παράλληλα με τον μηχανισμό έλεγχου δειγμάτων εκπαίδευσης με αρκετά μεγάλη σταθερότητα.

Για όλες τις ακολουθίες δημιουργήθηκε *ground truth* για το αντικείμενο υπό μελέτη. Στο Σχήμα 29 παρουσιάζεται το απόλυτο σφάλμα (σε *pixels*). Όπως μπορεί να παρατηρηθεί, η προτεινόμενη μέθοδος έχει πιο σταθερή απόδοση από τις άλλες μεθόδους παρακολούθησης πραγματικού χρόνου. Αυτό μπορεί επίσης να παρατηρηθεί στον Πίνακα 6, όπου εμφανίζεται το μέσο σφάλμα θέσης για όλα τα εμπλεκόμενα συστήματα παρακολούθησης. Αυτά τα αποτελέσματα αποδεικνύουν ότι ο προτεινόμενος μηχανισμός εκπαίδευσης μπορεί με ευκολία να ενσωματωθεί σε συστήματα παρακολούθησης.

Πίνακας 6. Αριθμητική παρουσίαση του μέσου σφάλματος (σε *pixel*) για τις **3** διαφορετικές εικονοσειρές. Η προτεινόμενη μέθοδος λειτουργεί παράλληλα με τον μηχανισμό έλεγχου δειγμάτων εκπαίδευσης και συγκρίνεται με τις υπόλοιπες μεθόδους.

Μέσο σφάλμα (σε <i>pixel</i>)	Εικονοσειρά 1 PETS	Εικονοσειρά 2 TIGER	Εικονοσειρά 3 MUG
<i>OAB1</i>	236,603	6,119	100,850
<i>OAB5</i>	263,301	45,861	143,882
<i>MIL</i>	74,148	6,595	40,321
<i>Color Tracker</i>	37,939	18,150	183,927
Προτεινόμενη Μέθοδος	11,536	5,763	15,353

Η σελίδα αυτή είναι σκόπιμα λευκή.

7

Μέθοδος Συσχέτισης Δεδομένων

Οι μέθοδοι συσχέτισης δεδομένων (*data association*) παίζουν σημαντικό ρόλο στα συστήματα παρακολούθησης αντικειμένων. Συγκεκριμένα, όταν παρακολουθούνται παραπάνω από ένα αντικείμενα στην σκηνή χρειάζεται ένας τρόπος, ο οποίος να μπορεί να συσχετίζει τα δεδομένα που περιγράφουν ξεχωριστά αντικείμενα (π.χ. ανθρώπους, αυτοκίνητα, μηχανήματα). Επίσης, σε αρκετές εφαρμογές χρειάζεται να κρατείται ένα ιστορικό, το οποίο περιέχει τα αντικείμενα που παρακολουθήθηκαν κατά την διάρκεια της λειτουργίας του συστήματος. Σε αυτή την περίπτωση ο στόχος αυτών των εφαρμογών παρακολούθησης συμπίπτει με τις πολυμεσικές τεχνικές ανάκτησης και οργάνωσης δεδομένων (*content-based retrieval*).

Συγκεκριμένα, αυτές οι πολυμεσικές τεχνικές (βλ. Κεφάλαιο 4.2 για μία παρουσίαση της τρέχουσας βιβλιογραφίας επάνω στο θέμα) χρησιμοποιούν διάφορους αλγορίθμους που ως στόχο έχουν την οργάνωση μιας βάσης δεδομένων που αποτελείται από πολυμεσικά δεδομένα (π.χ. εικόνες) σε ομάδες που μοιράζονται κοινά χαρακτηριστικά. Συγκεκριμένα, σε ένα σύστημα το οποίο αποτελείται από ένα μεγάλο αριθμό πολυμεσικών δεδομένων, η εφαρμογή αλγορίθμων ομαδοποίησης

παίζει ιδιαίτερα σημαντικό ρόλο, ειδικά στην περίπτωση όπου ο χρήστης επιθυμεί να αναζητήσει γρήγορα και αξιόπιστα τα δεδομένα που τον αφορούν. Η περιγραφή των δεδομένων μπορεί να γίνει είτε με την βοήθεια των πρωτογενών χαρακτηριστικών (*low level descriptors*), όπως χρώμα, σχήμα, μέγεθος αντικειμένων κ.τ.λ., είτε με την βοήθεια των σημασιολογικών χαρακτηριστικών (*semantics*).

Η προτεινόμενη μέθοδος συσχέτισης δεδομένων μπορεί να οργανώνει δυναμικά (δηλαδή, σε πραγματικό χρόνο) σε ομάδες διάφορες εικόνες που προέρχονται από ένα σύστημα παρακολούθησης. Συγκεκριμένα, ο προτεινόμενος αλγόριθμος δεικτοδότησης εικόνων έχει την ευελιξία να είναι ανεξάρτητος από το περιεχόμενο της βάσης δεδομένων και τη δομή, ένα γεγονός το οποίο τον κάνει κατάλληλο για την συγκεκριμένη εφαρμογή. Το προτεινόμενο σχήμα δουλεύει ως ακολούθως. Αρχικά δημιουργείται μια τοπολογία κύβου η οποία αποτελείται από κελιά. Η δομή αυτή χρησιμοποιείται για την αντιστοίχιση του περιεχομένου της εικόνας και/ή του τμήματος βίντεο. Η δομή κελιού χρησιμοποιείται για τον ορισμό συστοιχιών (*clusters*) που μοιράζονται κοινό οπτικό περιεχόμενο. Στην συγκεκριμένη εφαρμογή, το οπτικό περιεχόμενο περιγράφεται χρησιμοποιώντας το πρότυπο *MPEG-7*.

Θα πρέπει να τονιστεί ότι υπάρχουν πολλοί *MPEG-7* περιγραφείς για την αναπαράσταση του πολυμεσικού περιεχομένου (χρώμα, υφή, σχήμα, κίνηση). Η αποδοτικότητα του κάθε περιγραφέα ποικίλλει ανάλογα με το σύνολο των δεδομένων. Όμως, η υλοποίηση ενός μεγάλου αριθμού *MPEG-7* περιγραφέων είναι μη αποδοτική όσον αφορά την υπολογιστική πολυπλοκότητα, όπως και για την ανάγκη λειτουργίας του συστήματος σε πραγματικό χρόνο. Για τον λόγο αυτό, πραγματοποιήθηκε εκτεταμένη μελέτη της βιβλιογραφίας προκειμένου να επιλεγεί ένας κατάλληλος συνδυασμός περιγραφέων για τη συγκεκριμένη προσέγγιση. Ο *CLD* είναι ιδανικός για την ανάκτηση εικόνων χάρη στο μέγεθός του και στο γεγονός ότι

αποδίδει καλά σε διαφορετικά σύνολα δεδομένων. Επιπροσθέτως, αποτελεί έναν από τους πιο έξυπνους περιγραφείς διότι όχι μόνο περιγράφει χρώματα, αλλά εντοπίζει και τη θέση τους στην εικόνα [74]. Οι *DCD* και *HTD* είναι ιδανικοί για την περιγραφή απλών και περίπλοκων σχημάτων και υφών σε μια εικόνα. Σύμφωνα με το [75] οι καλύτεροι περιγραφείς για χρήση συνδυασμών είναι οι *CLD*, *EHD*, *RSD*, *DCD* και *HTD*. Μιας και είναι απαραίτητη η παραγωγή σε πραγματικό χρόνο των *MPEG-7* περιγραφέων, ο τελευταίος περιγραφέας δεν υλοποιήθηκε λόγω της μεγάλης του υπολογιστικής πολυπλοκότητας. Επιπρόσθετα, σύμφωνα με το [75], «λόγω των φτωχών διακυμάνσεων και των λίγων στοιχείων σε αυτόν τον περιγραφέα (οκτώ *bins*) είναι δύσκολο να χρησιμοποιηθεί στις περισσότερες εφαρμογές, μιας και τα αποτελέσματα μπορεί να είναι διφορούμενα».

Η προτεινόμενη μέθοδος συσχέτισης δεδομένων βασίζεται στο *MPEG-7* και εκμεταλλεύεται διαλειτουργικότητα, μεταφερσιμότητα και προσβασιμότητα της αναπαραστώμενης πληροφορίας [52]. Επίσης, η εκμετάλλευση της προσέγγισης συστοιχιών (*clustering*) χρησιμοποιείται για τη βελτίωση του χρόνου απόκρισης ανακτήσεων λόγω του ότι ο χώρος εύρεσης σχετικού περιεχομένου ελαττώνεται σημαντικά. Αυτό το γεγονός θεωρείται ιδιαίτερα σημαντικό θέμα για εφαρμογές ανάκτησης βάσει περιεχομένου. Άλλωστε, όπως αναφέρθηκε στο Κεφάλαιο 4.2, η σειριακή αναζήτηση μπορεί να αποτελεί μια εξαιρετικά μη αποδοτική προσέγγιση, ιδιαιτέρως όταν πραγματεύεται με μεγάλες και ετερογενείς κατανεμημένες βάσεις δεδομένων.

Η προτεινόμενη μέθοδος έχει τα ακόλουθα πλεονεκτήματα. Πρώτον, έχει επεκτάσιμη συμπεριφορά όσον αφορά στην ενημέρωση περιεχομένου. Ο προτεινόμενος αλγόριθμος έχει γραμμική υπολογιστική πολυπλοκότητα $O(n)$ για τη διαγραφή ή

εισαγωγή εικόνων ή τμημάτων βίντεο από τις συστοιχίες αντιστοιχώς. Δεύτερον, σε σύγκριση με μερικούς από τους προηγουμένως αναφερθέντες αλγόριθμους, επιτρέπει τη δυναμική αναμόρφωση της δομής δείκτη. Συνεπώς, οι συστοιχίες δεν υπολογίζονται από την αρχή κατά την εισαγωγή δεδομένων.

7.1 Κλιμακούμενη Δεικτοδότηση Πολυμεσικού

Υλικού

7.1.1 Τοπολογία Κελιού

Αρχικά η τοπολογία κελιού δημιουργείται για την αποδοτική αναπαράσταση του περιεχομένου των (M) οπτικών περιγραφέων που περιγράφηκαν πριν. Η τοπολογία

αυτή δημιουργείται με έναν τρόπο που επιτρέπει τη δυναμική ενημέρωση δείκτη. Η κατασκευασμένη τοπολογία κελιού ανήκει στον M -διάστατο χώρο, στον οποίο κάθε

διάσταση αναφέρεται σε έναν περιγραφέα. Έστω ότι f_1^k είναι ο l -οστός περιγραφέας

του k -οστού πολυμεσικού περιεχομένου. Στην περίπτωση αυτή, ο πρώτος περιγραφέας

f_1^k αναφέρεται στον CLD , ο δεύτερος στον EHD και ο τρίτος στον RSD (βλ.

παράγραφο 7.1.2).

Έτσι στην περίπτωση αυτή το M είναι ίσο με 3. Συνεπώς, η τοπολογία κελιού ανήκει

στον 3 -διάστατο χώρο. Έστω ότι μ_i είναι η παράμετρος που αναφέρεται στο πλήθος

των διαχωρισμών του i -οστού περιγραφέα. Για παράδειγμα, ο EHD περιγραφέας είναι ένα 80 -διάστατο διάνυσμα χαρακτηριστικών. Οι πιθανές τιμές για όλα τα 80 στοιχεία του περιγραφέα αυτού κυμαίνονται στο διάστημα $[0\ 7]$ όπως φαίνεται και στο [56]. Συνεπώς, επιλέγοντας $\mu_i = 8$ για τον EHD περιγραφέα, 8 διαφορετικά κελιά για τη κβαντοποίηση όλων των 80 bins του EHD μπορούν να δημιουργηθούν. Το πρώτο κελί περιλαμβάνει τιμές στο διάστημα $[0\ 1)$, το δεύτερο στο διάστημα $[1\ 2)$ κ.τ.λ. Για έναν δεδομένο αριθμό διαχωρισμού μ_i του i -οστού περιγραφέα, το μέγεθος κελιού της i -οστής διάστασης δίνεται από την ακόλουθη εξίσωση,

$$g_i = \frac{MaxV_i - MinV_i + 1}{\mu_i} \quad (7.1)$$

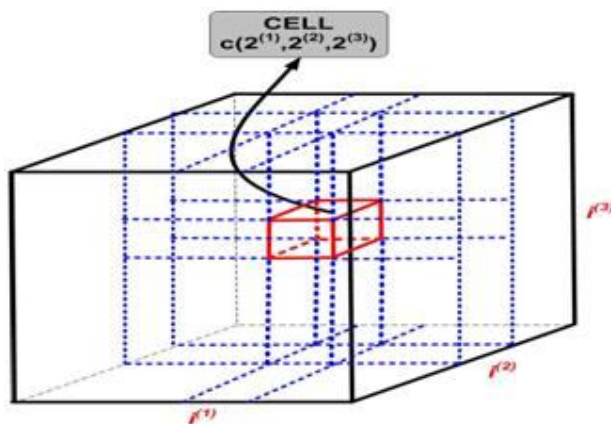
Στην παραπάνω εξίσωση, το $MaxV_i$ αναφέρεται στη μέγιστη δυνατή τιμή για όλα τα στοιχεία του i -οστού περιγραφέα, ενώ το $MinV_i$ αναφέρεται στην αντίστοιχη ελάχιστη τιμή.

Η Εξίσωση (7.1) ορίζει το μέγεθος ενός κελιού g_i σε μια τοπολογία μονοδιάστατου χώρου που αντιστοιχίζεται σε έναν περιγραφέα πολυμεσικού περιεχομένου. Υποθέτοντας ότι M περιγραφείς είναι διαθέσιμοι, μια δομή κελιού δημιουργείται σε έναν M -διάστατο χώρο. Στην περίπτωση, μια τοπολογία 3 -διάστατου υπέρ-ορθογωνίου δημιουργείται, στην οποία ο κάθε άξονας του υπέρ-ορθογωνίου

κβαντοποιείται σε διαστήματα μεγέθους g_i με $i = 1, 2, \dots, M$. Το συνολικό πλήθος των κελιών που δημιουργούνται δίνεται από,

$$K = \prod_{i=1}^M \mu_i \quad (7.2)$$

Ένα κελί στον M -διάστατο χώρο αναπαρίσταται μοναδικά από ένα σύνολο M δεικτών, ο καθένας εκ των οποίων δείχνει έναν δείκτη διαχωρισμού για τον περιγραφέα. Έστω $i^{(q)}$ με $q = 1, 2, \dots, M$ με το M να είναι το πλήθος των διαφορετικών δεικτών. Ο δείκτης $i^{(q)}$ παίρνει τιμές που κυμαίνονται στο $0, \dots, \mu_i^{-1}$. Στο συγκεκριμένο σενάριο, στο οποίο χρησιμοποιούνται 3 περιγραφείς ($M = 3$) και υποθέτωντας ότι τα επιλεγμένα τμήματα είναι 8 και 16 αντιστοίχως, η παράσταση $14^{(2)}$ αφορά στη 14-ιστό τμήμα του δεύτερου περιγραφέα, ενώ το $4^{(3)}$ στο 4ο τμήμα του 3ου περιγραφέα. Χρησιμοποιώντας τους δείκτες $i^{(q)}$, η πρόσβαση σε ένα κελί στον M -διάστατο χώρο γίνεται ως εξής, $c(i^{(1)}, i^{(2)}, \dots, i^{(M)})$. Για παράδειγμα, το κελί $c(2^{(1)}, 3^{(2)}, 4^{(3)})$ αναφέρεται στο 2ο, 3ο και 4ο τμήμα του 1ου, 2ου και 3ου περιγραφέα αντιστοίχως.



Σχήμα 30. Παράδειγμα χρησιμοποιούμενης τοπολογίας κελιού

Το Σχήμα 30 παρουσιάζει ένα παράδειγμα της προτεινόμενης τοπολογίας κελιού στην ειδική περίπτωση που έχουμε 3 περιγραφείς, και συνεπώς η αναπαράσταση του υπέρ-ορθογωνίου είναι στον **3-διάστατο** χώρο ($M = 3$) των *CLD*, *EHD* και *RSD* περιγραφέων.

7.1.2 Χαρτογράφηση Περιεχομένου στην Τοπολογία Κελιών

Χρησιμοποιώντας την συγκεκριμένη δομή το περιεχόμενο του K πολυμεσικού δεδομένου χαρτογραφείται στο αντίστοιχο κελί της τοπολογίας. Έστω ότι το k περιεχόμενο αντιπροσωπεύεται από M πολυδιάστατους περιγραφείς f_i^k . Το $f_{q,j}^k$ στοιχείο ορίζεται ως το j στοιχείο του q περιγραφέα. Εάν m_i ορίζεται ως το μήκος του διανύσματος f_i^k , τότε $j = 1, 2, \dots, m_i$. Συγκεκριμένα, το μήκος m_i εξαρτάται μόνο από τον τύπο του περιγραφέα και όχι από τα πολυμεσικά δεδομένα που περιέχει. Για

κάθε στοιχείο $f_{q,j}^k$ η κατάλληλη χαρτογράφηση (*index*) θα μπορούσε να επιτευχθεί χρησιμοποιώντας την παρακάτω εξίσωση,

$$\zeta(q) = \left\lfloor \frac{f_{q,j}^k}{g_q} \right\rfloor \quad (7.3)$$

Συγκεκριμένα, g_q είναι το μήκος του q περιγραφέα (βλ. Εξίσωση (7.1)). Η εξίσωση

7.3 υπολογίζεται για όλους τους δυνατούς συνδυασμούς των στοιχείων $f_{q,j}^k$ για κάθε

διαθέσιμο περιγραφέα έτσι ώστε να οριστεί το κατάλληλο κελί στο οποίο μπορεί να χαρτογραφηθεί η συγκεκριμένη εικόνα. Όπως περιγράφεται παραπάνω είναι ξεκάθαρο ότι για ένα συγκεκριμένο συνδυασμό μοναχά ένα κελί μπορεί να χαρτογραφηθεί. Επίσης χρησιμοποιείται ένας δείκτης για να μετρηθεί το ποσό των χαρτογραφημένων δεδομένων ανά κάθε κελί. Έτσι εάν σε ένα συγκεκριμένο κελί χαρτογραφηθούν ένα ή περισσότερα στοιχεία ο δείκτης αυξάνεται αντίστοιχα κατά ένα. Παρακάτω παρουσιάζεται ο αλγόριθμος ο οποίος εκτελεί την χαρτογράφηση των πολυμεσικών δεδομένων στα κελιά της τοπολογίας.

Πίνακας 7. Αλγόριθμος για την χαρτογράφηση κελιών.

1. Χαρτογράφηση στοιχείων (*CLD Array[]*, *EHD Array[]*, *RSD Array[]*)
2. Αρχικοποίηση του *Loop 1*, για $\forall f_{1,j}^k, f_{2,z}^k, z_{1,i}^k \in CLD[], EHD[]$ και *SHD[]*

αντίστοιχα,

a. Αρχικοποίηση του *Loop 2*

i. Για \forall κελί $c(i^{(1)}, i^{(2)}, i^{(3)}) \in$ στην τοπολογία έλεγχος,

1. Εάν $f_{1,j}^k, f_{2,r}^k, f_{3,i}^k \in c(i^{(1)}, i^{(2)}, i^{(3)})$

2. Τότε, αύξηση του περιεχομένου $c(i^{(1)}, i^{(2)}, i^{(3)})$ κατά

ένα

b. Τέλος του *Loop 2*

3. Τέλος του *Loop 1*

Θα πρέπει να τονιστεί ότι υπάρχουν δυο διαφορετικοί τρόποι για να χαρτογραφηθεί ένας αριθμός πολυμεσικών δεδομένων σε ένα κελί. Μία εικόνα χαρτογραφείτε σε ένα κελί μοναχά εάν υπάρχει ο συγκεκριμένος συνδυασμός στοιχείων ο οποίος αντιστοιχίζεται σε ένα συγκεκριμένο κελί, γεγονός που δημιουργεί μια προσέγγιση με ιδιαίτερη πολυπλοκότητα. Συγκεκριμένα στην περίπτωση πιθανού θορύβου η χαρτογράφηση των στοιχείων μπορεί να αλλοιωθεί. Για αυτό το λόγο μπορούμε να χαρτογραφήσουμε ένα στοιχείο σε ένα κελί εάν ο αριθμός των στοιχείων σε αυτό το κελί είναι μεγαλύτερος από ένα κατώφλι T . Έτσι εάν ο αριθμός των στοιχείων είναι μικρότερος από το κατώφλι T τότε, το συγκεκριμένο κελί ορίζεται ίσο με το λογικό μηδέν. Αντίθετα στην περίπτωση όπου τα κελιά με περιεχόμενο μεγαλύτερο από το κατώφλι T αυτά τα κελιά ορίζονται ίσα με το λογικό ένα. Έτσι για κάθε k εικόνα, χρησιμοποιώντας αυτήν την προσέγγιση δημιουργείται ένα διάνυσμα το οποίο αποτελείται από λογικές τιμές. Από εδώ και στο εξής αυτό το διάνυσμα λογικών τιμών θα περιγράφεται ως s^k . Θα πρέπει να σημειωθεί ότι η επιλογή του T εξαρτάται από την φύση της πολυμεσικής βάσης δεδομένων (ομογενής ή μη-ομογενής) όπως και

από την μορφή της υποβαλλόμενης (*query*) εικόνας. Στα πειράματα που διεξάχθηκαν για το T χρησιμοποιήθηκαν τιμές από 4 έως 400.

Η διαδικασία που περιγράφεται παραπάνω εκτελείται για κάθε εικόνα στην βάση δεδομένων. Το συγκεκριμένο κομμάτι του αλγορίθμου ολοκληρώνεται με την πρόσθεση **16-bit** στοιχείου ως *επικεφαλίδα*. Αυτή επικεφαλίδα βοηθά στο να αναγνωριστεί το όνομα της συγκεκριμένης εικόνας στα επόμενα επίπεδα του συστήματος. Εφόσον το λογικό διάνυσμα \mathfrak{s}^k θα αποθηκευτεί στην βάση δεδομένων του συστήματος, η **16-bit** επικεφαλίδα που θα προστεθεί, θα χρησιμοποιηθεί για την αναχώρηση του αρχείου. Παρακάτω περιγράφεται ο αντίστοιχος αλγόριθμος.

Πίνακας 8. Αλγόριθμος για την δημιουργία λογικών διανυσμάτων

1. Δημιουργία λογικών διανυσμάτων (εισαγωγή του *cell structure* $c(i^{(1)}, i^{(2)}, i^{(3)})$, με κατώφλι T , όνομα της εικόνας k)
2. Δημιουργία μονοδιάστατου λογικού διανύσματος με μέγεθος ίσο με τον αριθμό των κελιών ***NewSignatureVector = NULL***
 - a. Αρχικοποίηση του *Loop 2*, για \forall κελί $c(i^{(1)}, i^{(2)}, i^{(3)}) \in$ θέση j ,
 - i. Εάν, το περιεχόμενο του κελιού $c(i^{(1)}, i^{(2)}, i^{(3)}) > T$ τότε ορισμός της θέσης j στο ***NewSignatureVector*** ως ***true***
 - ii. Αλλιώς, ορισμός της θέσης j στο ***NewSignatureVector*** ως ***false***

b. Τέλος του *Loop 2*

4. Τέλος του *Loop 1* και μετατροπή του ονόματος της εικόνας k σε **16-bit**

7.2 Δυναμική Ομαδοποίηση Πολυμεσικών Στοιχείων

Αυτό το κομμάτι της εργασίας περιγράφει το κομμάτι του αλγορίθμου το οποίο είναι υπεύθυνο για την δημιουργία ομάδων οι οποίες μοιράζονται κοινά πολυμεσικά δεδομένα. Όπως αναφέραμε προηγουμένως, η μέθοδος εύρεσης δεδομένων σειριακά είναι αντιπαραγωγική. Επομένως δημιουργήθηκε ένας αλγόριθμος οργάνωσης περιεχομένου με στόχο να μειωθεί ο χρόνος και η πολυπλοκότητα της ανάκτησης δεδομένων (*clustering*) Έτσι κατά την είσοδο μίας υποβαλλόμενης εικόνας το σύστημα θα ψάχνει για κοινά δεδομένα ανάμεσα σε ορισμένες ομάδες (*content based clusters*) οι οποίες μοιράζονται κοινές πληροφορίες. Οι ομάδες αυτές δημιουργούνται βάση του λογικού διανύσματος \mathbf{s}^k , το οποίο περιγράφεται στην προηγούμενη ενότητα.

Συγκεκριμένα αυτές ομάδες διαμορφώνονται ως εξής. Ένα σύνολο ομάδων G_i , ορίζεται αρχικά εάν τουλάχιστον δύο εικόνες μοιράζονται ένα κοινό κελί c στον M -διάστατο χώρο που αναλύθηκε προηγουμένως στην ενότητα 7.2.1. Έστω λοιπόν S_i^k το i λογικό στοιχείο του διανύσματος \mathbf{s}^k , όπου $\mathbf{s}^k = [S_1^k, S_2^k, \dots, S_M^k]$. Έστω I το σύνολο όλων των διαθέσιμων δεικτών όλων των διαθέσιμων εικόνων. Τότε η ομάδα G_i ορίζεται ως

$$G_i = \{k \in I: S_i^k = 1\} \quad (7.4)$$

Η Εξίσωση (7.4) δείχνει ότι μια αρχική ομάδα G_i κατασκευάζεται από εκείνους τους δείκτες των οποίων όλα στοιχεία σε μία συγκεκριμένη θέση του διανύσματος s^k είναι ίσα με το λογικό ένα. Δηλαδή παρουσιάζουν τουλάχιστον ένα κοινό κελί c στην προαναφερθείσα τοπολογία. Αυτό μπορεί να εκφραστεί ως εξής,

$$S_i^k \text{ AND } S_i^m = 1 \forall k, m \in G_i \quad (7.5)$$

Έχοντας κατασκευάσει την ομάδα G_i , το διάνυσμα s^{G_i} της ομάδας G_i μπορεί επίσης να οριστεί. Αυτό το διάνυσμα θα έχει ακριβώς τις ίδιες διαστάσεις με την ομάδα s^k ,

$$s^{G_i} = s^u \text{ AND } s^v \dots \text{ AND } s^r \quad \forall u, v, \dots, r \in G_i \quad (7.6)$$

Στην Εξίσωση (7.6) παραπάνω u, v, \dots, r είναι οι δείκτες της ομάδας G_i . Επιπρόσθετα οι αρχικές ομάδες G_i , μπορούν να ενωθούν περαιτέρω εάν παρουσιάζουν παρακείμενες θέσεις στην παραπάνω τοπολογία. Η ομοιότητα βάσει χώρου εκφράζεται ως εξής. Έστω ότι H το σύνολο των αρχικών ομάδων G_i . Τότε οι ομάδες ενώνονται μαζί έτσι ώστε,

$$G_F = \{G_i \in H: (S_i^{G_i} \text{ AND } S_{i+1}^{G_{i+1}}) \text{ OR } (S_{i+1}^{G_i} \text{ AND } S_{i+1}^{G_{i+1}}) = 1\} \quad (7.7)$$

Πίνακας 9. Αλγόριθμος για την δημιουργία αρχικών ομάδων

1. Δημιουργία λογικών διανυσμάτων (εισαγωγή του *cell structure*)
-

$c(i^{(1)}, i^{(2)}, i^{(3)})$, με κατώφλι T , όνομα της εικόνας k

2. Αρχικοποίηση του *Loop 1* για την αναγνώριση κελιών.
 - a. Για s^k , αρχικοποίηση του *Loop 2*. Για την θέση j του s^k ,
 $s^1 \text{ AND } s^2 \dots s^k = \text{true}$ τότε,
 - i. Τοποθέτηση των δεδομένων $1, \dots, k$ στην ίδια κλάση G_i και δημιουργία διανύσματος.
 - b. Τέλος του *Loop 2*
3. Τέλος του *Loop 1*
4. Εάν, ο αριθμός περιεχομένων της κλάσης $n > 1$ τότε, αρχικοποίηση του *Loop 3* για τον εντοπισμό παρακείμενων περιοχών
 - i. Για την ομάδα i και την θέση j του διανύσματος G_i τότε,
 - ii. Ένωση των G_i σε μία νέα ομάδα G_F
5. Τέλος του *Loop 3*

Τα λογικά διανύσματα αποθηκεύονται τοπικά και είναι δυναμικά ανανεώσιμα κάθε φορά που νέα δεδομένα εισέρχονται ή αφαιρούνται. Η διαμόρφωση νέων ομάδων μπορεί να δημιουργηθεί με την εφαρμογή του λογικού **AND**. Ο αλγόριθμος που χρησιμοποιείται για την διαμόρφωση των αρχικών ομάδων παρουσιάζεται παρακάτω.

7.2.1 Μηχανισμός Δυναμικής Ανανέωσης Περιεχομένου

Η εισαγωγή νέων μεταδεδομένων στο σύστημα θα ξεκινήσει τον μηχανισμό ανανέωσης έτσι ώστε να οργανώσει σε ομάδες τα επιπρόσθετα πολυμεσικά δεδομένα.

Ο αλγόριθμος ο οποίος χρησιμοποιείται για την εισαγωγή των νέων πολυμεσικών δεδομένων στις ομάδες που έχουν ήδη δημιουργηθεί G_F παρουσιάζεται σε αυτό το κεφάλαιο. Συγκεκριμένα η νέα εικόνα ανήκει σε μια ομάδα G_F εάν,

$$s^{up} \text{ AND } s^{GF} = 1 \quad (7.8)$$

σε τουλάχιστον μία θέση j του λογικού διανύσματος. Στην περίπτωση όπου δεν βρεθούν κοινές θέσεις αυτή η συγκεκριμένη εικόνα εισέρχεται σε μια σειρά αναμονής. Ο αλγόριθμος που περιγράφει την ανανέωση των εικόνων παρουσιάζεται στο παρακάτω πίνακα.

Πίνακας 10. Αλγόριθμος για την δυναμική ανανέωση των ομάδων

1. Ανανέωση ομάδων (Εισαγωγή όλων των διανυσμάτων S^{GF} και εισαγωγή του
-

διανύσματος νέας εικόνας $S^{k_{wp}}$)

2. Αρχικοποίηση του δείκτη $runner = 0$
3. Αρχικοποίηση του *Loop 1* για την είσοδο μίας καινούργιας εικόνας στην ομάδα
 - a. Για $\forall S^{GF}$, αρχικοποίηση του *Loop 2*,
 - i. Για \forall θέση j του S^{GF} και $S^{k_{wp}}$, εάν στην θέση j , $S^{GF} OR S^{k_{wp}} = 1$, εισαγωγή της νέας εικόνας και αύξηση του δείκτη $runner$ κατά 1
 - b. Τέλος του *Loop 2*
4. Τέλος του *Loop 1*
5. Τέλος του *Loop 3*

7.3 Δυναμική Ανάκτηση Πολυμεσικών Δεδομένων

Στις προηγούμενες ενότητες παρουσιάστηκε το κομμάτι του συστήματος που οργανώνει τα πολυμεσικά δεδομένα σε μια συγκεκριμένη τοπολογία. Το σύστημα τώρα είναι έτοιμο να δεχτεί εικόνες (*queries*), έτσι ώστε να βρει τα δεδομένα από την βάση δεδομένων τα οποία σχετίζονται με αυτήν την εικόνα.

Αρχικά δημιουργείται ένα διάνυσμα για την εικόνα (*query*) ακριβώς με τον τρόπο που δημιουργήθηκε το διάνυσμα προηγουμένως s_{query} . Το διάνυσμα s_{query} έχει ακριβώς την ίδια μορφή με το διάνυσμα s^k . Το διάνυσμα s_{query} συγκρίνεται με όλα τα

διανύσματα των ομάδων. Το σύστημα επιλέγει εκείνα τα διανύσματα που έχουν τα περισσότερα κοινά κελιά με το διάνυσμα της υποβαλλόμενης εικόνας μιας και είναι πιθανό αυτή η ομάδα να έχει κοινά δεδομένα με την εικόνα. Κοινές ομάδες βρίσκονται χρησιμοποιώντας το λογικό **AND**. Οι υποψήφιες ομάδες επιλέγονται αν και εφόσον αν

$$s^{G_j} \text{ AND } s_{\text{query}} \neq 0 \quad (7.9)$$

Η Εξίσωση (7.9) παρουσιάζει ότι οι ομάδες οι οποίες δεν έχουν κοινά στοιχεία με την υποβαλλόμενη εικόνα απορρίπτονται. Αντίθετα στην περίπτωση όπου πολλές κοινές ομάδες παρουσιάζονται, τότε επιλέγονται αυτές οι ομάδες που έχουν τα περισσότερα κοινά στοιχεία με την υποβαλλόμενη εικόνα. Αυτές οι ομάδες που ανακτήθηκαν παρουσιάζονται στο τελικό χρήστη. Θα πρέπει να σημειωθεί ότι τα πλεονεκτήματα αυτής της προσέγγισης είναι ότι η απόκριση της ανάκτησης παρουσιάζει ιδιαίτερα χαμηλή υπολογιστική πολυπλοκότητα.

7.4 Πειραματικά Αποτελέσματα

Τα πειράματα πραγματοποιήθηκαν σε ένα υποσύνολο της βάσης δεδομένων *Corel*. Η βάση δεδομένων *Corel* είναι μία από τις πιο διαδεδομένες βάσεις για πειραματισμό και χρησιμοποιείται κυρίως για την αξιολόγηση της απόδοσης διαφόρων αλγορίθμων. Αυτή η βάση δεδομένων αποτελείται από **60,000** εικόνες μεσαίας ανάλυσης. Το υποσύνολο που χρησιμοποιήθηκε για τη διεξαγωγή των πειραμάτων αποτελείται από **12,000** εικόνες και έχει επιλεγεί προσεκτικά με τρόπο ώστε να συνδυαστούν τα διάφορα είδη των οπτικών χαρακτηριστικών όπως το χρώμα, την υφή και το σχήμα. Αρκετά πειράματα πραγματοποιήθηκαν προκειμένου να εξακριβωθεί η αποτελεσματικότητα του προτεινόμενου μοντέλου. Ειδικότερα, **1,000** τυχαίες ερωτήσεις (*queries*) υποβλήθηκαν στο σύστημα και τα αποτελέσματα αξιολογήθηκαν βάσει *α)* της υπολογιστικής αποδοτικότητας, *β)* του μέτρου σύγκρισης *precision/accuracy*.

Για τη διαδικασία αξιολόγησης, τα αντικειμενικά κριτήρια θεσπίστηκαν με στόχο να καταστεί δυνατή η δίκαιη σύγκριση της προτεινόμενης μεθόδου με άλλες αντίστοιχες μεθόδους στην τρέχουσα βιβλιογραφία. Συγκεκριμένα, τα κριτήρια αξιολόγησης επικεντρώθηκαν στις εξής δύο παραμέτρους *α)* στο υπολογιστικό κόστος που απαιτείται για την εκτέλεση μιας έρευνας σε μία πολυμεσική βάση δεδομένων, *β)* στην απόδοση ανάκτησης δεδομένων.

7.4.1 Παρουσίαση Αντικειμενικών Κριτηρίων

Ένα από τα κριτήρια που χρησιμοποιήθηκαν κατά την διαδικασία αξιολόγησης είναι η αναλογία βελτίωσης (*improvement ratio-IR*) του χρόνου απόκρισης. *IR* ορίζεται ως το κλάσμα του χρόνου που απαιτείται για τον εντοπισμό των σχετικών δεδομένων με τη σειριακή προσέγγιση κατά το χρόνο που απαιτείται για την προτεινόμενη

προσέγγιση. Επίσης είναι μια κανονικοποιημένη μετρική που εκφράζει το ποσοστό βελτίωσης του χρόνου απόκρισης. Δηλαδή,

$$IR = t_{\text{seq}} / t_{\text{parallel}} \quad (7.10)$$

Η Εξίσωση (7.10) παρουσιάζει την υπολογιστική απόδοση της προτεινόμενης μεθόδου. Ωστόσο, δεν μπορεί να εκφράσει πλήρως την απόδοση στην διαδικασία ανάκτησης δεδομένων από την άποψη της υπολογιστικής πολυπλοκότητας.

Για το λόγο αυτό, εισήχθηκε ένα νέο κριτήριο αξιολόγησης. Αυτό ονομάζεται απόδοση ανάκτησης (*retrieval efficiency-RE*) και χρησιμοποιείται για να μετρήσει τη διαφορά μεταξύ των αποτελεσμάτων που λαμβάνονται μέσω ανάκτησης του προτεινόμενου συστήματος και των αποτελεσμάτων που προκύπτουν από την διαδοχική αναζήτηση. Ειδικότερα, για ένα σταθερό αριθμό των ανακτήσεων, εκτιμάται ότι όλα τα στοιχεία είναι κοινά μεταξύ της προτεινόμενης και της διαδοχικής προσέγγισης. Επίσης, αυτός ο λόγος απόκλισης μπορεί να θεωρηθεί ως μια μέτρηση ομοιότητας για τα αποτελέσματα ανάκτησης. Με άλλα λόγια, η *RE* ορίζεται ως,

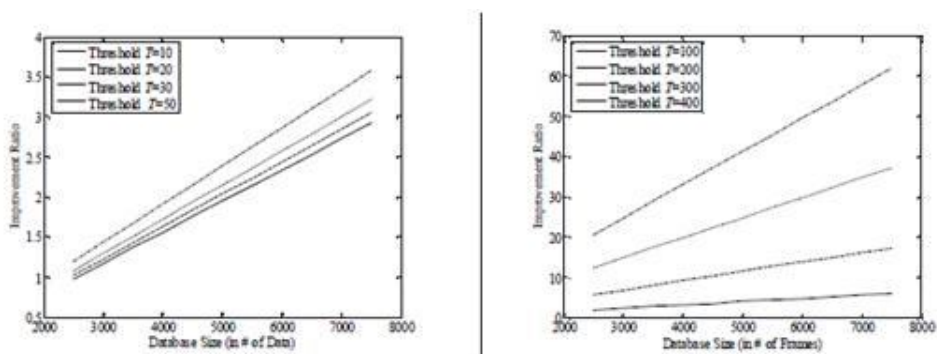
$$RE = W / L \quad (7.11)$$

όπου *L* είναι ο αριθμός των αποτελεσμάτων ανάκτησης που λαμβάνονται από τη διαδοχική προσέγγιση, ενώ *W* είναι ο αριθμός των κοινών εικόνων που ανακτήθηκαν από την προτεινόμενη μέθοδο με αυτές της διαδοχικής μεθόδου.

7.4.2 Παρουσίαση Αποτελεσμάτων

Το σύνολο δεδομένων που επιλέχθηκε για τα πειράματα αποτελείται από **12,000** εικόνες. Για όλες αυτές τις εικόνες, τρεις *MPEG-7* περιγραφείς (δηλαδή, *EHD*, *CLD* και *RSD*) χρησιμοποιήθηκαν για να περιγράψουν το περιεχόμενο των εικόνων. Θα πρέπει να τονιστεί ότι το προτεινόμενο σύστημα είναι επεκτάσιμο σε σχέση με τον αριθμό και τον τύπο των περιγραφέων, κάτι που σημαίνει ότι επιπλέον *MPEG-7* περιγραφές θα μπορούσαν να χρησιμοποιηθούν. Η αποτελεσματικότητα του προτεινόμενου συστήματος αξιολογήθηκε με την υποβολή **1,000** ερωτήσεων (*queries*).

Στο Σχήμα 31 παρακάτω παρουσιάζεται η βελτίωση του χρόνου απόκρισης σε σχέση με το μέγεθος της βάσης δεδομένων, για διαφορετικές τιμές του κατωφλίου **T**. Όπως μπορεί να παρατηρηθεί ο λόγος βελτίωσης του χρόνου απόκρισης αυξάνεται ανάλογα με το μέγεθος της βάσης δεδομένων.

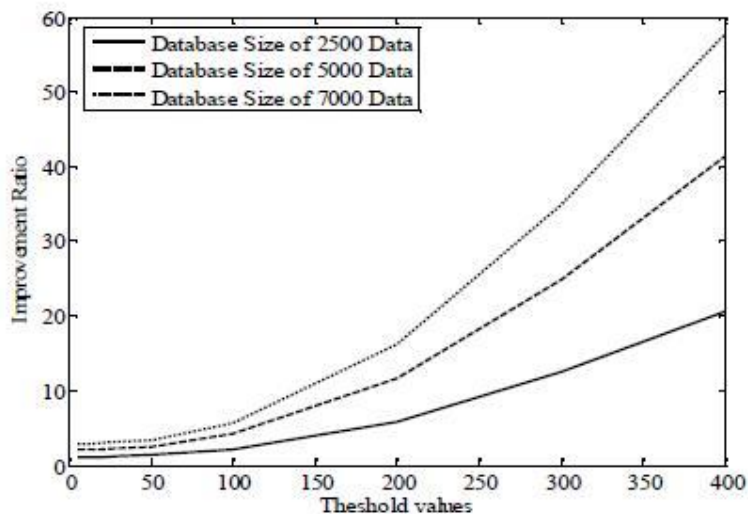


Σχήμα 31. Παρουσίαση του λόγου βελτίωσης έναντι του μεγέθους της βάσης δεδομένων για διαφορετικά κατώφλια **T: 10, 20, 30, 50, 100, 200, 300** και **400**

Επίσης, στο Σχήμα 32 παρακάτω απεικονίζεται η αναλογία βελτίωσης χρόνου απόκρισης σε σχέση με τις τιμές κατωφλίου **T**. Από την μορφή του γραφήματος, συμπεραίνεται ότι η αύξηση της υπολογιστικής απόδοσης αυξάνεται δραματικά σε

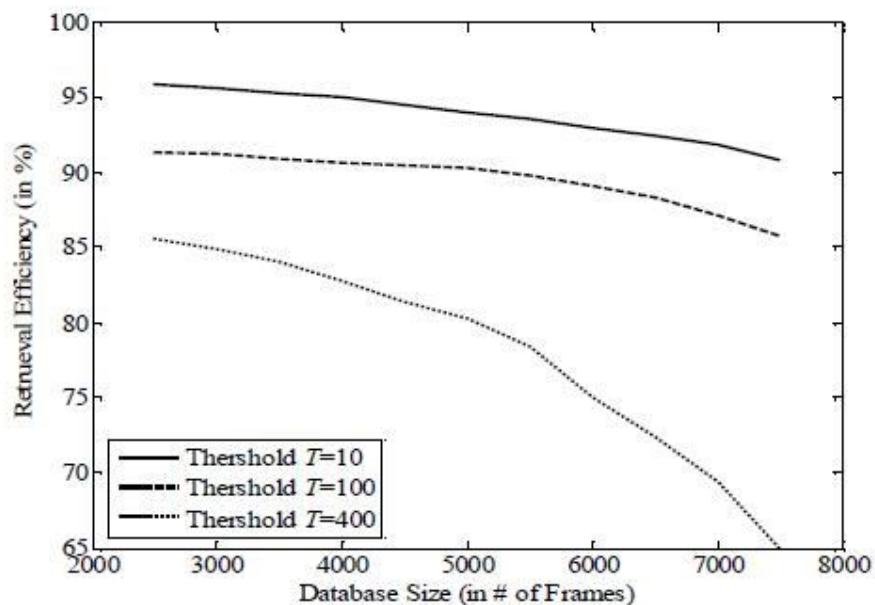
υψηλές τιμές του κατωφλίου T , ανεξαρτήτως του μεγέθους της βάσης δεδομένων.

Επιπλέον, η αύξηση του δείκτη βελτίωσης είναι περισσότερο εμφανής σε μεγάλα σύνολα δεδομένων ($> 7,000$).



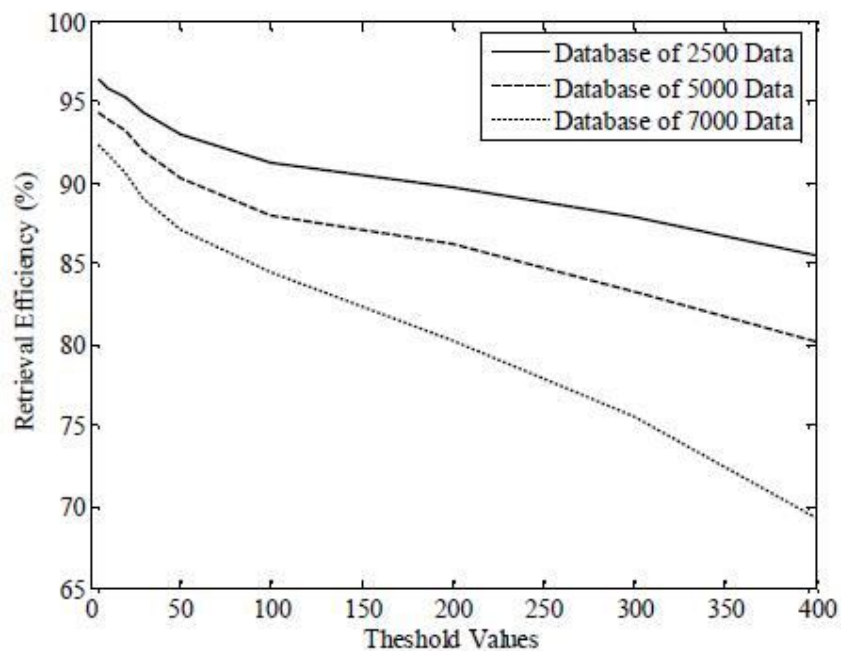
Σχήμα 32. Ο λόγος βελτίωσης έναντι του κατωφλίου T για διαφορετικά μεγέθη βάσης δεδομένων

Τα ακόλουθα αποτελέσματα επικεντρώνονται στη απόδοση της RE . Συγκεκριμένα, ο στόχος των παρακάτω πειραμάτων είναι η εξακρίβωση ότι η μείωση του υπολογιστικού κόστους οδηγεί σε μείωση της απόδοσης ανάκτησης.



Σχήμα 33. Η απόδοση ανάκτησης έναντι του μεγέθους της βάσης δεδομένων

Στο Σχήμα 33 παρουσιάζεται μια γραφική παράσταση η οποία δείχνει την απόδοση ανάκτησης (RE) σε σχέση με το μέγεθος της βάσης δεδομένων για διαφορετικές τιμές κατωφλίου T . Στην περίπτωση που οι τιμές του κατωφλιού είναι χαμηλές, η μείωση της αποδοτικότητας έχει ασήμαντο βαθμό. Ωστόσο, για τις υψηλές τιμές του κατωφλιού η απόδοση ανάκτησης μειώνεται σημαντικά. Οι υψηλές τιμές του κατωφλιού T οδηγούν σε υψηλά ποσοστά βελτίωσης της υπολογιστικής πολυπλοκότητας. Παρόμοια συμπεράσματα μπορούν να εξαχθούν από το Σχήμα 34 που φαίνεται παρακάτω.



Σχήμα 34. Η απόδοση ανάκτησης έναντι του κατωφλιού για διαφορετικά μεγέθη της βάσης δεδομένων

Τα πειραματικά αποτελέσματα που παρουσιάστηκαν, δείχνουν ότι η προτεινόμενη μέθοδος διατηρεί μία γραμμική απόκριση $O(n)$ σε σχέση με τις δυναμικές προσθήκες και αφαιρέσεις δεδομένων. Ωστόσο οι περισσότερες προσεγγίσεις που προτείνονται στη βιβλιογραφία (π.χ. *M-tree* [42]) έχουν $O(n \log n)$ χρόνο κατασκευής, μόνο στην περίπτωση ομοιόμορφων δεδομένων. Επίσης, η υπολογιστική πολυπλοκότητα της *HCT* [53] κάπου μεταξύ $O(n \log n)$ και $O(n^2)$.

8

Σύνοψη

Στο συγκεκριμένο κεφάλαιο περιλαμβάνεται η σύνοψη της διατριβής και τα συμπεράσματα που εξήχθησαν κατά την εκπόνησή της, η συνεισφορά και η καινοτομία που επιδεικνύει στον αντίστοιχο ερευνητικό χώρο, και συζητούνται θέματα μελλοντικής εργασίας και επέκτασης των ερευνητικών αποτελεσμάτων.

8.1 Συνεισφορά – Καινοτομία

Αυτή η διδακτορική διατριβή παρουσίασε την ροή αντικειμένου, η οποία είναι ένα μοντέλο εκτίμησης κίνησης αντικειμένων. Επάνω σε αυτή την προσέγγιση προτάθηκε ένα αξιόπιστο σύστημα παρακολούθησης αντικειμένων παράλληλα με ένα μηχανισμό ελέγχου δειγμάτων εκπαίδευσης. Επίσης, παρουσιάστηκε μία μέθοδος συσχέτισης δεδομένων, η οποία μπορεί να οργανώνει δυναμικά σε ομάδες (δηλαδή, σε πραγματικό χρόνο), διάφορες εικόνες που προέρχονται από ένα σύστημα παρακολούθησης.

Θα πρέπει να σημειωθεί ότι η εξαγωγή οπτικών πληροφοριών από εικόνες ή βίντεο είναι μία δύσκολη διαδικασία, δεδομένου ότι διάφορα μαθηματικά μοντέλα δεν μπορούν με ευκολία να ερμηνεύσουν πολύπλοκα δεδομένα. Για αυτό το λόγο οι αλγόριθμοι παρακολούθησης κάνουν σφάλματα λόγω της ύπαρξης θορύβου. Έτσι η αυτόματη παρακολούθηση αντικειμένων εξακολουθεί να είναι ένα δύσκολο

πρόβλημα προς επίλυση. Αντίθετα με το μεγαλύτερο μέρος της τρέχουσας βιβλιογραφίας, η παρούσα διατριβή παρουσίασε μία καινοτόμο τεχνική, η οποία μπορεί να λειτουργεί με αξιοπιστία κάτω από ιδιαίτερα δύσκολες συνθήκες, όπως σε ένα βιομηχανικό περιβάλλον, όπου τόσο η εμφάνιση του αντικειμένου όσο και η κίνηση του περιβάλλοντος αλλάζει σημαντικά. Επάνω σε αυτό το πλαίσιο αναπτύχθηκε ένας μηχανισμός μάθησης πραγματικού χρόνου, ο οποίος μπορεί να επαληθεύει ένα δείγμα εκπαίδευσης και να απορρίπτει τα δείγματα εκείνα που περιέχουν θορυβώδη χαρακτηριστικά. Πιο συγκεκριμένα η συνεισφορά και καινοτομία επικεντρώνεται στα εξής σημεία:

- **Προτείνεται μια καινοτόμος μέθοδος για τον υπολογισμό της μετατόπισης ενός αντικειμένου σε μια εικονοσειρά.** Η τεχνική αυτή ονομάζεται ροή αντικειμένου (*object flow*) και έχει την ικανότητα να προσδιορίζει το πεδίο κίνησης ενός κινούμενου αντικειμένου, με παρόμοιο τρόπο με την οπτική ροή. Παράλληλα μπορεί να απορρίπτει άλλες άσχετες κινήσεις στο εσωτερικό μιας σκηνής (π.χ. άλλα κινούμενα αντικείμενα ή κινούμενο φόντο). Οι ιδιότητες της προτεινόμενης μεθόδου παρουσιάστηκαν ποιοτικά στο Σχήμα 17. Επίσης, τα ποσοτικά αποτελέσματα έδειξαν ότι η ροή αντικειμένου έχει μικρότερη μέση τιμή σφάλματος και διακύμανσης από την μέθοδο οπτικής ροής.
- **Προτείνεται μία αξιόπιστη προσέγγιση παρακολούθησης αντικειμένων που βασίζεται στην μέθοδο ροής αντικειμένου.** Θα πρέπει να σημειωθεί ότι η συνήθης πρακτική για τους αλγορίθμους παρακολούθησης είναι να επιλέγεται ως μοντέλο κίνησης η δυναμική κατάσταση του αντικειμένου. Ωστόσο, όταν το αντικείμενο που παρακολουθείται κάνει μία απότομη κίνηση, η μοντελοποίηση της κίνησης μέσω της δυναμικής κατάστασης μπορεί να είναι αναποτελεσματική. Ως εκ τούτου, η επιλογή ενός

αποτελεσματικού μοντέλου κίνησης είναι ιδιαίτερος σημαντική για την σωστή λειτουργία του αλγορίθμου παρακολούθησης. Η προτεινόμενη μέθοδος υλοποιήθηκε στα πλαίσια ενός φίλτρου σωματιδίων και χρησιμοποιεί την ροή αντικειμένου τόσο ως μοντέλο κίνησης όσο ως μοντέλο περιγραφής αντικειμένου.

- **Στο προτεινόμενο αλγόριθμο παρακολούθησης, εισάγεται ένα επιπλέον μοντέλο περιγραφής αντικειμένου, με σκοπό να αντιμετωπίσει προβλήματα που σχετίζονται με αναξιόπιστες ανιχνεύσεις αντικειμένου.** Αυτό το μοντέλο περιγραφής χρησιμοποιεί μια παραγωγική *prior* μέθοδο. Θα πρέπει να σημειωθεί ότι διάφορες γραμμικές παραγωγικές μέθοδοι ταξινόμησης δεν μπορούν να αντιμετωπίσουν τις δραματικές αλλαγές στην εμφάνιση του αντικειμένου που παρακολουθείται. Μία παραγωγική μέθοδος που μπορεί να αντιμετωπίσει τέτοια προβλήματα είναι η *GPLVM*, η οποία ανακαλύπτει χαμηλών διαστάσεων υπό-επίπεδα (*low dimensional manifold*) χρησιμοποιώντας ένα μικρό αριθμό δειγμάτων εκπαίδευσης. Σε αυτή την προσέγγιση το αντικείμενο υπό μελέτη υποβάλλεται σε αρκετές αλλαγές στην εμφάνιση του. Έτσι η μέθοδος *GPLVM* μπορεί να παρέχει μία αποτελεσματική περιγραφή της κατάστασης της εμφάνισης του.
- **Προτείνεται ένας αποτελεσματικός μηχανισμός ελέγχου δειγμάτων εκπαίδευσης σε πραγματικό χρόνο.** Θα πρέπει να σημειωθεί ότι, το κλειδί για την καλή απόδοση ενός συστήματος παρακολούθησης, το οποίο ανανεώνεται σε πραγματικό χρόνο, είναι η ανάπτυξη ενός αξιόπιστου μηχανισμού επαλήθευσης του δείγματος εκπαίδευσης. Ο προτεινόμενος μηχανισμός έχει τη δυνατότητα να απορρίπτει τα δείγματα εκπαίδευσης που περιέχουν θορυβώδη χαρακτηριστικά (π.χ. επικαλύψεις από άλλα αντικείμενα κινούμενα η στατικά, βλ. Σχήμα 23), χρησιμοποιώντας τις πληροφορίες από τα

διαγράμματα απόκρισης (δηλαδή, τις αποκρίσεις ταξινομητή, βλ. Σχήμα 35 (β), (δ), (ζ) και (θ)) για την περιοχή που περικλείει ένα δείγμα εκπαίδευσης. Ο συγκεκριμένος μηχανισμός χρησιμοποιείται για να την δυναμική ανανέωση της ροής αντικειμένου.

- **Προτείνεται μία καινοτόμος μέθοδος συσχέτισης δεδομένων, η οποία μπορεί να οργανώνει δυναμικά σε ομάδες, διάφορες εικόνες που προέρχονται από ένα σύστημα παρακολούθησης.** Συγκεκριμένα, ο προτεινόμενος αλγόριθμος δεικτοδότησης εικόνων έχει την ευελιξία να είναι ανεξάρτητος από το περιεχόμενο της βάσης δεδομένων και τη δομή. Αυτό το γεγονός τον κάνει κατάλληλο για την συγκεκριμένη εφαρμογή παρακολούθησης. Επίσης, η προτεινόμενη μέθοδος βασίζεται στο *MPEG-7* και εκμεταλλεύεται την διαλειτουργικότητα, μεταφερσιμότητα και προσβασιμότητα της αναπαραστώμενης πληροφορίας. Επίσης, η εκμετάλλευση της προσέγγισης συστοιχιών (*clustering*) χρησιμοποιείται για τη βελτίωση του χρόνου απόκρισης ανακτήσεων, λόγω του ότι ο χώρος εύρεσης σχετικού περιεχομένου ελαττώνεται σημαντικά. Αυτό το γεγονός θεωρείται ιδιαίτερα σημαντικό θέμα για εφαρμογές ανάκτησης βάσει περιεχομένου.

8.2 Μελλοντική Εργασία

Η μελλοντική εργασία θα επικεντρωθεί στην κλιμάκωση της προτεινόμενης προσέγγισης, ιδιαίτερα με στόχο την ενσωμάτωση της σε συστήματα παρακολούθησης, που λειτουργούν σε καθημερινή βάση και ελέγχουν πολυσύχναστες σκηνές. Πιο συγκεκριμένα η παρούσα ερευνητική εργασία μπορεί να επεκταθεί προς τις εξής κατευθύνσεις :

- **Επανεξέταση του αλγορίθμου *Boosting* πραγματικού χρόνου**, με στόχο να μπορεί να ανανεώνεται χρησιμοποιώντας χρονική πληροφορία από πολλαπλές

εικονοσειρές. Αυτό θα έδινε μία ολοκληρωμένη και ενιαία μέθοδο ροής αντικειμένου, η οποία θα εμπλουτιζόταν με τους κατάλληλους μηχανισμούς ελέγχου δειγμάτων εκπαίδευσης.

- **Επανεξέταση του αλγορίθμου *GPLVM*, με στόχο να μπορεί να ανανεώνεται σε πραγματικό χρόνο.** Θα πρέπει να σημειωθεί ότι αυτό είναι μία δύσκολη διαδικασία μίας και ο αλγόριθμος *Scaled Conjugate Gradients (SCG)* χρησιμοποιεί πολύπλοκες και χρονοβόρες στον υπολογισμό μαθηματικές διαδικασίες. Ωστόσο κάτι τέτοιο θα μπορούσε να παρακαμφθεί με τις κατάλληλες ευριστικές (*heuristics*) προσεγγίσεις.
- **Υλοποίηση της προτεινόμενης προσέγγισης οργάνωσης πολυμεσικών δεδομένων σε κατανεμημένο υπολογιστικό περιβάλλον (*Cloud Computing*).** Αυτή η τεχνική θα μείωνε περαιτέρω την υπολογιστική πολυπλοκότητα και τον χρόνο απόκρισης και οργάνωσης δεδομένων.

Η σελίδα αυτή είναι σκόπιμα λευκή.

Βιβλιογραφικές Αναφορές

- [1] Yu, J., Amores, J., Sebe, N., Radeva, P., Tian, Q.: Distance learning for similarity estimation. *IEEE Trans. on PAMI* (2008)
- [2] Hertz, T., Bar-Hillel, A., Weinshall, D.: Learning distance functions for image retrieval. In: *Proc. CVPR. Volume 2.* (2004) 570–577
- [3] Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: *Proc. CVPR. Volume I.* (2001) 511–518
- [4] Grabner, H., Bischof, H.: On-line boosting and vision. In: *Proc. CVPR. Volume 1.* (2006) 260–267
- [5] Hertz, T., Bar-Hillel, A., Weinshall, D.: Learning distance functions for image retrieval. In: *Proc. CVPR. Volume 2.* (2004) 570–577
- [6] Werlberger, M., Trobin, W., Pock, T., Wedel, A., Cremers, D., Bischof, H.: Anisotropic Huber-L1 optical flow. In: *BMVC, London, UK* (2009)
- [7] Prisacariu, V., Reid, I.: fasthog - a real-time gpu implementation of hog. Technical Report 2310/09, (Department of Engineering Science, Oxford University)
- [8] Kanade, T., Lucas, B.: An iterative image registration technique with an application to stereo vision. In: *Proc. Int. Joint Conf. on Artificial Intelligence.* (1981) 674–679
- [9] Freund, Y., Schapire, R.: A short introduction to boosting. In: *Journal of Japanese Society for Artificial Intelligence*, 14(5):771-780, (1999)
- [10] K. Tieu and P. Viola. Boosting image retrieval. In *Proc. CVPR*, pages 228–235, 2000.

- [11] R. Lienhart and J. Maydt.: An extended set of haar-like features for object detection. In Proceedings International Conference on Image Processing, pages 900-903, 2002.
- [12] Rabiner, L.R.: A tutorial on hidden Markov models and selected applications in speech recognition. In Proceedings of IEEE, pages 257 – 286, 1989
- [13] Sanjeev Arulampalam, Simon Maskell, Neil Gordon and Tim Clapp.: A tutorial on Particle Filters for Online Nonlinear/Non Gaussian Bayesian Tracking. In Transactions of IEEE on signal processing, vol. 50, no. 2, February 2002.
- [14] Vasilios Anagnostopoulos, Sotirios Chatzis, Constantinos Lalos, Anastasios D. Doulamis, Dimitrios I. Kosmopoulos, Theodora A. Varvarigou, Helmut Neuschmied, Georg Thallinger, Stuart E. Middleton, Matthew Addis, Eduardo Bustos, Fabrizio Giorgini: A Cross Media Platform for Personalized Leisure & Entertainment: The POLYMNIA Approach. AXMEDIS 2006: 283-290.
- [15] Vasilios Anagnostopoulos, Dimitrios I. Kosmopoulos, Anastasios D. Doulamis is, Alexandros Makris, Constantinos Lalos, and Theodora A. Varvarigou, "Automated production of personalized video content for visitors of thematic parks", 2nd IET Conference on Intelligent Environments, Volume 1, Pages: 173-181 , July 2006
- [16] Constantinos Lalos, Nikolaos Doulamis, Anastasios Doulamis, Kleopatra Konstanteli, and Theodora Varvarigou, "An Innovative Content-based Indexing Technique with Linear Response suitable for Pervasive Environments," IEEE International Conference on Content –based Multimedia Indexing, London, UK, 2008
- [17] Li, Y., Ai, H., Yamashita, T., Lao, S., Kawade, M.: Tracking in low frame rate video: A cascade particle filter with discriminative observers of different life spans. In: In CVPR (2007)
- [18] Lu, W.L., Okuma, K., Little, J.J.: Tracking and recognizing actions of multiple hockey players using the boosted particle filter. Image Vision Comput. 27, 189–205 (2009)

- [19] Khan, Z., Balch, T., Dellaert, F.: An MCMC-Based Particle Filter for Tracking Multiple Interacting Targets. In: In ECCV (2004)
- [20] Breitenstein, M.D., Reichlin, F., Leibe, B., Koller-Meier, E., Gool, L.V.: Robust tracking-by-detection using a detector confidence particle filter. In: In ICCV (2009)
- [21] Leibe, B., Schindler, K., Gool, L.V.: Coupled detection and trajectory estimation for multi-object tracking. In: In ICCV (2007)
- [22] Yang, M., Lv, F., Xu, W., , Gong, Y.: Detection driven adaptive multi-cue integration for multiple human tracking. In: In ICCV (2009)
- [23] Jean-Marc Odobez Daniel Gatica-Perez, S.B.: Embedding motion in model-based stochastic tracking. IEEE Transactions on Image Processing 15, 3515–3531 (2006)
- [24] Kwon, J., Lee, K., Park, F.: Visual tracking via geometric particle filtering on the affine group with optimal importance functions. In: In CVPR (2009)
- [25] Andriluka, M., Roth, S., Schiele, B.: People-tracking-by-detection and people-detection-by-tracking. In: In CVPR (2008)
- [26] Collins, R., Liu, Y., Leordeanu, M.: Online selection of discriminative tracking features. PAMI 27, 1631–1643 (2005)
- [27] Leonardis, A., Bischof, H.: Robust recognition using eigenimages. CVIU 78, 99–118 (2000)
- [28] Turk, M., Pentland, A.: Face recognition using eigenfaces. In: In CVPR, pp. 586–591 (1991)
- [29] Woodley, T., Stenger, B., Cipolla, R.: Tracking using online feature selection and a local generative model. In: In BMVC (2007)
- [30] Liu, B., Yang, L., Huang, J., Meer, P., Gong, L., Kulikowski, C.: Robust and fast collaborative tracking with two stage sparse optimization. In: In ECCV (2010)
- [31] P. Perez, C. Hue, J. Vermaak and M. Gangnet Color-Based Probabilistic Tracking, European Conference on Computer Vision, pp 661-675, 2002

- [32] Chunhua Shen, Anton van den Hengel, and Anthony Dick "Probabilistic Multiple Cue Integration for Particle Filter Based Tracking", Proc. VIIth Digital Image Computing: Techniques and Applications, Sun C., Talbot H., Ourselin S. and Adriaansen T.(Eds.), 10-12 Dec. 2003, Sydney
- [33] Zia Khan, Tucker Balch and Frank Dellart,. A Rao-Blackwellized Particle Filter for EigenTracking
- [34] Xinyu Xu and Baoxin Li., Rao-Blackwellized Filter For tracking with Application in visual Surveillance, Proceedings of the 2nd Joint IEEE international Workshop on VS-PETS, Beijing, October 15-16 ,2005
- [35] Horn, B., Schunck, B.: Determining optical flow. *Artificial Intelligence* 17 (1981) 185–203
- [36] Brox, T., Bregler, C., Malik, J.: Large displacement optical flow. In: Proc. CVPR. (2009)
- [37] Sun, D., Roth, S., Lewis, J., Black, M.: Learning optical flow. In: Proc. ECCV. (2008)
- [38] Shi, J., Malik, J.: Motion segmentation and tracking using normalized cuts. In: Proc. ICCV. (1998)
- [39] Ali, S., Shah, M.: A lagrangian particle dynamics approach for crowd flow segmentation and stability analysis. In: Proc. CVPR. (2007)
- [40] Santner, J., Werlberger, M., Mauthner, T., Paier, W., Bischof, H.: FlowGames. In: 1st Int. Workshop on CVCG in conjunction with CVPR. (2010)
- [41] Stalder, S., Grabner, H., Gool, L.V.: Beyond semi-supervised tracking: Tracking should be as simple as detection, but not simpler than recognition. In: Proc. IEEE WS on On-line Learning for Computer Vision. (2009)
- [42] J.L. Bentley, "Multi-dimensional Binary Search Trees used for Associative Searching," *ACM Proc. Commun.* Vol. 18, No. 9, pp. 509-517, 1975.
- [43] A. Gutman, "R-trees: A Dynamic Index Structure for Spatial Searching," *in Proc. ACM SIGMOD*, p.. 45-57, 1984.

- [44] N. Beckman, H.-P. Kriegel, R. Schneider and B. Seeger, "The R*-trees: An Efficient and Robust Access Method for Points and Rectangles," *ACM SIGMOD*, pp. 322-331, 1990
- [45] S. Berchtold, D.A. Kein, and H.-P. Kriegel, "The X-tree: An Index Structure for High Dimensional Data," *In Proc. of 22th Int. Con. Very Large Databases*, 1996.
- [46] A. Schmidt, S. Spiekermann, A. Gershman, and F. Michahelles, Real-World Challenges of Pervasive Computing," *IEEE Pervasive Computing*, Vol. 5, No. 3, pp. 91-93, July-Sept. 2006.
- [47] X. Chen, M. O'Grady, G. O'Hare, "Towards a Pervasive Mobile Multimedia Experience," *IEEE CISIM* , pp. 209-212, June 2007
- [48] X. Hu, U. Y. Ogras, N. H. Zamora, and R. Marculescu, "Data Partitioning Techniques for Pervasive Multimedia Platforms," *IEEE Inter. Conf. on Multimedia and Expo*, Vol. 2, pp. 1035-1038, June 2004.
- [49] X. Chen, M. J. O'Grady, G.M.P. O'Hare, "Towards a Pervasive Mobile Multimedia Experience," *IEEE 6th International Conference on Computer Information Systems and Industrial Management Applications (CISIM)* , pp. 209-212, June 2007.
- [50] D. Linner, F. Kirsch, I. Radosch, and S. Steglich, "Context aware multimedia provisioning for pervasive games," *Seventh IEEE International Symposium on Multimedia*, pp 1-9, Dec. 2005
- [51] J.L. Bentley, "Multi-dimensional Binary Search Trees used for Associative Searching," *ACM Proc. Commun.* Vol. 18, No. 9, pp. 509-517, 1975.
- [52] S.-c S. Cheung and A. Zakhor, "Fast Similarity Search and Clustering of Video Sequences on the World-Wide-Web," *IEEE Trans. on Multimedia*, Vol. 7, No. 3, pp. 524-537, June 2005
- [53] S. Kiranyaz and M. Gabbouj, "Hierarchical Cellular Tree: An Efficient Indexing Scheme for Content-Based Retrieval on Multimedia Databases," *IEEE Trans. Multimedia*, vol. 9, No. 1, pp. 102-119, 2007.

- [54] G. Wang, X. Zhou, Bin Wang, B. Qiao, D. Han, "A Hyper plane based Indexing Technique for High-dimensional Data," *Information Sciences* Vol. 177, pp. 2255–2268, 2007
- [55] Peter Stanchev, David Green Jr, and Boyan Dimitrov, "MPEG-7: The multimedia content description interface", *International Journal "Information Theories & Applications"* vol. 11
- [56] B.S. Manjunath, J.-R. Ohm, Vinod V. Vasudevan, A. Yamada, "Color and texture descriptors" *IEEE Transactions on CSVT*, Vol.11, No. 6, June 2001.
- [57] N. Oza and S. Russell. Online bagging and boosting. In *Proc. Artificial Intelligence and Statistics*, pages 105–112, 2001.
- [58] N. Oza and S. Russell. Experimental comparisons of online and batch versions of bagging and boosting. In *Proc. 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2001.
- [59] Lawrence, N.D.: Probabilistic non-linear principal component analysis with Gaussian process latent variable models. *Journal of Machine Learning Research* 6, 1783–1816 (2005)
- [60] Geiger, A., Urtasun, R., Darrell, T.: Rank priors for continuous non-linear dimensionality reduction. In: *In CVPR* (2009)
- [61] Wang, J.M., Fleet, D.J., Member, S., Hertzmann, A.: Gaussian process dynamical models for human motion. *PAMI* (2007)
- [62] Lawrence, N.D., , Quinonero Candela, J.: "Local distance preservation in the GP-LVM through backconstraints. In: *Proceedings of the 23rd international conference on Machine learning, ICML* (2006)
- [63] Katja Nummiaro, Esther Koller-Muller and Luc Van Gool,: "An adaptive Color-Based Particle Filter" Submitted to Elsevier Science, 19 September 2002
- [64] Junyi Zuo, Chunhui Zhao, Yongmei Cheng and Hongcai,: "Zhang Particle Filter Based Visual Tracking Using New Observation Model", *IEEE International Conference on Automation and Logistics*, 18-21, 2007

- [65] Yogesh Rathi, Namrata Vaswani and Allen Tannenbaum,: “A generic framework for tracking using particle filter with dynamics shape prior” IEEE transactions on image processing, vol. 16, no.5 , May 2007
- [66] Jianyu Wang, Debin Zhao, Shiguang Shan and Wen Gao,: “Approximating Inference on Complex Motion Models using Multi-Model Particle Filter”, Lecture Notes in Computer Science, Springer Berlin / Heidelberg, vol. 3332/2005, Advances in Multimedia Information Processing - PCM 2004
- [67] Christopher Bishop Pattern recognition and machine learning” 2006 Springer Science + Business Media LLC ISBN-10: 0-387-31073-8
- [68] Zia Khan , Tucker Balch and Frank Dellart Efficient Particle Filter- Based Tracking of Multiple Interacting Targets using and MRF- based motion
- [69] Voulodimos, A., Kosmopoulos, D., Vasileiou, G., Sardis, E., Doulamis, A., Anagnostopoulos, V., Lalos, C., Varvarigou, T.: A dataset for workflow recognition in industrial scenes. In: IEEE Image Processing (ICIP), International Conference on (2011)
- [70] Doulamis, A.: Dynamic tracking re-adjustment: a method for automatic tracking recovery in complex visual environments. Multimedia Tools and Applications 50, 49–73 (2010)
- [71] Panagiotakis, C., Doulamis, A., Tziritas, G.: Equivalent key frames selection based on iso-content principles. IEEE Trans. Cir. and Sys. for Video Technol. 19, 447–451 (2009)
- [72] Phung, S.L., Bouzerdoun, A., Chai, D.: A novel skin color model in YCbCr color space and its application to human face detection. Image Processing. International Conference on 1, 289–292 (2002)
- [73] Babenko, B., Yang, M.H., S. Belongie, S.: Visual tracking with online multiple instance learning. In: In CVPR (2009)

[74] H. Lu, B. Chin Ooi, H. Tao Shen and X. Xue, "Hierarchical Indexing Structure for Efficient Similarity Search in Video Retrieval," *IEEE Trans. on Knowledge And Data Engineering*, Vol. 18, No. 11, pp. 1544-1559, Non. 2006

[75] D. Halkos, A. Yannopoulos, A. Doulamis, N. Doulamis, G. Mourkousis, K. Tsiara, T. Varvarigou, "A Secure Framework Exploiting Content Guided and Automated Algorithms for Real Time Video Searching," *Multimedia Tools and Applications*



Ο Δρ. **Κωνσταντίνος Γ. Λάλος** γεννήθηκε την 30^η Σεπτεμβρίου 1982 στην Αθήνα. Τον Ιούνιο του 2000 αποφοίτησε από το 4^ο Ενιαίο Λύκειο Γλυφάδας. Τον Ιούνιο του 2003 έλαβε δίπλωμα Ηλεκτρονικού και Τηλεπικοινωνιών Μηχανικού του Πανεπιστημίου του Μπέρμινγκχαμ (Αγγλία). Επίσης, τον Σεπτέμβριο του 2005 έλαβε μεταπτυχιακό δίπλωμα με θέμα «Συστήματα τηλεπικοινωνιών και επεξεργασία σήματος» από την σχολή Ηλεκτρονικών και Ηλεκτρολόγων Μηχανικών του Πανεπιστημίου του Μπρίστολ (Αγγλία). Η διπλωματική του εργασία είχε ως θέμα τον αυτόματο διαχωρισμό ηχητικών πηγών (Blind Source Separation).

Στον χρόνο που ακολούθησε μετά την αποφοίτηση του εργάστηκε ως ερευνητής με αντικείμενο την όραση υπολογιστών, στο επιστημονικό έργο POLYMNIA (IP project), το οποίο χρηματοδοτήθηκε από την Ευρωπαϊκή Ένωση. Λόγω της επιτυχημένης ερευνητικής του πορείας του δόθηκε μία θέση υποψηφίου διδάκτορα, η οποία και εγκρίθηκε στις 29/11/2005 από την επιστημονική επιτροπή του Εθνικού Μετσόβιου Πολυτεχνείου.

Κατά τη διάρκεια της εκπόνησης της διδακτορικής του διατριβής, ο Κ. Λάλος εργάστηκε σε Ευρωπαϊκά Προγράμματα, στα οποία του δόθηκε η δυνατότητα να εμβαθύνει σε ερευνητικά θέματα, άμεσα συνδεδεμένα με το αντικείμενο της διατριβής. Περαιτέρω, έλαβε μέρος σε επιστημονικά συνέδρια και συνεργάστηκε με διάφορους οργανισμούς, εταιρίες και πανεπιστημιακές ομάδες. Τα

αποτελέσματα της ερευνητικής του εργασίας παρουσιάστηκαν σε διεθνή συνέδρια και δημοσιεύθηκαν σε έγκριτα επιστημονικά περιοδικά. Ο Κ. Λάλος είναι μέλος του Τεχνικού Επιμελητηρίου της Ελλάδος (ΤΕΕ) από το 2006.