ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΕΦΑΡΜΟΣΜΕΝΩΝ ΜΑΘΗΜΑΤΙΚΩΝ ΚΑΙ ΦΥΣΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

ΔΙΑΤΜΗΜΑΤΙΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ

«ΜΑΘΗΜΑΤΙΚΗ ΠΡΟΤΥΠΟΠΟΙΗΣΗ
σε ΣΥΓΧΡΟΝΕΣ ΤΕΧΝΟΛΟΓΙΕΣ και την ΟΙΚΟΝΟΜΙΑ»

ΑΝΑΠΤΥΞΗ ΚΑΙ ΑΞΙΟΛΟΓΗΣΗ ΜΕΘΟΔΟΛΟΓΙΑΣ ΓΙΑ ΤΗΝ ΠΑΡΑΚΟΛΟΥΘΗΣΗ ΠΟΛΛΑΠΛΩΝ ΑΝΤΙΚΕΙΜΕΝΩΝ ΣΕ ΑΚΟΛΟΥΘΙΕΣ ΕΙΚΟΝΩΝ

ΑΘΗΝΑ ΨΑΛΤΑ
ΑΡΙΘΜΟΣ ΜΗΤΡΩΟΥ: 09315042

ΕΠΙΒΛΕΠΩΝ ΚΑΘΗΓΗΤΗΣ: ΚΩΝΣΤΑΝΤΙΝΟΣ ΚΑΡΑΝΤΖΑΛΟΣ

ΑΘΗΝΑ, 24/10/2017

NATIONAL TECHNICAL UNIVERSITY OF ATHENS

SCHOOL OF APPLIED MATHEMATICAL AND PHYSICAL SCIENCES

MASTER THESIS

MSc in «MATHEMATICAL MODELING IN MODERN TECHNOLOGIES AND THE ECONOMICS»

DEVELOPMENT AND VALIDATION OF A METHODOLOGY FOR MULTIPLE OBJECT TRACKING

ATHENA PSALTA
STUDENT ID: 09315042

SUPERVISOR: KONSTANTINOS KARANTZALOS

ATHENS, 24/10/2017

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΕΦΑΡΜΟΣΜΕΝΩΝ ΜΑΘΗΜΑΤΙΚΩΝ ΚΑΙ ΦΥΣΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

ΔΙΑΤΜΗΜΑΤΙΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ

«ΜΑΘΗΜΑΤΙΚΗ ΠΡΟΤΥΠΟΠΟΙΗΣΗ σε ΣΥΓΧΡΟΝΕΣ ΤΕΧΝΟΛΟΓΙΕΣ και την ΟΙΚΟΝΟΜΙΑ»

ΑΝΑΠΤΥΞΗ ΚΑΙ ΑΞΙΟΛΟΓΗΣΗ ΜΕΘΟΔΟΛΟΓΙΑΣ ΓΙΑ ΤΗΝ ΠΑΡΑΚΟΛΟΥΘΗΣΗ ΠΟΛΛΑΠΛΩΝ ΑΝΤΙΚΕΙΜΕΝΩΝ ΣΕ ΑΚΟΛΟΥΘΙΕΣ ΕΙΚΟΝΩΝ

ΑΘΗΝΑ ΨΑΛΤΑ
ΑΡΙΘΜΟΣ ΜΗΤΡΩΟΥ: 09315042

ΤΡΙΜΕΛΗΣ ΕΞΕΤΑΣΤΙΚΗ ΕΠΙΤΡΟΠΗ:

……..	…….	……..
ΚΑΡΑΝΤΖΑΛΟΣ ΚΩΝΣΤΑΝΤΙΝΟΣ	ΑΡΓΙΑΛΑΣ ΔΗΜΗΤΡΙΟΣ	ΚΑΡΡΑΣ ΓΕΩΡΓΙΟΣ
Επ. Καθηγητής Σ.Α.Τ.Μ. Ε.Μ.Π.	Καθηγητής Σ.Α.Τ.Μ. Ε.Μ.Π.	Καθηγητής Σ.Α.Τ.Μ. Ε.Μ.Π.

ΑΘΗΝΑ, Οκτώβριος 2017

..... Αθηνά Β. Ψάλτα

Διπλωματούχος Αγρονόμος και Τοπογράφος Μηχανικός Ε.Μ.Π.

"If I have seen further it is by standing on ye shoulders of Giants"

-Isaac Newton-

Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή μου κ. Καράντζαλο για την καθοδήγηση που μου παρείχε κατά τη διάρκεια της μεταπτυχιακής μου εργασίας. Ακόμη, θα ήθελα να ευχαριστήσω τον Β. Τσιρώνη για τη βοήθεια που μου παρείχε, καθώς και την οικογένειά μου και τα άτομα που βρίσκονται κοντά μου για την αμέριστη στήριξή τους κατά τη διάρκεια των σπουδών μου.

ΠΕΡΙΛΗΨΗ

Η οπτική παρακολούθηση (visual tracking) αποτελεί ένα σημαντικό πρόβλημα της Όρασης Υπολογιστών για το οποίο τις τελευταίες δεκαετίες η ερευνητική κοινότητα δείχνει έντονο ενδιαφέρον. Τόσο η παρακολούθηση ενός αντικειμένου, όσο και η παρακολούθηση πολλαπλών αντικειμένων σε μία ακολουθία εικόνων είναι μια πολύπλοκη διαδικασία και ειδικότερα για την τελευταία οι αλγόριθμοι αιχμής απέχουν ακόμη από την πλήρη επίλυση του προβλήματος. Στη παρούσα εργασία μελετάται ενδελεχώς η διεθνής βιβλιογραφία γύρω από το θέμα αυτό και αναπτύσσεται ένας αλγόριθμος παρακολούθησης πολλαπλών στόχων.

Η βιβλιογραφική αναδρομή χωρίζεται σε δύο μέρη. Πρώτα μελετάται το πρόβλημα της παρακολούθησης ενός αντικειμένου τόσο ως προς τη δυσκολία του όσο και ως προς τον τρόπο προσέγγισης διάφορων αλγορίθμων αιχμής για την επίλυσή του. Στο δεύτερο μέρος αναλύεται το πρόβλημα της παρακολούθησης πολλαπλών αντικειμένων και μελετώνται οι σχετικές προσεγγίσεις της βιβλιογραφίας. Ειδικότερα, τονίζονται οι διαφορές τόσο στη μεθοδολογική προσέγγιση όσο και στη δυσκολία για το συγκεκριμένο πρόβλημα έναντι του προβλήματος του πρώτου μέρους. Επιπρόσθετα, γίνεται ιδιαίτερη αναφορά στη χρήση τεχνητών νευρωνικών δικτύων για την επίλυση του προβλήματος και παρουσιάζεται αναλυτικά ο αλγόριθμος αιχμής MDP (Markov Decision Processes).

Στο δεύτερο σκέλος της εργασίας αναπτύσσεται και αξιολογείται ένας πρωτότυπος αλγόριθμος οπτικής παρακολούθησης πολλαπλών αντικειμένων. Αρχικά, παρουσιάζεται η αναλυτική περιγραφή του αλγορίθμου με ιδιαίτερη έμφαση αφενός στην “παρακολούθηση μέσω ανιχνεύσεων” (tracking-by-detection) προσέγγιση που υιοθετήθηκε αφετέρου στα επιμέρους υποσυστήματα του αλγορίθμου. Ειδικότερα, αναλύονται τα μοντέλα κίνησης και εμφάνισης που χρησιμοποιούνται, καθώς και η σύνθετη διαδικασία αντιστοίχισης των δεδομένων (στόχων-ανιχνεύσεων) που βασίζεται σε κινηματικά, φασματικά και χωρικά κριτήρια.

Ο αλγόριθμος που αναπτύχθηκε αξιολογείται στη συνέχεια στο σύνολο δεδομένων 2DMOT15 και οι επιδόσεις του αντιπαρατίθενται με τις αντίστοιχες μιας απλοϊκής υλοποίησης και του αλγορίθμου MDP. Η αξιολόγηση βασίζεται τόσο σε ποιοτικά χαρακτηριστικά όσο και σε ποσοτικούς δείκτες με έμφαση στις επίσημες μετρικές CLEAR του αντίστοιχου διαγωνισμού MOT Challenge. Συνολικά ο αλγόριθμος που αναπτύχθηκε εμφανίζει σχετικά ικανοποιητικά αποτελέσματα δεδομένης της δυσκολίας του προβλήματος και του στόχου της εργασίας, που δεν είναι άλλος παρά την εμβάθυνση σε αυτό το πεδίο και όχι την ανάπτυξη ενός αλγορίθμου αιχμής. Τέλος, προκύπτουν ορισμένα συμπεράσματα τόσο όσον αφορά τη συμπεριφορά του αναπτυχθέντα αλγορίθμου όσο και περί των μελλοντικών κατευθύνσεων που φαίνεται ότι πρόκειται να ακολουθήσει η κοινότητα.

ABSTRACT

Master Thesis

"Development and validation of a methodology for multiple object tracking"

Athena V. Psalta

National Technical University of Athens

School of Applied Mathematical and Physical Sciences

Visual tracking is a heavily studied problem in Computer Vision main due to is numerous and diverse applications. Both single and multiple object tracking are complex procedures and as a result the state-of-the-art algorithms do not completely address these problems. In the current thesis, the relative literature is exhaustively studied and a prototype multiple object tracking algorithm is developed.

In the first part of the thesis, the problem of single object tracking is discussed via addressing the nature of the problem and a selection of modern approaches are presented. Next, the problem of multiple object tracking follows and its main differences concerning both the methodologies and difficulties are stressed out. Moreover, the trend of employing Neural Networks, such as RNN, LSTM and CNN, for solving the multiple object tracking problem is emphasized and the state-of-the-art algorithm MDP (Markov Decision Processes) is analyzed in depth.

The second part of the thesis is dedicated to the design, development and validation of a prototype multiple object tracking algorithm. First, a "tracking-by-detection" approach is used and multiple cues are co-evaluated, such as the appearance model, the motion model and the spatial interaction model. Detections are assigned to tracks using a complex stepwise data association formulation of the problem that is based on the Hungarian Algorithm.

The tracking algorithm that was developed is tested and evaluated on the 2DMOT15 dataset of MOT Challenge and its results are compared with the respective ones of MDP algorithm and another simplistic approach. The performance of all the above algorithms is evaluated using CLEAR metrics and qualitative criteria. All in all, the overall performance of the developed tracking algorithm is acceptable given the difficulty of the problem and the goals of the thesis, which are no other than an early introduction to this field. Finally, certain conclusions for the algorithms that were tested and some future directions are presented.

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

ΚΕΦΑΛΑΙΟ 1 : ΕΙΣΑΓΩΓΗ

1.1. Πλαίσιο της εργασίας

Τα τελευταία χρόνια η ραγδαία ανάπτυξη της τεχνολογίας έχει συμβάλλει στην διαρκή παρουσία των φωτογραφικών μηχανών στην καθημερινότητα των ανθρώπων λόγω των ολοένα και αυξανόμενων δυνατοτήτων τους σε συνδυασμό με τη χαμηλή τιμή τους. Γι' αυτό το λόγο η **Όραση Υπολογιστών** (Computer Vision), στόχος της οποίας είναι η μοντελοποίηση και η ανάπτυξη αλγορίθμων που επιτρέπουν στους υπολογιστές να κατανοούν ένα σύνολο εικόνων που απεικονίζουν τον πραγματικό κόσμο, έχει γίνει ένα ιδιαίτερα δημοφιλές ερευνητικό πεδίο.

Ορισμένα ερευνητικά πεδία της Όρασης Υπολογιστών είτε σχετικά απλά, όπως η αφαίρεση θορύβου σε μία εικόνα, είτε πιο πολύπλοκα, όπως η αναγνώριση προσώπου, υπάρχουν στη καθημερινότητα του ανθρώπου με άμεσο ή έμμεσο τρόπο. Παρ' όλα αυτά, η δυνατότητα κατανόησης του περιβάλλοντος χώρου και ερμηνείας της οπτικής πληροφορίας που διαθέτει ο άνθρωπος εξακολουθεί να είναι ανώτερη από αυτή που προσφέρεται από τα υφιστάμενα συστήματα. Τέτοιες εφαρμογές υψηλού επιπέδου ως προς τη πολυπλοκότητά τους αποτελούν η αναγνώριση αντικειμένων (object detection), η σημασιολογική κατάτμηση (semantic segmentation) και η ταξινόμηση μιας εικόνας (image classification). Αυτή η μεταπτυχιακή εργασία ασχολείται με ένα επίσης πολύπλοκο πρόβλημα αυτού του επιστημονικού πεδίου που αποτελεί η παρακολούθηση των κινούμενων αντικειμένων σε μία δυναμική σκηνή (βίντεο), μία διαδικασία η οποία συχνά αναφέρεται ως **παρακολούθηση πολλαπλών αντικειμένων** (multiple object tracking).

Η παρακολούθηση πολλαπλών αντικειμένων κατά τη διάρκεια ενός βίντεο δεν υποδεικνύει απλώς την τοποθεσία τους σε κάθε χρονικό βήμα, αλλά δίνει επίσης την δυνατότητα της *πλήρους ανακατασκευής των τροχιών τους*. Βάσει αυτής της πληροφορίας μπορούν να αναλυθούν η δυναμική συμπεριφορά κάθε στόχου αλλά και των όποιων αλληλεπιδράσεων που μπορεί να υφίστανται μεταξύ των στόχων, καθώς και υπάρχει η δυνατότητα πρόβλεψης της μελλοντικής συμπεριφοράς των στόχων. Σε αυτό το σημείο θα σχολιαστεί εν συντομία το **κίνητρο** των ερευνητών για την ανάπτυξη ολοκληρωμένων συστημάτων παρακολούθησης πολλαπλών αντικειμένων και κατά πόσο ένα τέτοιο σύστημα μπορεί να συσχετιστεί με την επιστήμη, το χώρο του θεάματος και την καθημερινή ζωή των ανθρώπων. Μερικές εφαρμογές, λοιπόν, της οπτικής παρακολούθησης είναι οι εξής :

- **Εφαρμογές επιτήρησης και οπτικής παρακολούθησης (surveillance)** : Βάζοντας στην άκρη όλα τα θέματα ιδιωτικότητας και τα κοινωνικά και ηθικά ζητήματα που ανακύπτουν με την παρακολούθηση του πληθυσμού από τις κάμερες CCTV, είναι προφανές ότι η «χειροκίνητη» παρακολούθηση ενός πλήθους ανθρώπων είναι μια εξαιρετικά δύσκολη διαδικασία. Επομένως, η ανάγκη για

την ανάπτυξη αξιόπιστων συστημάτων παρακολούθησης για ένα πλήθος στόχων (ανθρώπους, ζώα, οχήματα κλπ.) μεγαλώνει ολοένα και περισσότερο. Ωστόσο, αξίζει να αναφερθεί ότι ένα τέτοιο σύστημα είναι ιδιαίτερα χρήσιμο και σε περιπτώσεις μη πραγματικού χρόνου (offline), διότι μέσω μιας τέτοιας ανάλυσης προκύπτουν οι ανακατασκευασμένες τροχιές των αντικειμένων. Μία τέτοια πληροφορία καθίσταται, για παράδειγμα, πολύ χρήσιμη για περιπτώσεις σχεδιασμού καλύτερης χωροθέτησης των εξόδων κινδύνου, αφού μπορούν να ανακαλυφθούν τα μέρη όπου τα πλήθη των ανθρώπων συσσωρεύονται. Επίσης, ένα τέτοιο σύστημα μπορεί να βοηθήσει από διαφημιστικής απόψεως αφού μπορεί να αναλυθεί η συμπεριφορά των ανθρώπων μέσα σε ένα κατάστημα ή σε έναν εμπορικό δρόμο.

- **Εφαρμογές στη ρομποτική (robotics)** : Η πλοήγηση των αυτόνομων ρομπότ σε ένα στατικό περιβάλλον αντιμετωπίζεται σχετικά εύκολα με τον εντοπισμό των εμποδίων που μπορεί να αντιμετωπίσει. Ωστόσο, η πλοήγηση σε ένα μεταβαλλόμενο περιβάλλον επιβάλει την πλήρη γνώση της κίνησης των αντικειμένων γύρω από το ρομπότ προκειμένου αυτό να μπορεί να αποφύγει τα εμπόδια που του εμφανίζονται. Έτσι, λοιπόν, η ανάγκη ανάπτυξης συστημάτων παρακολούθησης πολλαπλών αντικειμένων για την πλοήγηση των αυτόνομων ρομπότ καθίσταται απαραίτητη για τη περίπτωση πραγματικών συνθηκών.
- **Εφαρμογές στον αθλητισμό** : Η παρακολούθηση της τροχιάς των αθλητών σε ένα παιχνίδι παρέχει χρήσιμες πληροφορίες για τη φυσική απόδοση του καθενός από τους παίκτες και για την στρατηγική της ομάδας. Σε αυτό τον τομέα έχουν ήδη αναπτυχθεί αρκετά συστήματα πολλαπλής παρακολούθησης, όμως η ολοένα και αυξανόμενη ανάγκη της βιομηχανίας του αθλητισμού για όσο το δυνατόν ακριβέστερη πληροφορία «γεννά» προκλήσεις για τους ερευνητές.
- **Εφαρμογές επαυξημένης πραγματικότητας (Augmented Reality)** : Οι υφιστάμενες εφαρμογές επαυξημένης πραγματικότητας συχνά περιορίζονται σε αντικείμενα που μπορούν να μοντελοποιηθούν εκ των προτέρων. Ο πυρήνας, θα λέγαμε, αυτών των εφαρμογών είναι ένα εύρωστο σύστημα παρακολούθησης των στόχων που συνήθως εξαρτάται από ένα στάδιο εκπαίδευσης σε μη πραγματικό χρόνο (offline). Συνεπώς, η δυνατότητα ενός συστήματος παρακολούθησης πολλαπλών αντικειμένων που θα είναι εύρωστο ακόμη και στη περίπτωση στόχων με τυχαίο σχήμα χωρίς να έχει προηγηθεί κάποιο στάδιο εκπαίδευσης χαίρει μεγάλου ενδιαφέροντος με πιθανές εφαρμογές στη gaming βιομηχανία, στην διαφήμιση, στον τουρισμό και σε ιατρικές εφαρμογές.
- **Εφαρμογές ασφάλειας των οδηγών (road safety)** : Παρά τη δυνατότητα εξοπλισμού των σύγχρονων αυτοκινήτων με high-end αισθητήρες για την υποβοήθηση του οδηγού σε επικίνδυνες καταστάσεις, προκειμένου να καταστεί επιτυχημένη η πλοήγηση ενός οχήματος απαιτείται η παρακολούθηση όλων των κοντινών του αντικειμένων. Αυτό είναι ένα πρόβλημα που θα μπορούσε να προσεγγιστεί με την τοποθέτηση μίας ή πολλών καμερών πάνω σε ένα όχημα που

θα παρακολουθεί τα κοντινά αντικείμενα και θα συνδέεται με ένα σύστημα υποβοήθησης του οδηγού που θα του παρέχει πληροφορίες για τα κοντινά οχήματα και τους πεζούς. Ακόμη, ένα σύστημα παρακολούθησης των αυτοκινήτων σε κάποιο δρόμο μπορεί να παρέχει χρήσιμες πληροφορίες για την όποια ένδειξη επικίνδυνης συμπεριφοράς από κάποιο από αυτά.

- **Βιοεπιστήμες** : Η αξιόπιστη παρακολούθηση ενός πλήθους ζώων είναι αρκετά σημαντική, αφού ένα τέτοιο σύστημα μπορεί να παρέχει σημαντικές πληροφορίες για την κοινωνική τους συμπεριφορά ως ένα σύνολο. Αυτό το αντικείμενο είναι ιδιαίτερα δύσκολο ερευνητικά αφού στη περίπτωση παρακολούθησης ενός είδους ζώου η εμφάνιση των περισσοτέρων είναι πανομοιότυπη. Ως αποτέλεσμα, τα συστήματα παρακολούθησης που αναπτύσσονται γι' αυτό τον σκοπό πρέπει να βασίζονται κυρίως σε δυναμικά μοντέλα προκειμένου να διαχωρίζονται οι στόχοι μεταξύ τους. Επιπρόσθετα, στον τομέα της μικροβιολογίας τα συστήματα παρακολούθησης πολλαπλών αντικειμένων σε μοριακό επίπεδο χαίρουν ιδιαίτερου ερευνητικού ενδιαφέροντος. Η μελέτη της δυναμικής κίνησης των διάφορων σωματιδίων είναι εξαιρετικά κρίσιμη για την ανακάλυψη νέων φαρμάκων, όμως η ανάπτυξη ενός συστήματος παρακολούθησης για τέτοιους λόγους έχει αρκετές δυσκολίες. Η μη δυνατότητα υιοθέτησης αλγορίθμων που χρησιμοποιούνται σε εφαρμογές του πραγματικού κόσμου, η κακή - σε γενικές γραμμές- ποιότητα των βίντεο που παρέχονται από τα μικροσκόπια και η ομοιότητα αρκετών όμοιων αντικειμένων προς παρακολούθηση είναι μερικές μόνο από τις προκλήσεις που καλείται ένα σύστημα παρακολούθησης να αντιμετωπίσει. Συνεπώς, γίνεται αντιληπτή η ανάγκη του τομέα των βιοεπιστημών για την ανάπτυξη τέτοιων συστημάτων παρακολούθησης.

Εικόνα 1.1 : Παραδείγματα εφαρμογών της παρακολούθησης πολλαπλών αντικειμένων. (α) αποφυγή ατυχημάτων, (β) παρακολούθηση επικίνδυνης συμπεριφοράς από CCTV κάμερες, (γ) ρομποτική, (δ)-(ε) μελέτη της συμπεριφοράς των ζώων, (ζ) παρακολούθηση μικροοργανισμών

Εικόνα 1.2 : Παράδειγμα αλγορίθμου παρακολούθησης ενός πλήθους ανθρώπων που χρησιμοποιείται για τον εντοπισμό και την παρακολούθηση εγκαταλελειμμένων και πιθανόν επικίνδυνων αντικειμένων.

Εικόνα 1.3 : Παράδειγμα χρήσης ενός αλγορίθμου παρακολούθησης για διαφημιστικούς σκοπούς που στοχεύει στην εκτίμηση της οπτικής παρατήρησης των δύο ανθρώπων που βρίσκονται μπροστά σε μία βιτρίνα ενός καταστήματος.

1.2. Σκοπός της εργασίας

Η παρούσα μεταπτυχιακή εργασία επικεντρώνεται στην ανάπτυξη μεθοδολογίας για την παρακολούθηση πολλαπλών αντικειμένων (πχ ανθρώπων) σε ένα βίντεο. Αναλύεται το θεωρητικό υπόβαθρο και οι τεχνολογίες αιχμής, ενώ περιγράφονται αναλυτικά οι δυνατότητες της μεθοδολογίας, οι παράμετροί του και περιορισμοί του. Επίσης, περιγράφονται τα αποτελέσματα της διαδικασίας αξιολόγησης της μεθοδολογίας και η σύγκριση τους με τα αντίστοιχα άλλων αλγορίθμων παρακολούθησης που συναντώνται στη βιβλιογραφία ποσοτικά και ποιοτικά.

Κίνητρο της εργασίας αποτέλεσε το έντονο ερευνητικό ενδιαφέρον της επιστημονικής κοινότητας στο συγκεκριμένο πρόβλημα της οπτικής παρακολούθησης. Συγκεκριμένα, η παρακολούθηση πολλαπλών αντικειμένων αποτελεί ένα ανοιχτό ερευνητικά ζήτημα, αφού λόγω της πολυπλοκότητάς του ακόμη και πρόσφατοι αλγόριθμοι στην αιχμή της τεχνολογίας (state-of-the-art) δεν λύνουν επαρκώς το πρόβλημα. Οι πολλές εφαρμογές ενός τέτοιου συστήματος αποτέλεσαν, επίσης, ισχυρό κίνητρο για την μελέτη του συγκεκριμένου προβλήματος.

Συνοψίζοντας, βασική επιδίωξη αποτέλεσε η μελέτη του προβλήματος της οπτικής παρακολούθησης και ο σχεδιασμός, ανάπτυξη και αξιολόγηση μεθοδολογίας για την οπτική παρακολούθηση η οποία να αξιοποιεί εξειδικευμένα μοντέλα εμφάνισης.

1.3. Δομή της εργασίας

Η παρούσα εργασία αποτελείται από έξι κεφάλαια. Στο πρώτο (παρών) κεφάλαιο παρουσιάζεται το πλαίσιο στο οποίο βασίστηκε η εργασία τονίζοντας το κίνητρο των ερευνητών για την επίλυση του προβλήματος της παρακολούθησης πολλαπλών αντικειμένων, καθώς και αναφέρεται ο στόχος αυτής της εργασίας.

Το δεύτερο κεφάλαιο αναφέρεται στην οπτική παρακολούθηση ενός αντικειμένου. Αρχικά, εστιάζεται στις δυσκολίες αυτού του προβλήματος και στη συνέχεια αναφέρονται οι διάφορες κατηγοριοποιήσεις τέτοιων αλγορίθμων παρακολούθησης που συναντώνται και τέλος παρουσιάζεται μια συνοπτική αναφορά σε ορισμένους αλγορίθμους της βιβλιογραφίας.

Το τρίτο κεφάλαιο αφορά το πρόβλημα της παρακολούθησης πολλαπλών αντικειμένων. Σε πρώτη φάση, αναλύονται οι δυσκολίες αυτού του προβλήματος και οι διαφορές του με την παρακολούθηση ενός αντικειμένου. Έπειτα, πραγματοποιείται μία συνοπτική αναφορά σε διάφορους αλγορίθμους αιχμής της βιβλιογραφίας και αναλύεται εκτενώς ο αλγόριθμος παρακολούθησης MDP.

Στο τέταρτο κεφάλαιο περιγράφεται η μεθοδολογική προσέγγιση του πρωτότυπου αλγορίθμου παρακολούθησης πολλαπλών αντικειμένων που αναπτύχθηκε στη παρούσα εργασία. Αρχικά, παρουσιάζεται η κύρια δομή του αλγορίθμου και τα στάδια προεπεξεργασίας που απαιτούνται. Στη συνέχεια, αναλύονται ενδελεχώς ο τρόπος υπολογισμού των κοστών αντιστοίχισης και η διαδικασία αντιστοίχισης των δεδομένων.

Στο πέμπτο κεφάλαιο παρουσιάζονται τα πειραματικά αποτελέσματα του αλγορίθμου που αναπτύχθηκε σε σύγκριση με άλλους δύο αλγορίθμους παρακολούθησης της βιβλιογραφίας. Ακολουθεί η αξιολόγηση του αλγορίθμου τόσο με ποσοτικά όσο και ποιοτικά κριτήρια.

Το έκτο κεφάλαιο της εργασίας αφορά τα γενικά συμπεράσματα που εξήχθησαν για τον αλγόριθμο παρακολούθησης που αναπτύχθηκε, αλλά και σχολιάστηκαν οι μελλοντικές κατευθύνσεις που φαίνεται ότι η επιστημονική κοινότητα πρόκειται να ακολουθήσει για την επίλυση του προβλήματος της παρακολούθησης πολλαπλών αντικειμένων.

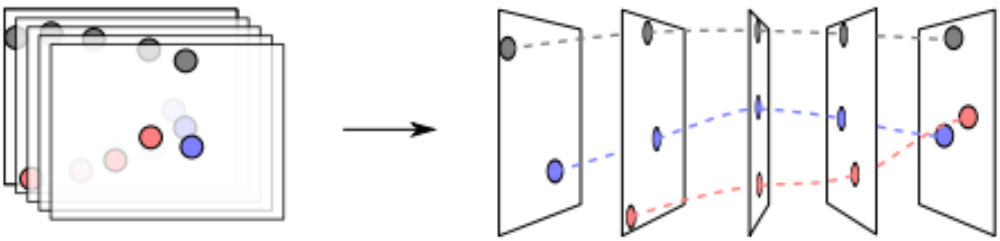
Στο παράρτημα της εργασίας αναλύεται συνοπτικά ο αλγόριθμος αντιστοίχισης του Munkres (Hungarian Algorithm) που χρησιμοποιήθηκε για τη διαδικασία αντιστοίχισης στόχων και ανιχνεύσεων στον αλγόριθμο παρακολούθησης που αναπτύχθηκε.

ΚΕΦΑΛΑΙΟ 2 : ΟΠΤΙΚΗ ΠΑΡΑΚΟΛΟΥΘΗΣΗ ΕΝΟΣ ΑΝΤΙΚΕΙΜΕΝΟΥ

Σε αυτό το κεφάλαιο πρόκειται να περιγραφεί αναλυτικά το πρόβλημα της παρακολούθησης ενός αντικειμένου, σχολιάζοντας τη δυσκολία του προβλήματος που εξαρτάται από ένα σύνολο παραγόντων. Στη συνέχεια, θα αναφερθούν διάφορες κατηγοριοποιήσεις αλγορίθμων που συναντώνται στη διεθνή βιβλιογραφία και θα περιγραφεί συνοπτικά ο state-of-the-art αλγόριθμος TLD.

2.1. Το πρόβλημα της παρακολούθησης ενός αντικειμένου

Ο όρος **οπτική παρακολούθηση (visual tracking)** αναφέρεται στη διαδικασία του χωρικού εντοπισμού ενός αντικειμένου ενδιαφέροντος σε μία ακολουθία εικόνων. Οι αλγόριθμοι που ασχολούνται με την επίλυση αυτού του προβλήματος συνήθως ακολουθούν δύο βασικές αρχές. Η πρώτη αναφέρεται στην ακριβή γνώση της αρχικής θέσης του στόχου, η οποία παρέχεται από τον χρήστη στο πρώτο καρέ του βίντεο, ενώ η δεύτερη αφορά το γεγονός ότι μόνο ένας στόχος παρακολουθείται κατά τη διάρκεια του βίντεο. Επιπρόσθετα, αξίζει να αναφερθεί ότι μέσω αυτής της διαδικασίας υπάρχει η δυνατότητα της πλήρους ανακατασκευής των τροχιών των στόχων και όχι απλά η γνώση της θέσης του καθενός από αυτούς σε κάθε καρέ ενός βίντεο. Έτσι, λοιπόν, μπορεί να υπολογιστεί η μελλοντική συμπεριφορά ενός αντικειμένου προβλέποντας βάσει της προηγούμενης συμπεριφοράς του, καθώς και να εξεταστούν οι όποιες αλληλεπιδράσεις υφίστανται μεταξύ των διαφορετικών αντικειμένων.



Εικόνα 2.1 : Σχηματική απεικόνιση της παρακολούθησης πολλαπλών στόχων. Δεδομένης μιας ακολουθίας εικόνων, στόχος αποτελεί η ανακατασκευή όλων των τροχιών των στόχων.

Ένα **ολοκληρωμένο σύστημα παρακολούθησης** ενός αντικειμένου απαιτεί αρχικά την ανίχνευση του στόχου, στη συνέχεια την παρακολούθηση σε πραγματικό χρόνο ή μη (online/offline) του στόχου σε κάθε καρέ του βίντεο και σε μερικές περιπτώσεις την ερμηνεία της τροχιάς του στόχου για την εξαγωγή συμπερασμάτων για την εκάστοτε εφαρμογή. Η παρακολούθηση ενός στόχου μπορεί να επιτευχθεί αφενός με στοχαστικό τρόπο μέσω της δυναμικής περιγραφής του αντικειμένου στον χώρο, αφετέρου με ντετερμινιστικό τρόπο μέσω της θέσπισης ορισμένων ευριστικών κανονισμών που περιορίζουν την κίνηση του αντικειμένου στο χώρο.

Το πρόβλημα της παρακολούθησης ενός αντικειμένου θεωρείται ένα δύσκολο πρόβλημα της Όρασης Υπολογιστών, γεγονός το οποίο μπορεί να προκύπτει από ένα σύνολο παραγόντων όπως :

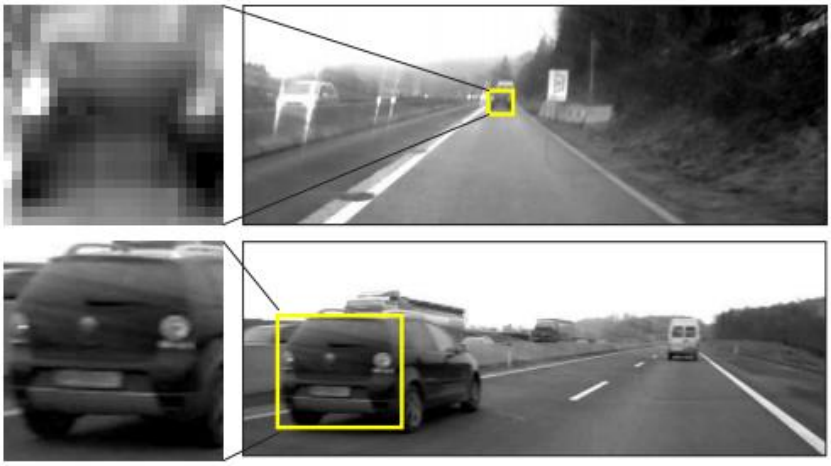
- **Αποκρύψεις του αντικειμένου προς παρακολούθηση** : Κατά τη διάρκεια ενός βίντεο το αντικείμενο μπορεί να αποκρύπτεται πλήρως ή μερικώς από άλλα αντικείμενα του υποβάθρου ως προς την οπτική της κάμερας, γεγονός το οποίο μπορεί να επηρεάσει τη διαδικασία παρακολούθησης.
- **Απαιτήσεις για παρακολούθηση σε πραγματικό χρόνο** : Η ανάπτυξη ενός αλγορίθμου με υψηλή χρονική αποδοτικότητα (επεξεργασία κάθε καρέ σε χρόνο το πολύ 1/30 sec) συνεπάγεται απλοποιήσεις στη πολυπλοκότητα του αλγορίθμου διατηρώντας παράλληλα την ακρίβεια της παρακολούθησης και πιθανώς εξειδικευμένο υλικό εξοπλισμό.
- **Αλλαγές φωτισμού της σκηνής** : Είτε κάτι τέτοιο συμβαίνει σε εξωτερικό ή εσωτερικό χώρο, η απεικόνιση του υποβάθρου επηρεάζεται σε μεγάλο βαθμό συνεπώς μπορεί να οδηγήσει σε λανθασμένους εντοπισμούς του αντικειμένου άρα και σε εσφαλμένη παρακολούθηση του στόχου.
- **Δυναμικό υπόβαθρο** : Ορισμένα μέρη του υποβάθρου της σκηνής μπορεί να εμπεριέχουν κίνηση, όπως για παράδειγμα η κίνηση των σύννεφων του ουρανού, των κλαδιών ενός δέντρου, του νερού από ένα συντριβάνι, όμως εκείνα πρέπει να θεωρηθούν ως υπόβαθρο και όχι ως ανιχνεύσεις του αντικειμένου προς παρακολούθηση. Τέτοιες κινήσεις μπορεί να είναι περιοδικές αλλά και ακανόνιστες (όπως τα φανάρια σε έναν δρόμο, η κίνηση των φύλλων των δέντρων), οπότε η δυναμική μοντελοποίηση αυτών των κινήσεων μπορεί να αποδειχθεί αρκετά απαιτητική.
- **Παρουσία σκιών** : Οι σκιές που μπορούν να δημιουργηθούν από τα κινούμενα αντικείμενα περιπλέκουν τη διαδικασία υπολογισμού του υποβάθρου, άρα και της διαδικασίας παρακολούθησης ενός αντικειμένου.
- **Σχετική κίνηση της κάμερας** : Στη περίπτωση ενός βίντεο που έχει ληφθεί από μία ασταθή κάμερα, η μερική κίνηση της σκηνής από το ένα καρέ στο άλλο δημιουργεί προβλήματα στη παρακολούθηση του αντικειμένου.
- **Παρουσία θορύβου στο βίντεο** : Κάτι τέτοιο δυσκολεύει αρκετά την εξαγωγή εύρωστων χαρακτηριστικών σε ένα σύστημα οπτικής παρακολούθησης.

- **Μεταβολές στην ταχύτητα του αντικειμένου** : Η ταχύτητα της κίνησης ενός αντικειμένου παίζει σημαντικό ρόλο στον εντοπισμό και την παρακολούθησή του. Η θεώρηση σύνθετων κινηματικών μοντέλων είναι αναγκαία για την ανάπτυξη ενός εύρωστου αλγόριθμου παρακολούθησης αφού είτε πρόκειται για ένα άκαμπτο αντικείμενο είτε για ένα παραμορφώσιμο οι απότομες μεταβολές στην ταχύτητά του δυσκολεύουν τη διαδικασία συνταύτισης χαρακτηριστικών.

Εικόνα 2.2 : Παράδειγμα δυσκολίας παρακολούθησης ενός κινούμενου αντικειμένου στη περίπτωση που υπάρχει απόκρυψή ή εξαφάνισή του από τη σκηνή.

Εικόνα 2.3 : Αλλαγές στην εμφάνιση των αντικειμένων ανάλογα με τον φωτισμό της σκηνής. Η επίδραση των σκιάσεων και του εναλλασσόμενου φωτισμού της σκηνής δυσκολεύει την εύρωστη μοντελοποίηση των αντικειμένων.

Εικόνα 2.4 : Παράδειγμα μερικής μεταβολής της κίνησης του αντικειμένου



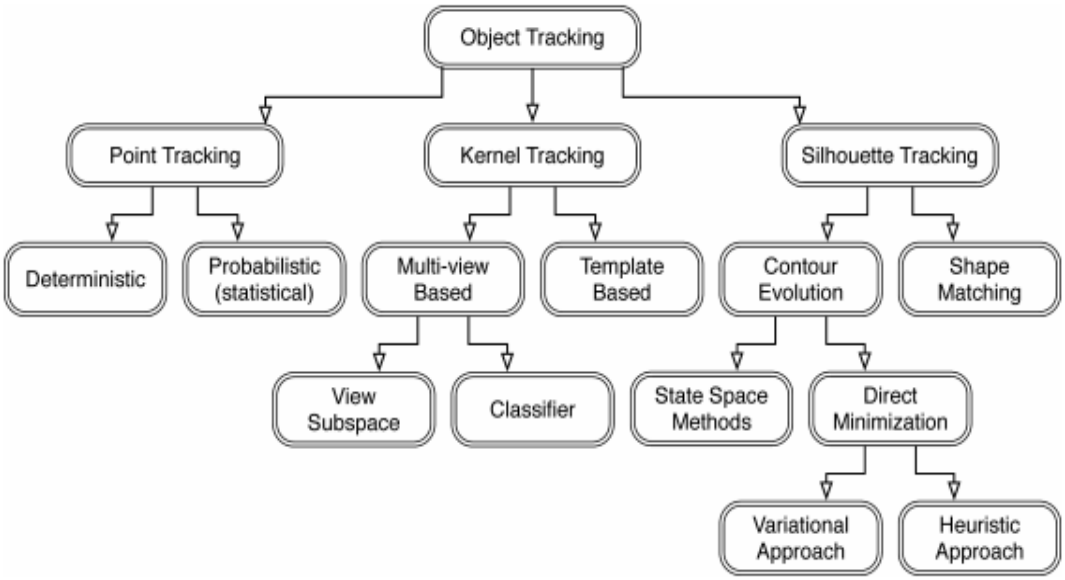
Εικόνα 2.5 : Παράδειγμα εναλλαγής της κλίμακας του παρακολουθούμενου αντικειμένου

Εικόνα 2.6 : Παράδειγμα αλλαγής της εμφάνισης του αντικειμένου προς παρακολούθηση από διαφορετικές οπτικές γωνίες.

2.2. Ανασκόπηση της βιβλιογραφίας

Σε αυτό το σημείο θα αναφερθούν οι διάφορες κατηγοριοποιήσεις αλγορίθμων παρακολούθησης ενός αντικειμένου που συναντώνται και στη συνέχεια θα γίνει μια αναφορά σε ορισμένους αλγορίθμους της βιβλιογραφίας.

Εξαιρετική σημασία για ένα σύστημα παρακολούθησης αποτελεί η θεωρούμενη κατηγορία αναπαράστασης του εκάστοτε αντικειμένου, αφού είναι εκείνη που περιορίζει τον δυναμικό του χαρακτήρα. Στη διεθνή βιβλιογραφία συναντώνται τρεις κατηγορίες παρακολούθησης. Η πρώτη αναφέρεται στη **παρακολούθηση σημείων** (point tracking) όπου τα αντικείμενα που εντοπίζονται σε διαδοχικά καρέ αναπαρίστανται από σημεία και η συσχέτιση αυτών βασίζεται στη προηγούμενη κατάσταση του αντικειμένου όσον αφορά τη θέση και την κίνηση. Οι αλγόριθμοι που βασίζονται σε αυτή τη προσέγγιση χωρίζονται στις *ντετερμινιστικές μεθόδους* όπου χρησιμοποιούνται ποιοτικοί ευριστικοί κανόνες για τον περιορισμό της κίνησης του αντικειμένου και στις *στοχαστικές μεθόδους* που λαμβάνουν υπόψη και την αβεβαιότητα της κίνησης του αντικειμένου. Η δεύτερη κατηγορία αναπαράστασης αναφέρεται στην **παρακολούθηση πυρήνα** (kernel tracking), όπου με τον όρο αυτό εννοείται η εμφάνιση και το σχήμα του αντικειμένου. Στους αλγορίθμους αυτής της κατηγορίας πραγματοποιείται η εκτίμηση της κίνησης του αντικειμένου κατά την αλληλουχία εικόνων είτε μέσω υπολογισμού της οπτικής ροής είτε μέσω παραμετρικών μετασχηματισμών (π.χ. μετάθεση, στροφή, αφινικός μετασχηματισμός). Οι αλγόριθμοι αυτοί διαφέρουν ανάλογα με την επιλογή της αναπαράστασης εμφάνισης του αντικειμένου (template based), τον αριθμό των αντικειμένων προς παρακολούθηση (multi-view based) και τη μέθοδο που χρησιμοποιείται για την εκτίμηση της κίνησης του αντικειμένου. Η τελευταία κατηγορία αναπαράστασης αφορά την **παρακολούθηση σιλουέτας** ή αλλιώς **περιγράμματος** (silhouette tracking) κατά την οποία απαιτείται η ακριβής περιγραφή της περιοχής του επιπέδου της εικόνας για το αντικείμενο προς παρακολούθηση. Περίπλοκα αντικείμενα όπως το ανθρώπινο σώμα μπορούν να περιγραφούν με ακρίβεια σε μία αλληλουχία εικόνων μέσω αυτών των μεθόδων. Οι δύο πιο συνήθεις πρακτικές για την περιγραφή του περιγράμματος ενός αντικειμένου αποτελεί η θεώρηση κάποιου μοντέλου *ενεργών καμπυλών* (active contours) και η *ταύτιση σχήματος* (shape matching). Στο σχήμα που ακολουθεί [33] παρουσιάζεται συγκεντρωτικά η ταξινόμηση των μεθόδων παρακολούθησης ενός αντικειμένου.



Εικόνα 2.7 : Κατηγοριοποίηση των μεθόδων παρακολούθησης βάσει της αναπαράστασης ενός αντικειμένου [33]

Επιπρόσθετα, οι αλγόριθμοι παρακολούθησης που συναντώνται στη διεθνή βιβλιογραφία μπορούν να κατηγοριοποιηθούν σε αυτούς που χρησιμοποιούν έναν **αλγόριθμο ανίχνευσης** και σε αυτούς που χρησιμοποιούν **άλλες τεχνικές** για να εντοπίσουν το κινούμενο αντικείμενο. Όσον αφορά τους τελευταίους, τέτοιες τεχνικές μπορεί να είναι ο υπολογισμός της οπτικής ροής και η εύρεση του παρασκηνίου (background) και του προσκηνίου (foreground) της εικόνας. Για παράδειγμα, στην Εικόνα 2.8 παρουσιάζονται οι ανιχνεύσεις που προκύπτουν από έναν αλγόριθμο ανίχνευσης και εκείνες που προκύπτουν μετά από εκτίμηση της οπτικής ροής και παρασκηνίου της εικόνας. Ωστόσο, οι περισσότεροι αλγόριθμοι χρησιμοποιούν τη λογική της "παρακολούθησης μέσω ανιχνεύσεων" (tracking-by-detection), σύμφωνα με την οποία σε πρώτη φάση οι ανιχνεύσεις που παρέχονται ως δεδομένα εισόδου αξιοποιούνται για την εξαγωγή ορισμένων χαρακτηριστικών ώστε να μοντελοποιηθεί το κινούμενο αντικείμενο. Σε επόμενο βήμα, αντιστοιχίζεται η αναπαράσταση του αντικειμένου με τις ανιχνεύσεις του επόμενου καρέ και ανανεώνεται το μοντέλο αναπαράστασης του αντικειμένου.

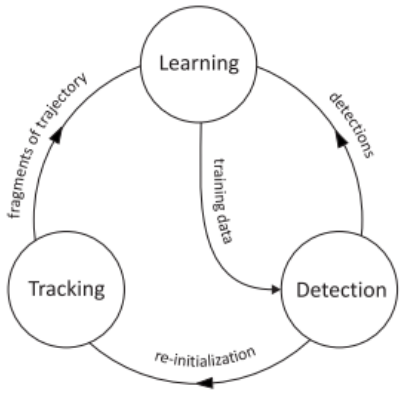
Εικόνα 2.8 : Διαφορετικές προσεγγίσεις για την ανίχνευση των κινούμενων αντικειμένων

Σε γενικές γραμμές, στην επιστημονική κοινότητα παρατηρείται μεγάλη ποικιλία στον τρόπο προσέγγισης του προβλήματος της παρακολούθησης ενός αντικειμένου ανάλογα με τον τρόπο αναπαράστασης του κινούμενου αντικειμένου, με τη χρήση ή όχι ενός αλγορίθμου ανίχνευσης και την εφαρμογή. Στον Πίνακα 2.1 παρουσιάζονται διάφοροι αλγόριθμοι παρακολούθησης ενός αντικειμένου που συναντώνται στη διεθνή βιβλιογραφία. Επίσης, αναφέρονται ο τρόπος αναπαράστασης του αντικειμένου και η μέθοδος πρόβλεψης της κίνησης του στόχου που χρησιμοποιούν, καθώς και η αποδοτικότητα του αλγορίθμου σε ταχύτητα (καρέ ανά δευτερόλεπτο).

Αλγόριθμος	Αναπαράσταση	Μέθοδος πρόβλεψης	FPS
CRF [38]	L, IH	Particle filter	109
LOT [39]	L, χρώμα	Particle filter	0,7
IVT [40]	H, PCA, GM	Particle filter	33,4
VTS [41]	L, SPCA, GM	Markov Chain Monte Carlo	5,7
ORIA [42]	H, T, GM	Local optimum search	9
OAB [43]	H, Haar, DM	Dense sampling search	22,4
TLD [37]	L, BP, DM	Dense sampling search	28,1

Πίνακας 2.1 : Διάφοροι αλγόριθμοι παρακολούθησης ενός αντικειμένου [44], όπου L τοπική αναπαράσταση, IH ιστόγραμμα, H ολιστική, PCA η Ανάλυση Κυρίων Συνιστωσών, GM γενετικό μοντέλο, DM διαχωριστικό (discriminative) μοντέλο και BP δυαδικό πρότυπο.

Ένας αρκετά δημοφιλής αλγόριθμος παρακολούθησης ενός αντικειμένου της διεθνούς βιβλιογραφίας είναι ο **TLD** [37] που ακολουθεί μία λογική παρακολούθησης μέσω ανίχνευσης. Πρόκειται για ένα ολοκληρωμένο σύστημα παρακολούθησης που ταυτόχρονα αναλαμβάνει την παρακολούθηση, την ανίχνευση και την εκμάθηση του μοντέλου κίνησης και εμφάνισης του στόχου που έχει επιλεχθεί αρχικά χειροκίνητα από τον χρήστη. Αυτός ο αλγόριθμος έχει αρκετά καλές επιδόσεις αφού μπορεί να λειτουργεί σε πραγματικό χρόνο, να αντιμετωπίζει «δύσκολα» βίντεο όπου η διαδικασία παρακολούθησης συχνά αποτυγχάνει με άλλους αλγορίθμους και να υποβοηθάει τη διαδικασία ανίχνευσης μέσω της εκμάθησης που επιτελείται. Συνοπτικά, η διαδικασία παρακολούθησης αναλαμβάνει να εκτιμά την κίνηση του αντικειμένου σε διαδοχικά καρέ δεδομένου ότι το αντικείμενο είναι ορατό και η κίνησή του είναι σχετικά μικρή, δηλαδή σε λίγα εικονοστοιχεία. Η διαδικασία ανίχνευσης αντιμετωπίζει κάθε καρέ ανεξάρτητα και πραγματοποιείται η εύρεση όλων των εμφανίσεων του αντικειμένου που έχουν υπάρξει στο παρελθόν, ενώ η διαδικασία εκμάθησης αναλαμβάνει να παρατηρεί τις επιδόσεις της ανίχνευσης και της παρακολούθησης και παράγει παραδείγματα αποφυγής των όποιων λαθών έχουν προκύψει. Συνολικά, η λειτουργία του TLD παρουσιάζεται στην Εικόνα 2.9.



Εικόνα 2.9 : Στάδια λειτουργίας του αλγορίθμου παρακολούθησης TLD [37]

ΚΕΦΑΛΑΙΟ 3 : ΟΠΤΙΚΗ ΠΑΡΑΚΟΛΟΥΘΗΣΗ ΠΟΛΛΑΠΛΩΝ ΑΝΤΙΚΕΙΜΕΝΩΝ

Η διαδικασία **παρακολούθησης πολλαπλών αντικειμένων (multi-object tracking)** σε ένα βίντεο αποτελεί ένα αρκετά πιο δύσκολο πρόβλημα σε σχέση με το πρόβλημα της παρακολούθησης ενός αντικειμένου. Συγκεκριμένα, σε ένα τέτοιο σύστημα απαιτείται η πλήρως αυτόματη λειτουργία του χωρίς αρχικοποίηση από τον χρήστη. Σε αυτή την περίπτωση αφενός ο αριθμός των στόχων είναι άγνωστος, αφετέρου το πλήθος αυτό μεταβάλλεται κατά τη διάρκεια ενός βίντεο, καθώς μερικοί νέοι στόχοι εμφανίζονται και άλλοι εξαφανίζονται από το οπτικό πεδίο του βίντεο. Σκοπός, δηλαδή, ενός τέτοιου συστήματος είναι η ακριβής επανασύσταση της τροχιάς κάθε στόχου που κινείται ελεύθερα σε ένα βίντεο, ή με άλλα λόγια ο ακριβής χωρικός εντοπισμός κάθε αντικειμένου ενδιαφέροντος σε κάθε χρονικό βήμα μιας δυναμικής απεικόνισης. Σε αυτό το εδάφιο πρόκειται να αναλυθεί περαιτέρω το πρόβλημα της παρακολούθησης πολλαπλών αντικειμένων, καθώς και θα αναφερθούν ορισμένοι αλγόριθμοι της διεθνούς βιβλιογραφίας.

3.1. Το πρόβλημα της παρακολούθησης πολλαπλών αντικειμένων

Το ζήτημα της οπτικής παρακολούθησης πολλαπλών αντικειμένων σε ένα βίντεο είναι ακόμη ένα ανοιχτό πρόβλημα ερευνητικά, αφού δεν έχει εφευρεθεί ένα αρκετά ακριβές, αποδοτικό και εύρωστο σύστημα. Σε αυτό το κεφάλαιο, λοιπόν, πρόκειται να αναλυθούν οι λόγοι για τους οποίους οι προκλήσεις που αντιμετωπίζει κανείς για την ανάπτυξη ενός ολοκληρωμένου συστήματος παρακολούθησης πολλαπλών αντικειμένων είναι πολλές και μάλιστα αρκετά περισσότερες σε σχέση με τη περίπτωση της παρακολούθησης ενός αντικειμένου.

Ένα από τα πιο βασικά προβλήματα στη παρακολούθηση πολλαπλών αντικειμένων είναι η **μοντελοποίηση** των αντικειμένων προς παρακολούθηση, δηλαδή η περιγραφή του κάθε αντικειμένου με τέτοιο τρόπο ώστε να είναι εύληπτο από έναν υπολογιστή. Η οπτική περιγραφή των αντικειμένων προς παρακολούθηση πρέπει αφενός να είναι αρκετά γενική ώστε να μπορεί να εντοπίσει αντικείμενα ίδιας κατηγορίας, αφετέρου να μπορεί να διακρίνει το κάθε αντικείμενο από το άλλο. Για παράδειγμα, στη περίπτωση της παρακολούθησης ενός πλήθους ανθρώπων σε ένα βίντεο ζητείται η μοντελοποίηση των ανθρώπων με τέτοιο τρόπο ώστε να είναι δυνατή η ταυτόχρονη παρακολούθηση ανθρώπων με διαφορετικό σχήμα, μέγεθος και χρώμα, αλλά ταυτόχρονα και η διάκριση του καθενός ξεχωριστά. Η εύρεση μιας κατάλληλης οπτικής περιγραφής των αντικειμένων προς παρακολούθηση με αυτό τον τρόπο είναι ένα αρκετά δύσκολο

πρόβλημα και οι μέθοδοι που χρησιμοποιούνται συνήθως από τους ερευνητές περιλαμβάνουν την περιγραφή χαρακτηριστικών χρώματος, υφής, κίνησης, σχήματος και υποβάθρου (background).

Η **αλλαγή της εμφάνισης** των αντικειμένων προς παρακολούθηση σε ένα βίντεο είναι ένα σύνηθες φαινόμενο και αποτελεί ένα βασικό πρόβλημα για την ανάπτυξη αλγορίθμων παρακολούθησης. Τα περισσότερα αντικείμενα σε γενικές γραμμές εμφανίζονται διαφορετικά κατά τη διάρκεια ενός βίντεο ανάλογα με τη γωνία θέασης με αποτέλεσμα να είναι δύσκολη η ακριβής περιγραφή τους. Επιπρόσθετα, η εκάστοτε αλλαγή της εμφάνισης των αντικειμένων μπορεί να προκληθεί λόγω προοπτικής, δηλαδή σχετικά με το πόσο κοντά βρίσκονται αυτά στη κάμερα. Κάτι τέτοιο μπορεί να προκαλέσει προβλήματα στη μοντελοποίηση των αντικειμένων διότι συνήθως δεν είναι γνωστή η απόσταση του αντικειμένου από την κάμερα. Εκτός από την περίπτωση δυνατότητας ανακατασκευής 3D μοντέλων των αντικειμένων, η αλλαγή της εμφάνισης των αντικειμένων, λοιπόν, είναι ένα απαιτητικό κομμάτι των αλγορίθμων που δυσκολεύει το έργο των ερευνητών.

Σημαντικά προβλήματα για την ορθή λειτουργία των αλγορίθμων παρακολούθησης πολλαπλών αντικειμένων αποτελούν οι **αλλαγές φωτισμού της σκηνής**. Αυτές οι αλλαγές επηρεάζουν την εμφάνιση των αντικειμένων αφού το καθένα έχει διαφορετικό χρώμα που ανταποκρίνεται διαφορετικά από τις συνθήκες φωτισμού της σκηνής με αποτέλεσμα να γίνονται λάθη κατά την αντιστοίχιση των στόχων με τις ανιχνεύσεις. Ακόμη, οι σκιές των κινούμενων αντικειμένων δημιουργούν σοβαρά προβλήματα στη μοντελοποίηση των αντικειμένων αφού χαρακτηριστικά όπως η κίνηση, το σχήμα και η εικόνα υποβάθρου (background) συνήθως δίνουν λανθασμένες αποκρίσεις, καθώς και μπορούν να επηρεάσουν το χρώμα του αντικειμένου.

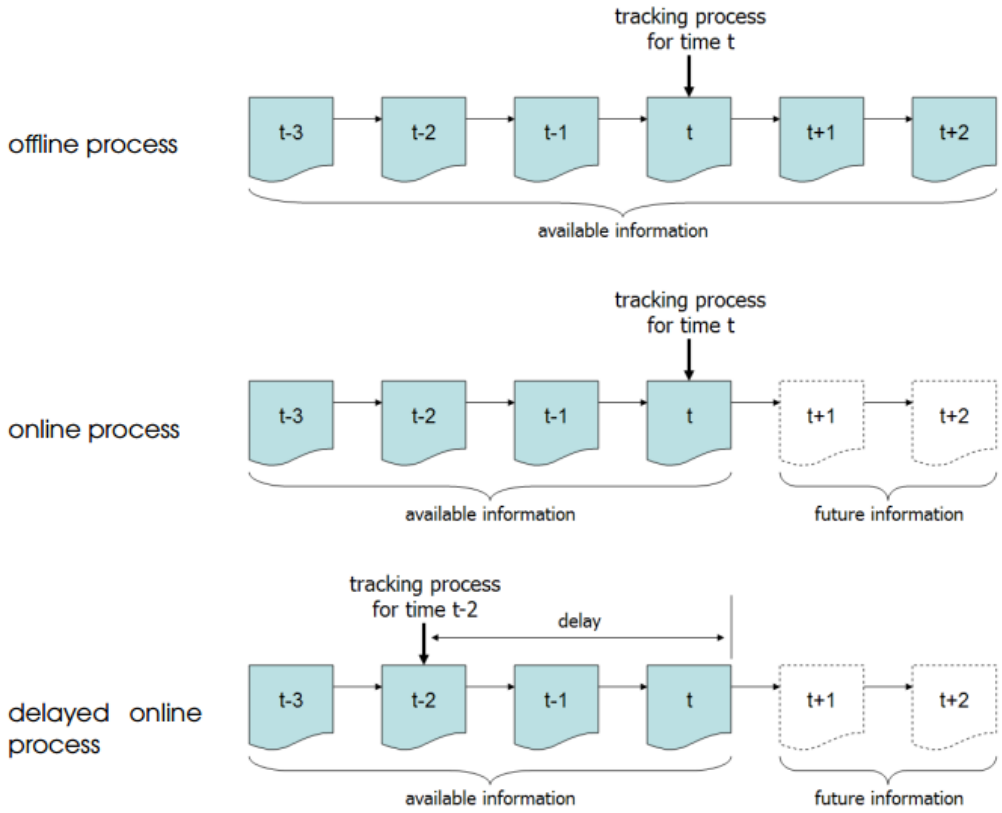
Ένα ακόμη κλασσικό πρόβλημα της παρακολούθησης πολλαπλών αντικειμένων αποτελεί η **παρουσία αποκρύψεων** (occlusions). Οι αποκρύψεις των αντικειμένων συμβαίνουν όταν ένα αντικείμενο, κινούμενο ή μη, που βρίσκεται πιο μπροστά από ένα άλλο σε σχέση με το οπτικό πεδίο της κάμερας. Κάτι τέτοιο μπορεί να προκαλέσει σοβαρά προβλήματα κατά την αντιστοίχιση των στόχων με τα αντικείμενα που έχουν εντοπιστεί (data association) γι' αυτό το λόγο πρέπει οι αλγόριθμοι παρακολούθησης που αναπτύσσονται να μπορούν να αντιμετωπίζουν τέτοια προβλήματα για να θεωρηθούν αξιόπιστοι.

Όλα τα προβλήματα που αναφέρθηκαν μέχρι τώρα είναι κοινά ως ένα βαθμό με τα προβλήματα που αντιμετωπίζουν οι αλγόριθμοι παρακολούθησης ενός αντικειμένου. Ωστόσο, το ζήτημα της παρακολούθησης πολλαπλών αντικειμένων είναι αρκετά πολύπλοκο με αποτέλεσμα να υπάρχουν αρκετά επιπλέον προβλήματα σε σχέση με την περίπτωση παρακολούθησης ενός αντικειμένου. Το πρώτο πρόβλημα που προσπαθούν να αντιμετωπίσουν οι αλγόριθμοι παρακολούθησης πολλαπλών αντικειμένων σχετίζεται με την **μοντελοποίηση των αντικειμένων**, δηλαδή κατά πόσο είναι εφικτή η ανα-

παράσταση των αντικειμένων υπό μία κοινή πλατφόρμα. Ορισμένοι αλγόριθμοι προτιμούν να αναπαριστούν κάθε αντικείμενο ξεχωριστά, ενώ άλλοι δημιουργούν μία κοινή αναπαράσταση για τα αντικείμενα που αλλάζει ανάλογα με τα χαρακτηριστικά τους.

Το δεύτερο σημαντικό πρόβλημα της παρακολούθησης πολλαπλών αντικειμένων αποτελεί η **αντιστοίχιση των δεδομένων** (data association), δηλαδή η αντιστοίχιση των ανιχνεύσεων με τους κατάλληλους στόχους. Θεωρώντας έναν πίνακα κόστους που εμπεριέχει όλες τις ανιχνεύσεις και στόχους για κάθε καρέ του βίντεο, η διαδικασία αντιστοίχισης ουσιαστικά αφορά την εύρεση εκείνων των αντιστοιχίσεων που ελαχιστοποιούν το συνολικό κόστος αντιστοιχίσεων. Ο πιο γνωστός αλγόριθμος αντιστοίχισης που συναντάται στη βιβλιογραφία αποτελεί ο λεγόμενος Hungarian Algorithm ή αλλιώς αλγόριθμος του Munkres [31], ο οποίος αναλύεται στο παράρτημα της παρούσας εργασίας. Άλλες δημοφιλείς μέθοδοι που συναντώνται στη διεθνή βιβλιογραφία είναι οι MHT [35] και JPDA [36], ενώ μία σχετικά πρόσφατη προσπάθεια αποτελεί ο αλγόριθμος HyperBoost [34].

Τέλος, αξίζει να αναφερθεί η επιπλέον δυσκολία που προστίθεται στη διαδικασία παρακολούθησης πολλαπλών στόχων όταν η εκάστοτε εφαρμογή απαιτεί αυτή να γίνεται σε **πραγματικό χρόνο** (online). Στη περίπτωση αυτή η πληροφορία που λαμβάνεται από τον αλγόριθμο πρέπει να είναι αρκετή ώστε να είναι διαθέσιμα τα αποτελέσματα παρακολούθησης των στόχων πριν από το αμέσως επόμενο καρέ του βίντεο. Οι αλγόριθμοι παρακολούθησης σε πραγματικό χρόνο έχουν διαθέσιμα στη προηγούμενη και την παρούσα πληροφορία του βίντεο, ενώ οι αλγόριθμοι που δεν εφαρμόζονται σε πραγματικό χρόνο (offline) έχουν διαθέσιμη ολόκληρη την ακολουθία του βίντεο. Ωστόσο, υπάρχει και μία ακόμη κατηγορία αλγορίθμων οι οποίοι εφαρμόζοντας μία μικρή χρονική καθυστέρηση (delayed online) μπορούν να συμπεριφέρονται ως αλγόριθμοι σε πραγματικό χρόνο. Στην Εικόνα 3.1 παρουσιάζονται οι διαφοροποιήσεις αυτών των αλγορίθμων. Συμπερασματικά, ενώ στη περίπτωση παρακολούθησης ενός αντικειμένου συναντώνται διάφοροι αλγόριθμοι που τρέχουν σε πραγματικό χρόνο, όσον αφορά τη παρακολούθηση πολλαπλών αντικειμένων δεν συναντάται ακόμη ίδια πληθώρα αλγορίθμων.



Εικόνα 3.1 : Διαδικασίες παρακολούθησης αντικειμένων σε πραγματικό ή μη χρόνο

3.2. Ανασκόπηση της βιβλιογραφίας

Η παρακολούθηση πολλαπλών αντικειμένων λόγω διαφόρων παραγόντων που αναφέρθηκαν στο προηγούμενο εδάφιο αποτελεί ένα αρκετά δύσκολο πρόβλημα, το οποίο είναι ανοιχτό ερευνητικά. Σε αυτό το εδάφιο, λοιπόν, θα πραγματοποιηθεί μία συνοπτική παρουσίαση ορισμένων state-of-the-art αλγορίθμων και θα εξαχθούν ορισμένα συμπεράσματα για τη γενικότερη κατεύθυνση που ακολουθεί η κοινότητα για να λύσει το πρόβλημα.

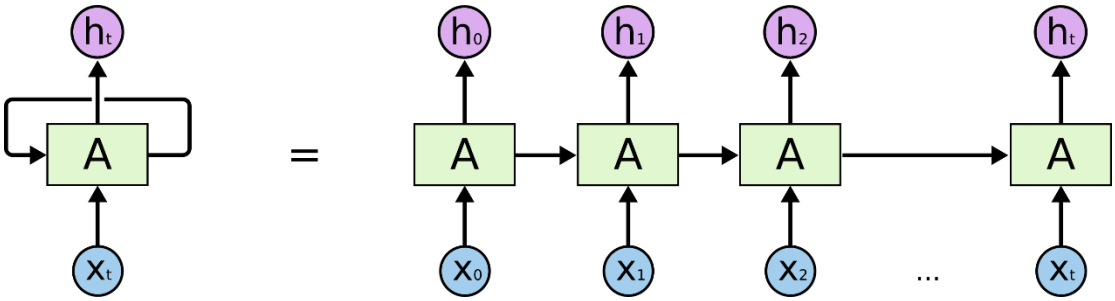
3.2.1. Νευρωνικά Δίκτυα και Οπτική Παρακολούθηση

Τα τελευταία χρόνια η διείσδυση του κλάδου της Μηχανικής Μάθησης (Machine Learning) και ιδίως των Τεχνητών Νευρωνικών Δικτύων (Neural Networks) στους αλγορίθμους παρακολούθησης πολλαπλών αντικειμένων είναι ιδιαίτερα έντονη. Η βασική αιτία για την οποία προτιμάται μία τέτοια προσέγγιση από τους ερευνητές οφείλεται κατά κύριο λόγο στην επιτυχία τεχνικών με νευρωνικά δίκτυα σε άλλα πεδία της Όρασης Υπολογιστών, όπως η αναγνώριση δράσεων, η ταξινόμηση εικόνων και η ανίχνευση αντικειμένων. Επιπλέον, ο ολοένα και αυξανόμενος όγκος των δεδομένων

βίντεο που συνοδεύονται από τα αντίστοιχα δεδομένα αληθείας και διατίθενται ελεύθερα στο διαδίκτυο έχει παίξει καθοριστικό ρόλο στη στροφή της επιστημονικής κοινότητας στην ανάπτυξη αλγορίθμων που χρησιμοποιούν νευρωνικά δίκτυα. Στη βιβλιογραφία συναντάται πληθώρα διαφορετικών αρχιτεκτονικών νευρωνικών δικτύων, όμως σε γενικές γραμμές οι ερευνητές προτιμούν να χρησιμοποιούν κυρίως μία συγκεκριμένη από αυτές που πιστεύεται ότι προσφέρει τα περισσότερα οφέλη.

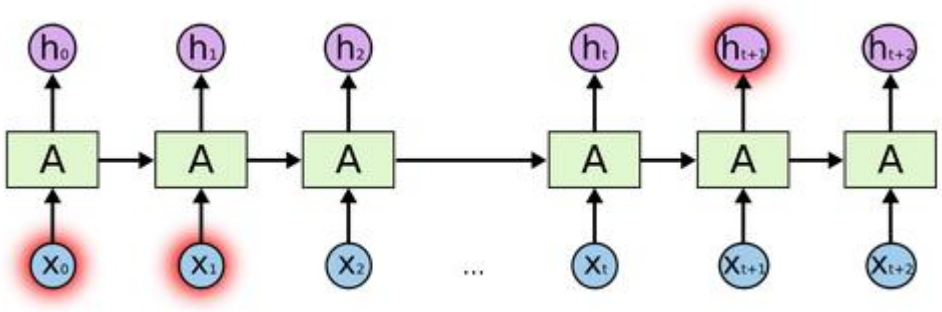
Όπως γνωρίζουμε, η επεξεργασία διαδικασίας της σκέψης του ανθρωπίνου εγκεφάλου δεν ξεκινά από την αρχή κάθε δευτερόλεπτο που περνά και προκειμένου να συντεθεί μία πληροφορία βασιζόμαστε στη πρότερη μνήμη που υφίσταται. Τα κλασικά νευρωνικά δίκτυα δεν μπορούν να υλοποιήσουν αυτή τη λογική, γεγονός που τα καθιστά αρκετά προβληματικά για τη περίπτωση παρακολούθησης πολλαπλών αντικειμένων, αφού είναι προφανές ότι για την εξαγωγή ποιοτικής πληροφορίας για κάποιο καρέ του βίντεο χρειαζόμαστε σχεδόν πάντα κάποια πληροφορία από προηγούμενα καρέ. Η αρχιτεκτονική νευρωνικών δικτύων, λοιπόν, που συναντάται σε όλο και περισσότερες πρόσφατες δημοσιεύσεις και μπορεί να αντιμετωπίσει αυτό το πρόβλημα είναι εκείνη των λεγόμενων **Αναδρομικών (ή Ανατροφοδοτούμενων) Νευρωνικών Δικτύων** (Recurrent Neural Networks – RNN). Χαρακτηρίζονται από εσωτερικές συνδέσεις ανατροφοδότησης (feedback connections) ή χρονικής υστέρησης (time delays) και μπορούν να χρησιμοποιούν την εσωτερική τους μνήμη για την επεξεργασία αυθαίρετων ακολουθιών δεδομένων εισόδου. Τέτοια δίκτυα χρησιμοποιούνται για την επίλυση προβλημάτων που απαιτούν την υλοποίηση μιας δυναμικής απεικόνισης του χώρου των διανυσμάτων εισόδου στο χώρο εξόδου, εν αντιθέσει για παράδειγμα με τη περίπτωση των MLP (Multi-Layer Perceptron) που απαιτούν αναμφίβολα μια στατική απεικόνιση.

Ένα χαρακτηριστικό παράδειγμα ενός απλού RNN παρουσιάζεται στην Εικόνα 3.2, όπου παρατηρούμε τον βρόγχο επανατροφοδότησης στον νευρώνα Α με δεδομένο εισόδου το x_t και τιμή εξόδου το h_t. Αυτός ο βρόγχος επιτρέπει την μετάδοση της πληροφορίας από το ένα χρονικό βήμα του δικτύου στο επόμενο με αποτέλεσμα τη διαχρονική «διατήρηση» της πληροφορίας. Παρά την φαινομενική διαφορετική προσέγγιση που ακολουθούν τα RNN, στη πραγματικότητα δεν διαφοροποιούνται τόσο έντονα σε σχέση με ένα κανονικό νευρωνικό δίκτυο. Ουσιαστικά ένα RNN μπορεί να θεωρηθεί ως πολλαπλά αντίγραφα του ίδιου δικτύου, το καθένα από τα οποία μεταβιβάζει ένα μήνυμα στο επόμενο. Έτσι, αν «ξεδιπλώσουμε» τον βρόγχο προκύπτει η ακόλουθη ισοδύναμη αρχιτεκτονική του ΑΝΔ, όπως παρουσιάζεται επίσης στην Εικόνα 3.2.



Εικόνα 3.2 : Παράδειγμα ενός απλού αναδρομικού νευρωνικού δικτύου.

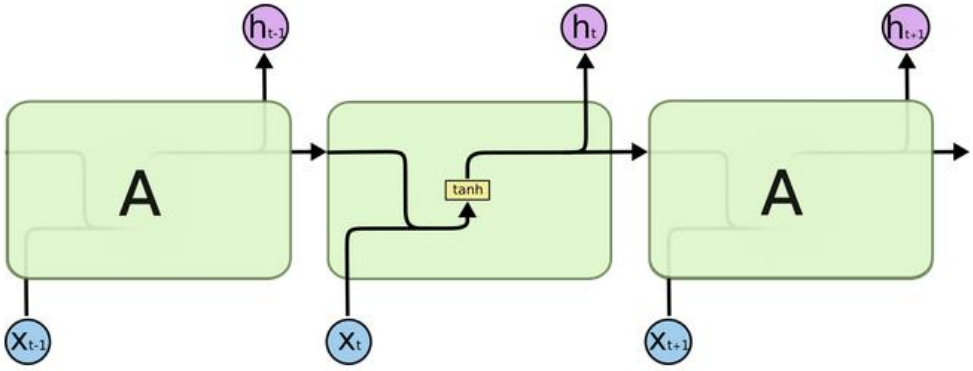
Όπως αναφέρθηκε προηγουμένως, το βασικό πλεονέκτημα των RNN έναντι άλλων αρχιτεκτονικών αποτελεί το γεγονός ότι για τη σύνθεση μιας πληροφορίας σε ένα στάδιο μπορεί να χρησιμοποιηθεί η πρότερη μνήμη του δικτύου. Ωστόσο, η δυνατότητα αυτή στη περίπτωση των RNN περιορίζεται μόνο στη χρησιμοποίηση σχετικά «πρόσφατης» πληροφορίας. Σε αρκετές εφαρμογές, όπως στη παρακολούθηση αντικειμένων για μεγάλο χρονικό διάστημα, προκειμένου να προκύψει μία ορθή πρόβλεψη για τον στόχο το δίκτυο απαιτεί τη πρόσβαση σε πληροφορία αρκετά παρελθοντικών χρονικών στιγμών. Ωστόσο, στη περίπτωση των RNN όταν το κενό μεταξύ της σχετικής πληροφορίας και της επανάληψης που απαιτείται για την εξαγωγή της πρόβλεψης γίνεται αρκετά μεγάλο, τότε τα δίκτυα αυτά αποτυγχάνουν να βγάλουν ορθά συμπεράσματα (Εικόνα 3.3). Θεωρητικά, τα RNN έχουν τη δυνατότητα να χειρίζονται τέτοιες περιπτώσεις μακροπρόθεσμων εξαρτήσεων (long-term dependencies) χρησιμοποιώντας προσεκτικά ορισμένες παραμέτρους και για περιορισμένο πλήθος προβλημάτων. Πρακτικά, ωστόσο, τα RNN δεν μπορούν να ανταποκριθούν με επιτυχία σε τέτοια προβλήματα, γεγονός το οποίο έχει εξεταστεί εκ βάθους από τους Hochreiter και Bengio (1994). Αυτοί απέδειξαν σε θεωρητικό και πειραματικό επίπεδο ότι η ελάττωση της παραγώγου (gradient descent) του κριτηρίου σφάλματος μπορεί να είναι ανεπαρκής για την εκπαίδευση του δικτύου σε περιπτώσεις μακροπρόθεσμων εξαρτήσεων.



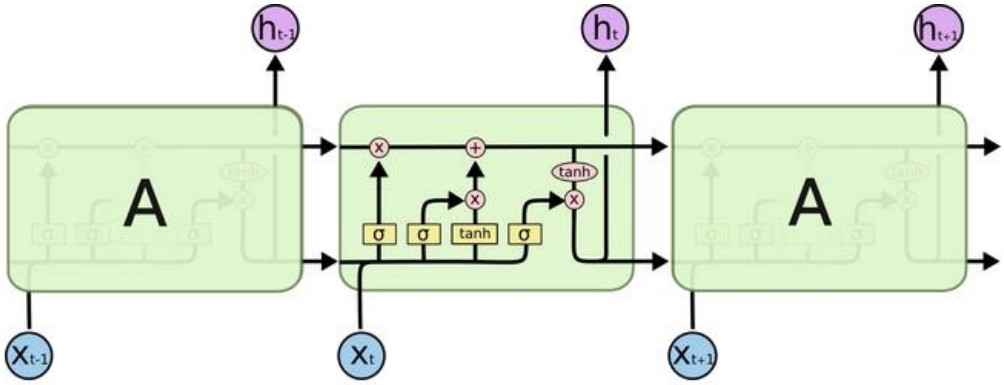
Εικόνα 3.3 : Αδυναμία των RNN για διαχείριση μακροπρόθεσμων εξαρτήσεων

Μία κατηγορία αναδρομικών νευρωνικών δικτύων που μπορεί να διαχειρίζεται καλύτερα τις μακροπρόθεσμες εξαρτήσεις και χρησιμοποιείται ολοένα και περισσότερο από τους ερευνητές για την παρακολούθηση πολλαπλών αντικειμένων είναι τα επονομαζόμενα **Δίκτυα Μακρο-Βραχυπρόθεσμης Μνήμης (Long-Short Term Memory - LSTM)**. Αναπτύχθηκαν αρχικά το 1997 από τους Hochreiter και Schmidhuber και έγιναν ιδιαίτερα δημοφιλή στην επιστημονική κοινότητα για την χρήση τους σε ευρεία προβλημάτων. Είναι ειδικά σχεδιασμένα ώστε να ικανοποιούν ακριβώς την ανάγκη μακροπρόθεσμης μνήμης του νευρωνικού δικτύου με αποτέλεσμα να μην δυσκολεύονται καθόλου στην εκμάθηση πληροφορίας για μεγάλα χρονικά διαστήματα.

Όλοι οι τύποι αρχιτεκτονικών αναδρομικών δικτύων έχουν τη μορφή μιας αλυσίδας επαναληπτικών ενοτήτων (modules) του νευρωνικού δικτύου, με τα πιο κλασσικά από τα οποία να έχουν μια πολύ απλή δομή, όπως ένα tanh επίπεδο. Όσον αφορά τα LSTM, διαθέτουν μια παρόμοια δομή σε γενικές γραμμές, όμως η επαναληπτική ενότητα είναι διαφορετική σε σχέση με τα κλασσικά RNN. Αντί, δηλαδή, να υφίσταται ένα απλό επίπεδο νευρωνικού δικτύου, στη περίπτωση των LSTM υπάρχουν τέσσερα τέτοια επίπεδα που αλληλεπιδρούν μεταξύ τους. Οι διαφορές μιας κλασσικής RNN αρχιτεκτονικής με μία LSTM αρχιτεκτονική παρουσιάζονται στις Εικόνες 3.4 και 3.5.



Εικόνα 3.4 : : Η επαναληπτική διαδικασία σε ένα κλασσικό RNN εμπεριέχει ένα επίπεδο



Εικόνα 3.5 : Η επαναληπτική διαδικασία σε ένα LSTM εμπεριέχει τέσσερα αλληλεπιδρώμενα επίπεδα

Σε γενικές γραμμές, οι αρχιτεκτονικές νευρωνικών δικτύων που κυριαρχούν σε αλγορίθμους παρακολούθησης πολλαπλών στόχων στη διεθνή βιβλιογραφία είναι αυτές που αναλύθηκαν, δηλαδή τα RNN και τα LSTM. Ωστόσο, εμφανής αλλά όχι σε τόσο μεγάλο βαθμό είναι η διείσδυση αρχιτεκτονικών **Συνελικτικών Νευρωνικών Δικτύων (Convolutional Neural Networks)**, γεγονός που κατά μεγάλο βαθμό οφείλεται στην επιτυχία αυτών των δικτύων σε εφαρμογές ταξινόμησης εικόνων και ανίχνευσης αντικειμένων. Τέλος, αξίζει να σημειωθεί ότι υπάρχουν δεκάδες παραλλαγές αυτών των αρχιτεκτονικών νευρωνικών δικτύων ανάλογα με τις ενδείξεις (cues) που έχουν αναπτύξει οι ερευνητές για τον αλγόριθμό τους.

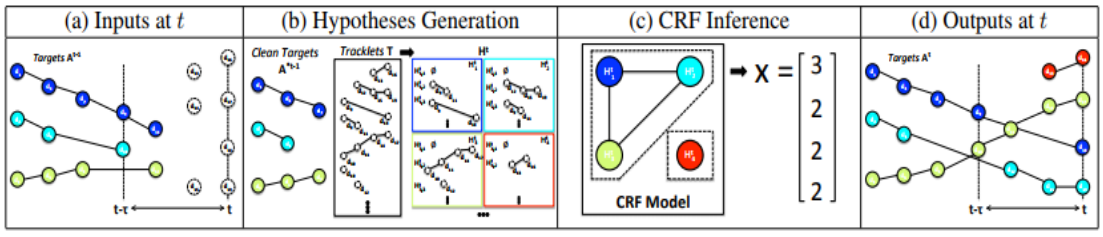
3.2.2. Ανασκόπηση της σχετικής βιβλιογραφίας

Στον παρακάτω πίνακα παρουσιάζονται οι καλύτεροι αλγόριθμοι παρακολούθησης πολλαπλών αντικειμένων για το σετ δεδομένων MOT16 μέχρι την 1η Μαρτίου 2016, όπως προκύπτει από τη δημοσίευση [46]. Επιπλέον, συνοδεύονται από διάφορα χαρακτηριστικά τους, όπως τους γεωμετρικούς μετασχηματισμούς που χρησιμοποιούν για να αντιμετωπίσουν την αφινικότητα μεταξύ των κουτιών περιγράμματος του ίδιου στόχου σε διαφορετικά καρέ, τη μέθοδο αντιστοίχισης των δεδομένων που αξιοποιούν και τα επιπλέον δεδομένα που απαιτούν, καθώς και αν χρησιμοποιούν κάποιο μοντέλο εμφάνισης ή αν εφαρμόζονται σε πραγματικό χρόνο. Στη συνέχεια του εδαφίου θα αναλυθούν ορισμένοι από τους αλγορίθμους του Πίνακα 3.1.

Αλγόριθμος	Γεωμετικοί μετασχηματισμοί	Μοντέλο εμφάνισης	Βελτιστοποίηση	Extra	Online
NOMT [10]	Τροχιές σημείων ενδιαφέροντος	✓	CRF	Οπτική ροή	✗
JMC [11]	DeepMatching	✓	Multicut	Non-NMS dets	✗
MDPNN16 [12]	RNN (κίνηση, εμφάνιση, αλληλεπιδράσεις)	✓	MRF	-	✓
oICF [13]	Μοντέλο κίνησης + MIL appearance	✓	Kalman filter	-	✓
MHT_DAM [14]	Regression classifier appearence	✓	MHT	-	✗
LINF1 [15]	Sparse representations appearance	✓	MCMC	-	✗
EAMTTpub [16]	2D αποστάσεις	✗	Particle filter	Non-NMS dets	✓
OVBT [17]	Δυναμική μέσω οπτικής ροής	✓	Variational EM	Οπτική ροή	✓
LTTSC-CRF [18]	SURF	✓	CRF	SURF	✗
LP2D [19]	2D αποστάσεις, IoU	✗	Global, LP	-	✗
TBD [20]	IoU + NCC	✓	Hungarian Algorithm	-	✗
CEM [21]	2D διαφορά ταχύτητας	✗	L-BFGS+greedy samping	-	✗
DP_NMS [22]	2D αποστάσεις	✗	k-shortest paths	-	✗
GMPHD_HDA [23]	HoG, ιστόγραμμα χρώματος	✓	Gaussian Mixture PHD filter	HoG	✓
SMOT [24]	Δυναμική στόχου	✗	Hankel Total LS	-	✗
JPDA_m [25]	Απόσταση Mahalanobis	✗	LP	-	✗

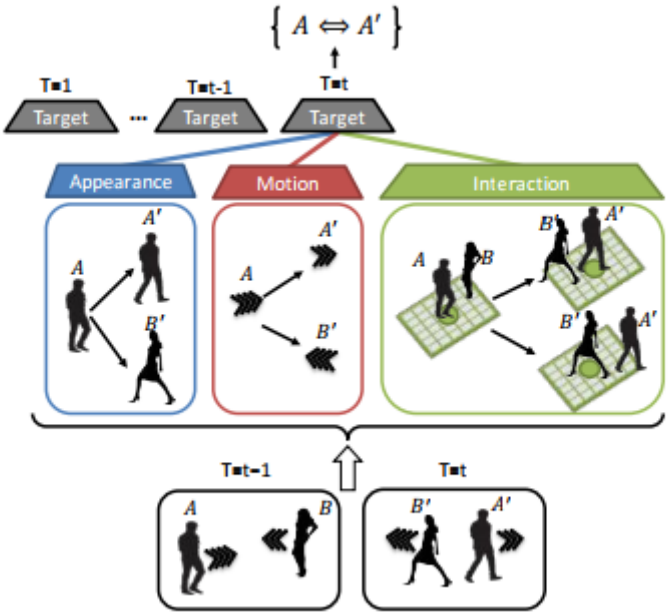
Πίνακας 3.1 : Οι καλύτεροι αλγόριθμοι παρακολούθησης με τα χαρακτηριστικά τους για το σετ δεδομένων MOT16 σύμφωνα με το [46].

Αναλυτικότερα, στον **αλγόριθμο NOMT** [10] οι ερευνητές προσπάθησαν να αντιμετωπίσουν δύο πολύ σημαντικές πτυχές του προβλήματος της παρακολούθησης πολλαπλών στόχων που είναι: α) ο σχεδιασμός ενός ακριβούς μέτρου αφινικότητας για το συσχετισμό ανιχνεύσεων και β) ο προσδιορισμός ενός αποδοτικού και ακριβούς αλγορίθμου παρακολούθησης πολλαπλών στόχων που εφαρμόζεται σε -σχεδόν- πραγματικό χρόνο. Όσον αφορά το πρώτο, αναπτύχθηκε ένας καινοτόμος περιγραφέας που ονομάζεται ALFD (Aggregated Local Flow Descriptor) που κωδικοποιεί τη σχετική κίνηση μεταξύ ενός ζεύγους χρονικά απόμερων ανιχνεύσεων χρησιμοποιώντας μακροπρόθεσμες τροχιές σημείων ενδιαφέροντος. Αξιοποιώντας αυτές τις τροχιές, ο περιγραφέας ALFD παρέχει ένα ιδιαίτερα εύρωστο μέτρο αφινικότητας που χρησιμοποιείται για την εκτίμηση της πιθανότητας αντιστοίχισης ανιχνεύσεων με στόχους ανεξαρτήτως της περίπτωσης εφαρμογής του αλγορίθμου. Επιπλέον, το ίδιο το πρόβλημα της παρακολούθησης αντιμετωπίστηκε από τους ερευνητές ως μία διαδικασία αντιστοίχισης των κινούμενων στόχων με τις ανιχνεύσεις σε ένα χρονικό παράθυρο, η οποία εκτελείται διαδοχικά σε κάθε καρέ του βίντεο. Ο αλγόριθμος αυτός καταφέρνει τόσο να είναι αποδοτικός όσο και να είναι ιδιαίτερα εύρωστος ενσωματώνοντας πολλαπλές ενδείξεις (cues) που συμπεριλαμβάνουν την μετρική ALFD, την δυναμική κίνηση του στόχου και της μακροπρόθεσμης κανονικοποίησης των τροχιών. Όπως θα παρατηρήσουμε σε επόμενο κεφάλαιο, ο αλγόριθμος NOMT καταφέρνει να είναι ο state-of-the-art μέχρι αυτή τη στιγμή πετυχαίνοντας, μάλιστα, 10% μεγαλύτερη ακρίβεια σε σχέση με τους επόμενους καλύτερους αλγορίθμους.



Εικόνα 3.6 : Σχηματική απεικόνιση του NOMT αλγορίθμου [10]. (α) Δεδομένου ενός συνόλου A^{t-1} στόχων και D_{t-r}^{t} ανιχνεύσεων, (β) η μέθοδος γεννά ένα σύνολο υποψήφιων υποθέσεων H^t χρησιμοποιώντας T tracklets. Κατασκευάζοντας ένα CRF μοντέλο (γ) επιλέγεται η πιο συνεπής λύση x και (δ) οι εξαγόμενοι στόχοι A^t επιλέγονται αυξάνοντας τους προηγούμενους στόχους A^{t-1} με τη λύση $H^t(\hat{x})$.

Οι περισσότεροι αλγόριθμοι παρακολούθησης πολλαπλών στόχων δεν συνδυάζουν διαφορετικές ενδείξεις (cues) με έναν συνεπή τρόπο για μεγάλο χρονικό διάστημα. Κάτι τέτοιο, όμως, υποστηρίζει ότι καταφέρνει ο **αλγόριθμος MDPNN** [12] που αναπτύχθηκε από το πανεπιστήμιο του Stanford και εφαρμόζεται μάλιστα σε πραγματικό χρόνο. Προκειμένου να αντιμετωπιστεί το πρόβλημα λανθασμένης αντιστοίχισης των στόχων με τις ανιχνεύσεις λόγω αποκρύψεων ή παρόμοιων χαρακτηριστικών εμφάνισης, χρησιμοποιήθηκε μία αρχιτεκτονική Αναδρομικών Νευρωνικών Δικτύων (Recurrent Neural Networks – RNN) που μπορεί να συνειδητοποιεί τις πολλαπλές ενδείξεις εντός ενός χρονικού παραθύρου. Μάλιστα, για το μοντέλο χωρικών αλληλεπιδράσεων μεταξύ των στόχων χρησιμοποιήθηκε μία ακόμη αρχιτεκτονική νευρωνικών δικτύων, τα οποία είναι μία υποκατηγορία των RNN και ονομάζονται Δίκτυα Μακρο-Βραχυπρόθεσμης Μνήμης (Long-Short Term Memory – LSTM). Με αυτό τον τρόπο, λοιπόν, επιτυγχάνεται η διόρθωση πολλών λαθών αντιστοίχισης και ανακτώνται οι παρατηρήσεις σε περίπτωση που αποκρύπτονται.



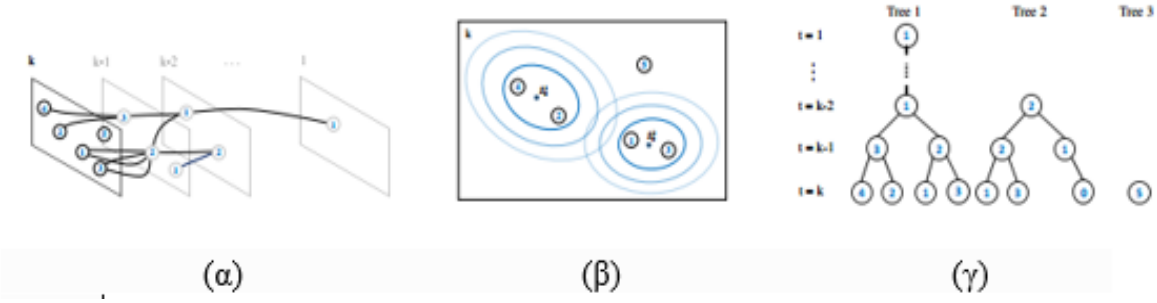
Εικόνα 3.7:Ο αλγόριθμος MDPNN βασίζεται σε μία δομή RNN (τραπέζια σχήματα) που μαθαίνει να κωδικοποιεί μακροπρόθεσμες χρονικά εξαρτήσεις μεταξύ πολλαπλών ενδείξεων (εμφάνιση, κίνηση και αλληλεπιδράσεις) [12].

Η παρακολούθηση πεζών σε πραγματικό χρόνο επωφελείται χρησιμοποιώντας ένα online δυναμικό μοντέλο εμφάνισης, αφού με αυτό τον τρόπο η διαδικασία της αντιστοίχισης στόχων και ανιχνεύσεων σαφέστατα μπορεί να επιτευχθεί με ακριβέστερο τρόπο. Δεδομένης, λοιπόν, της δημοφιλίας των χαρακτηριστικών ICF (Integral Channel Features) για γρήγορη ανίχνευση των πεζών μιας σκηνής, αναπτύχθηκε ο **αλγόριθμος oICF** [13] που αξιοποιεί ένα online μοντέλο εμφάνισης, το οποίο χρησιμοποιεί τα ίδια χαρακτηριστικά χωρίς να υπολογίζει κάποια άλλα. Ο αλγόριθμος αυτός

χρησιμοποιεί μία διαδικασία εκμάθησης πολλαπλών παραδειγμάτων (Multiple-Instance Learning - MIL) ώστε να εκπαιδεύσει το μοντέλο εμφάνισης κάθε ανθρώπου που καταφέρνει να τον διαχωρίζει από τους όμορούς του. Στην Εικόνα 3.8 παρουσιάζονται τα θετικά και αρνητικά δείγματα που δημιουργούνται για το δυναμικό μοντέλο εμφάνισης του αλγορίθμου.

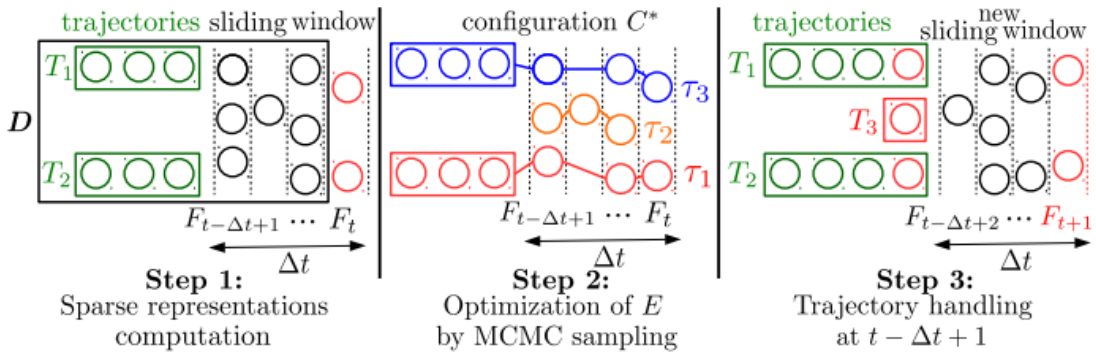
Εικόνα 3.8 : Οπτικοποίηση θετικών (πράσινα κουτιά) και αρνητικών (κόκκινα κουτιά) δειγμάτων που δημιουργεί ο αλγόριθμος oICF [13]. Με κίτρινο χρώμα εμφανίζεται η εκτιμώμενη θέση του στόχου.

Ο αλγόριθμος **MHT_DAM** [14] επαναπροσδιόρισε τον κλασσικό αλγόριθμο Παρακολούθησης Πολλαπλών Υποθέσεων (Multiple Hypotheses Tracking – MHT) υπό μία φιλοσοφία "παρακολούθησης μέσω ανιχνεύσεων". Η επιτυχία του αλγορίθμου MHT που πρωτοεμφανίστηκε τη δεκαετία του '90 βασίζεται σε μεγάλο βαθμό στη δυνατότητά του να διατηρεί μία μικρή λίστα πιθανών υποθέσεων, κάτι το οποίο μπορεί να επιτευχθεί με τους σημερινούς state-of-the-art αλγορίθμους ανίχνευσης. Οι ερευνητές προκειμένου να υλοποιήσουν τη δυνατότητα του αλγορίθμου για εκμετάλλευση υψηλότερης τάξης (σημασιολογικά) πληροφορίας, ανέπτυξαν μία μέθοδο για την online εκπαίδευση μοντέλων εμφάνισης για κάθε υπόθεση στόχου. Με αυτό τον τρόπο επιτυγχάνεται η εύκολη εκμάθηση αυτών των μοντέλων μέσω της τεχνικής των κανονικοποιημένων ελαχίστων τετραγώνων (Regularized Least Squares), η οποία απαιτεί μόνο λίγες παραπάνω λειτουργίες για κάθε "κλαδί" υπόθεσης.



Εικόνα 3.9 : Οπτικοποίηση του αλγορίθμου MHT. (α) Οι υποθέσεις των στόχων σε χρόνο k, (β) παραδείγματα διαδικασίας "φραγμού" της περιοχής δύο υποθέσεων στόχων που έχουν διαφορετικό κατώφλι d_{th} , (γ) τα αντίστοιχα δενδροδιαγράμματα των στόχων, όπου κάθε ακμή (node) συσχετίζεται με μία παρατήρηση στο (α).

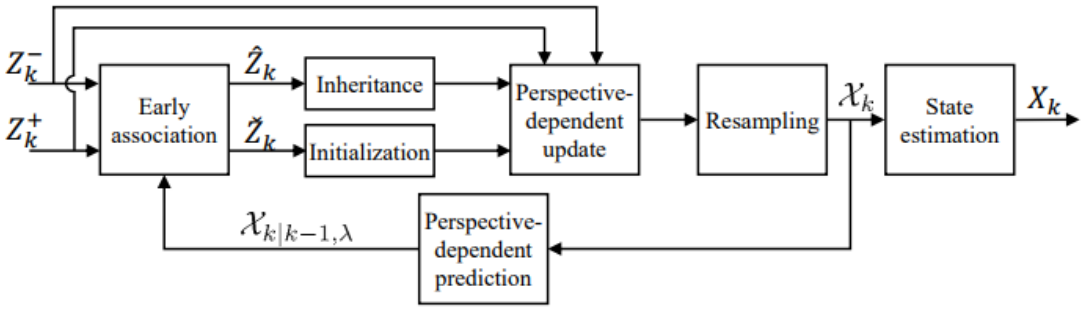
Επηρεασμένοι από την επιτυχία που αναγνωρίζουν οι αραιές αναπαραστάσεις στη περίπτωση της παρακολούθησης ενός αντικειμένου, οι Γάλλοι ερευνητές που ανέπτυξαν τον **αλγόριθμο LINF1** [15] αντιμετώπισαν τη διαδικασία αντιστοίχισης σε πολλά καρέ ως ένα πρόβλημα ελαχιστοποίησης ενέργειας. Σχεδίασαν, λοιπόν, μία ενέργεια που εκμεταλλεύεται αποδοτικά τις αραιές αναπαραστάσεις όλων των ανιχνεύσεων της σκηνής. Συνοπτικά, η λειτουργία του αλγορίθμου που βασίζεται στη διαδικασία που παρουσιάζεται στην Εικόνα 3.10 έχει ως εξής: (α) υπολογίζονται οι αραιές αναπαραστάσεις των ανιχνεύσεων (κύκλοι) από το τελευταίο καρέ, (β) η ολική ενέργεια E βελτιστοποιείται μέσω δειγματοληψίας MCMC (Markov Chain Monte Carlo) αποδίδοντας τον σχηματισμό C^* και (γ) οι τροχιές (τετράγωνα) υπολογίζονται οριστικά στο πρώτο καρέ του κινούμενου παραθύρου, ακολουθώντας C^*σχηματισμό.



Εικόνα 3.10 : Διαδικασία λειτουργίας του αλγορίθμου LINF1 [15].

Ο online **αλγόριθμος EAMTTpub** [16] εκμεταλλεύεται τόσο τις ανιχνεύσεις με μεγάλο επίπεδο εμπιστοσύνης όσο και εκείνες με μικρό επίπεδο εμπιστοσύνης διαμέσου ενός κοινού συστήματος. Οι ανιχνεύσεις με μεγάλο επίπεδο εμπιστοσύνης (strong detections) χρησιμοποιήθηκαν για αρχικοποίηση των στόχων και παρακολούθησή τους (label propagation), ενώ οι ανιχνεύσεις με μικρό επίπεδο εμπιστοσύνης

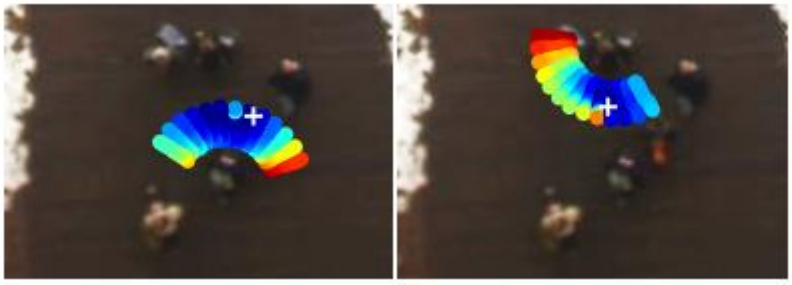
(weak detections) χρησιμοποιήθηκαν μόνο για την παρακολούθηση των υπάρχοντων στόχων στη σκηνή. Επιπρόσθετα, η διαδικασία αντιστοίχισης στόχων με ανιχνεύσεις εφαρμόστηκε αμέσως μετά τη διαδικασία πρόβλεψης της θέσης των στόχων αποφεύγοντας, έτσι, την ανάγκη για «ακριβές» υπολογιστικά διαδικασίες ανάθεσης ταυτότητας (label) στους στόχους. Τέλος, διενεργείται μία δειγματοληψία λαμβάνοντας υπόψη την προοπτική παραμόρφωση μεταξύ των παρατηρήσεων των στόχων. Κατά μέσο όρο ο αλγόριθμος εκτελείται στα 12 καρέ ανά δευτερόλεπτο, ενώ στην Εικόνα 3.11 παρουσιάζονται συνοπτικά τα στάδια λειτουργίας του.



Εικόνα 3.11 : Διαδικασία λειτουργίας του αλγορίθμου EAMTTpub [16].

Όσον αφορά τους αλγορίθμους παρακολούθησης ανθρώπων, τυπικά η οπτική παρακολούθηση βασίζεται σε ένα δυναμικό μοντέλο που προβλέπει την θέση του πεζού στη σκηνή λαμβάνοντας πληροφορία από το ιστορικό της τροχιάς του. Σε σενάρια μεγάλης πυκνότητας των πεζών στο οπτικό πεδίο του βίντεο η ύπαρξη ενός εύρωστου δυναμικού μοντέλου είναι καθοριστική διότι οι πιο ακριβείς προβλέψεις των τροχιών των στόχων επιτρέπουν την πιο ακριβή αντιστοίχιση των ανιχνεύσεων με τους στόχους. Τα "παραδοσιακά" δυναμικά μοντέλα προβλέπουν την θέση κάθε στόχου αποκλειστικά βασιζόμενα στις προηγούμενες θέσεις του στην εικόνα χωρίς να λαμβάνουν υπόψη τα υπόλοιπα (κινούμενα ή μη) αντικείμενα της εικόνας. Μία τέτοια προσέγγιση, όμως, αγνοεί μια πολύ σημαντική ιδιότητα της ανθρώπινης συμπεριφοράς : οι άνθρωποι οδηγούνται στον μελλοντικό τους προορισμό λαμβάνοντας υπόψη το περιβάλλον για πιθανές "συγκρούσεις" με άλλους πεζούς και προσαρμόζονται ανάλογα σε αρκετά προηγούμενο στάδιο ώστε να τους αποφύγουν. Κάτω από αυτή τη φιλοσοφία έχουν εμφανιστεί στη διεθνή βιβλιογραφία αρκετά πιο σύνθετα δυναμικά μοντέλα που βασίζονται σε **μοντέλα κοινωνικής συμπεριφοράς**, τα οποία όμως εφαρμόζονται κυρίως σε περιπτώσεις που υφίσταται υψηλή πυκνότητα πεζών στη σκηνή. Για παράδειγμα, στον αλγόριθμο παρακολούθησης YNWA [45] χρησιμοποιούνται δυναμικά μοντέλα πρόβλεψης που προσαρμόζονται βάσει της ανθρώπινης συμπεριφοράς. Όταν κάποιος πεζός κινείται κοντά σε άλλους ανθρώπους πολλοί παράγοντες επηρεάζουν τις βραχυπρόθεσμες κινήσεις του, συνεπώς περιορίζονται οι κινήσεις και η ταχύτητα του πεζού,

γεγονός το οποίο αυτός ο αλγόριθμος λαμβάνει υπόψη (Εικόνα 3.12). Παρόμοια λογική ακολουθείται και από τον αλγόριθμο LP2D [19].

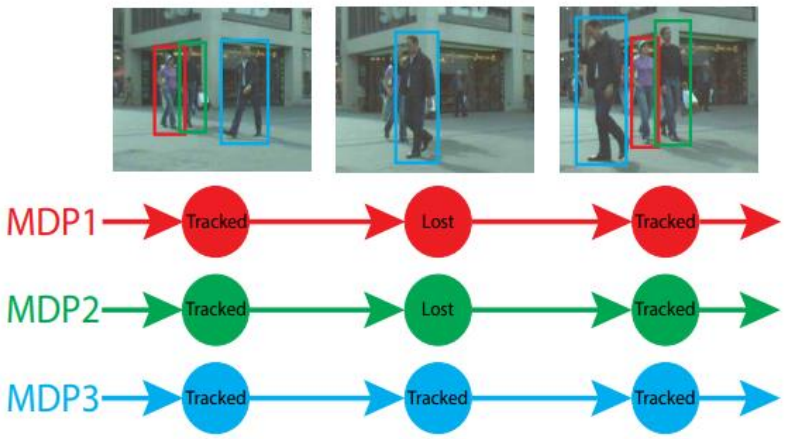


Εικόνα 3.12 : Το δυναμικό μοντέλο του αλγορίθμου YNWA [45] που δέχεται γνώσεις κοινωνικής συμπεριφοράς. Η μπλε περιοχή υποδεικνύει καλές επιλογές για την ταχύτητα, η κόκκινη περιοχή κακές επιλογές και ο λευκός σταυρός την πραγματική επιλογή που λαμβάνει ο στόχος.

Συμπερασματικά, παρατηρείται σε γενικές γραμμές πληθώρα διαφορετικών προσεγγίσεων για το πρόβλημα της παρακολούθησης πολλαπλών αντικειμένων χωρίς, όμως, κάποια να ξεχωρίζει από τις υπόλοιπες ως προς τα αποτελέσματα που προκύπτουν. Η λογική της "παρακολούθησης μέσω ανιχνεύσεων" (tracking-by-detection) φαίνεται ότι είναι η κυρίαρχη λογική που ακολουθείται από τους state-of-the-art αλγορίθμους μέχρι αυτή τη στιγμή. Ακόμη, οι πολύ καλές επιδόσεις αλγορίθμων που αξιοποιούν τα τεχνητά νευρωνικά δίκτυα σε άλλα πεδία της Όρασης Υπολογιστών σαφώς έχουν επηρεάσει και τους αλγορίθμους παρακολούθησης πολλαπλών αντικειμένων, αφού η συντριπτική τους πλειοψηφία πλέον χρησιμοποιεί τέτοιες τεχνικές. Κυρίαρχες αρχιτεκτονικές φαίνεται ότι είναι τα Αναδρομικά Νευρωνικά Δίκτυα (RNN) και τα Δίκτυα Μακρο-Βραχυπρόθεσμης Μνήμης (LSTM) για λόγους που αναλύθηκαν στο προηγούμενο εδάφιο. Μάλιστα, οι αλγόριθμοι που πετυχαίνουν τα καλύτερα αποτελέσματα επιλέγουν να συνδυάσουν διάφορες αρχιτεκτονικές νευρωνικών δικτύων και να τις μετατρέψουν έτσι ώστε το δίκτυο να "μαθαίνει" ακριβέστερα τα χαρακτηριστικά του μοντέλου εμφάνισης και κίνησης ενός στόχου, καθώς και των χωρικών αλληλεπιδράσεων μεταξύ των στόχων. Τέλος, σε αντίθεση με τους αλγορίθμους παρακολούθησης ενός αντικειμένου, παρατηρείται ότι η κοινότητα δεν έχει επικεντρωθεί ακόμη στην ανάπτυξη αλγορίθμων που εφαρμόζονται σε πραγματικό χρόνο πιθανότατα λόγω των χαμηλών επιδόσεων που παρατηρούνται στους state-of-the-art αλγορίθμους.

3.3. Αναλυτική περιγραφή του αλγόριθμου MDP

Σε αυτό το σημείο θα πραγματοποιηθεί μία εκτενής αναφορά στον αλγόριθμο MDP [8] που χρησιμοποιήθηκε, μάλιστα, ως μέτρο σύγκρισης για τον αλγόριθμο που αναπτύχθηκε σε αυτή την εργασία. Η διαδικασία παρακολούθησης πολλαπλών στόχων σε αυτό τον αλγόριθμο μοντελοποιείται ως ένα πρόβλημα λήψης αποφάσεων Μαρκοβιανών Διαδικασιών Απόφασης (Markov Decision Processes – MDPs). Συγκεκριμένα, η «διάρκεια ζωής» κάθε αντικειμένου ξεχωριστά στην ακολουθία εικόνων μοντελοποιείται με μία ΜΔΑ. Η βασική ιδέα της μεθόδου έγκειται στο γεγονός ότι κάθε ΜΔΑ θεωρείται στο μοντέλο ως μία ξεχωριστή διαδικασία (agent) που έχει στόχο να εντοπίσει ένα στόχο. Συνεπώς, για τον εντοπισμό πολλαπλών αντικειμένων χρησιμοποιούνται πολλές ΜΔΑ όπου για καθεμία από τις οποίες ακολουθείται μία διαδικασία «εξαναγκασμένης» εκμάθησης (reinforcement learning) βάσει συγκεκριμένων κανόνων (policies).



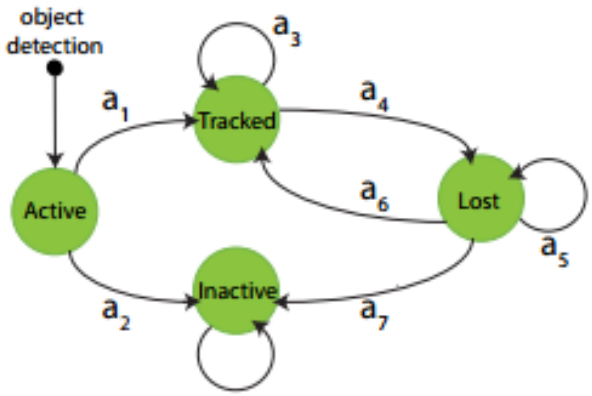
Εικόνα 3.13 : Μοντελοποίηση του προβλήματος της παρακολούθησης πολλαπλών στόχων μέσω Μαρκοβιανών Διαδικασιών Απόφασης από τον αλγόριθμο MDP [8].

Η εκμάθηση μιας πολιτικής για την ΜΔΑ πρακτικά είναι ισάξια με την εκμάθηση μιας συνάρτησης για την συσχέτιση των δεδομένων (data association), ενώ η πολιτική εκμάθησης πραγματοποιείται με έναν εξαναγκασμένο τρόπο έτσι ώστε το μοντέλο να επωφελείται από τα προτερήματα της εκμάθησης σε πραγματικό (online) ή μη (offline) χρόνο για την συσχέτιση των δεδομένων. Σε πρώτο βήμα, η εκμάθηση πραγματοποιείται σε μη πραγματικό χρόνο έτσι ώστε να αξιοποιείται η πληροφορία που παρέχεται από τις πραγματικές (ground truth) τροχιές. Προκειμένου η κάθε ΜΔΑ να είναι σε θέση να πάρει μία απόφαση που θα βασίζεται τόσο στο τρέχον καρέ όσο και στα προηγούμενα καρέ του στόχου, η εκμάθηση στη μέθοδο αυτή λαμβάνει χώρα κατά τη διάρκεια του εντοπισμού αντικειμένων σε δεδομένα εκπαίδευσης. Πιο συγκεκριμένα, δεδομένων της πραγματικής τροχιάς ενός στόχου και μιας αρχικής συνάρτησης ομοιότητας, η κάθε ΜΔΑ προσπαθεί αρχικά να ανιχνεύσει τον στόχο και συλλέγει

πληροφορίες ανατροφοδότησης (feedback) από τα δεδομένα αληθείας. Συνεπώς, ανάλογα με την πληροφορία ανατροφοδότησης που λαμβάνει η κάθε ΜΔΑ ανανεώνει τη συνάρτηση ομοιότητας ώστε να βελτιωθεί η παρακολούθηση του στόχου. Αξίζει να σημειωθεί ότι η συνάρτηση ομοιότητας ανανεώνεται μόνο στη περίπτωση όπου η ΜΔΑ κάνει λάθος στη φάση αντιστοίχισης των δεδομένων ανίχνευσης με τα δεδομένα τροχιών παρακολούθησης, γεγονός το οποίο επιτρέπει τη συλλογή «δύσκολων» δεδομένων εκπαίδευσης για την εκμάθηση της συνάρτησης ομοιότητας. Τέλος, η εκμάθηση λαμβάνει τέλος όταν η κάθε ΜΔΑ ανιχνεύει επιτυχώς τον στόχο. Ακόμη, το μοντέλο αυτό μπορεί να χειρίζεται τη γέννηση/θάνατο και την εμφάνιση/εξαφάνιση κινούμενων στόχων αντιμετωπίζοντάς τα ως μεταβάσεις καταστάσεων (state transitions) στο MDP ενώ παράλληλα αξιοποιούνται οι υπάρχουσες μέθοδοι online εντοπισμού μεμονωμένων αντικειμένων.

Μαρκοβιανές Διαδικασίες Απόφασης (Markov Decision Processes)

Η «διάρκεια ζωής» κάθε στόχου στο βίντεο μοντελοποιείται με μία ΜΔΑ, η οποία αποτελείται από την πλειάδα $(S, A, T(\cdot), R(.))$, όπου $s \in S$ είναι η κατάσταση (state) του στόχου, $a \in A$ είναι μία δράση (action) που μπορεί να εφαρμοστεί στον στόχο, $T: S\ x\ A \rightarrow S$ είναι η συνάρτηση που περιγράφει την επίδραση μίας δράσης σε μία κατάσταση του στόχου και $R: S\ x\ A \rightarrow \mathbb{R}$ είναι η συνάρτηση «επιβράβευσης» (rewarding function) της επίδρασης μίας δράσης σε μία κατάσταση του στόχου.



Εικόνα 3.14 : Ο χώρος καταστάσεων μιας Μαρκοβιανής Διαδικασίας Απόφασης στον αλγόριθμο MDP [8]

Όπως παρουσιάζεται στην παραπάνω εικόνα, ο χώρος S των καταστάσεων μιας MDP διαιρείται σε τέσσερις υποκαταστάσεις:

$$S = S_{active} \bigcup S_{tracked} \bigcup S_{lost} \bigcup S_{inactive}$$

Οι δράσεις a_i που μπορούν να επιτευχθούν μεταξύ των τεσσάρων υποκαταστάσεων είναι επτά και καθορίζονται ντετερμινιστικά. Ακόμη, σημειώνεται ότι η συνάρτηση «επιβράβευσης» T των δράσεων δεν δίνεται εκ των προτέρων αλλά απαιτείται να επιτευχθεί η εκμάθησή της από τα δεδομένα.

Η αρχική κατάσταση για κάθε στόχο που εντοπίζεται είναι η **ενεργή (active)**. Ένας ενεργός στόχος μπορεί να **παρακολουθείται (tracked)** ή να είναι **ανενεργός (inactive)**. Στην ιδεατή περίπτωση μία ορθή και μία λανθασμένη απόκριση του ανιχνευτή (detector) για έναν στόχο πρέπει να έχει ως αποτέλεσμα τη μεταβολή της κατάστασης του στόχου από ενεργή σε κατάσταση που παρακολουθείται και ανενεργή κατάσταση αντίστοιχα. Σύμφωνα με το μοντέλο, ένας στόχος που παρακολουθείται μπορεί να συνεχίσει να βρίσκεται σε αυτή την κατάσταση ή να μεταβεί σε μία κατάσταση που έχει **χαθεί (lost)** στην περίπτωση όπου για κάποιο λόγο (π.χ. αποκρύψεις) έχει εξαφανιστεί από το πεδίο όρασης της κάμερας. Αντίστοιχα, ένας χαμένος στόχος μπορεί είτε να συνεχίσει να βρίσκεται σε αυτή την κατάσταση είτε να μεταβεί στην κατάσταση που παρακολουθείται αν εμφανιστεί ξανά στο βίντεο είτε να μεταβεί στην ανενεργή κατάσταση αν έχει χαθεί για αρκετά μεγάλο χρονικό διάστημα από το βίντεο. Τέλος, η ανενεργή κατάσταση αφορά έναν στόχο που έχει εξαφανιστεί τελείως από το βίντεο.

Πολιτικές εκμάθησης των ΜΔΑ (policy learning)

Γενικά, μία πολιτική εκμάθησης $\pi : S \rightarrow A$ σε μία ΜΔΑ αποτελεί η συνάρτηση απεικόνισης του χώρου των καταστάσεων S στον χώρο των δράσεων A. Σκοπός της είναι η εύρεση εκείνης της πολιτικής που μεγιστοποιεί την συνάρτηση «επιβράβευσης» $R: S\ x\ A\ \rightarrow \mathbb{R}$.

Όσον αφορά την **πολιτική σε μία ενεργή κατάσταση *s***, μία ΜΔΑ αποφασίζει για την μετακίνηση μιας ανίχνευσης αντικειμένου σε στόχο που παρακολουθείται ή σε έναν ανενεργό στόχο. Για τον λόγο αυτό εκπαιδεύεται μία Μηχανή Διανυσμάτων Υποστήριξης (SVM) δυαδικής μορφής ώστε ένας στόχος να ταξινομηθεί ως εντοπισμένος ή ανενεργός χρησιμοποιώντας ένα κανονικοποιημένο 5D διάνυσμα χαρακτηριστικών $\varphi_{active}(s)$ που εμπεριέχει 2D συντεταγμένες, πλάτος, ύψος και score του στόχου. Σημειώνεται ότι μία λάθος απόκριση από τον ανιχνευτή μπορεί να οδηγήσει σε λανθασμένη ταξινόμηση κατά την οποία ο στόχος θα έχει μεταβεί στην εντοπισμένη κατάσταση, όμως το μοντέλο αναμένεται να λειτουργήσει ορθά και στη συνέχεια να οδηγήσει αυτό τον στόχο στην χαμένη κατάσταση.

Στη περίπτωση της κατάστασης που ένας **στόχος παρακολουθείται**, η MDP καλείται να αποφασίσει για το αν πρέπει ο στόχος να συνεχίσει να παρακολουθείται ή αν πρέπει να μεταφερθεί σε μία κατάσταση που έχει χαθεί. Αυτή η διαδικασία απόφασης πρακτικά αφορά την παρακολούθηση ενός αντικειμένου, αφού κάθε ΜΔΑ, όπως αναφέρθηκε, αντιπροσωπεύει έναν στόχο του βίντεο. Γι' αυτό τον λόγο οι ερευνητές χρησιμοποιούν τον αλγόριθμο παρακολούθησης TLD [37], ο οποίος δουλεύει μέσω αντιστοίχισης των προτύπων (templates), δηλαδή κουτιών που περιγράφουν γεωμετρικά τον στόχο, μεταξύ των διαδοχικών καρέ του βίντεο. Όταν ένα αντικείμενο μεταφέρεται στη κατάσταση που παρακολουθείται, το πρότυπο του στόχου αρχικοποιείται από το κουτί που προκύπτει από τον αλγόριθμο εντοπισμού και στη συνέχεια η ΜΔΑ που αφορά αυτόν τον στόχο συλλέγει τα πρότυπα των καρέ που εντοπίστηκε ο στόχος. Οι τελευταίες αντιπροσωπεύουν το ιστορικό παρακολούθησης του στόχου και πρόκειται να είναι χρήσιμες για τη λήψη αποφάσεων στη κατάσταση που ο στόχος έχει χαθεί.

Εικόνα 3.15 : Η εμφάνιση του στόχου αναπαρίσταται από πρότυπα που βασίζονται στον υπολογισμό της οπτικής ροής σε διαδοχικά καρέ μέσω πυκνής δειγματοληψίας σημείων ενδιαφέροντος εντός του κουτιού περιγράμματος του στόχου [8]. Η ποιότητα της οπτικής ροής χρησιμοποιείται από τον αλγόριθμο για να αποφασίσει για την ποιότητα της πρόβλεψης.

Πιο συγκεκριμένα για την **παρακολούθηση του στόχου**, αρχικά υπολογίζεται η οπτική ροή για πυκνά και ομοιόμορφα κατανεμημένα σημεία των προτύπων που αφορούν διαδοχικά καρέ μέσω της πολυκλιμακωτής μεθόδου Lucas-Kanade. Στη συνέχεια χρησιμοποιείται η μετρική σφάλματος Forward-Backward ώστε να εκτιμηθεί με τη χρήση ενός κατωφλίου πόσο σταθερή είναι η πρόβλεψη. Ωστόσο, δεν αποτελεί εύρωστη τεχνική η λήψη μιας απόφασης για την παρακολούθηση του στόχου αποκλειστικά με τη γνώση της οπτικής ροής. Γι' αυτό τον λόγο οι ερευνητές επιλέγουν να βοηθήσουν το μοντέλο τους με τη λήψη περαιτέρω πληροφορίας από τον ανιχνευτή που χρησιμοποιούν, βασιζόμενοι ότι μία λανθασμένη ανίχνευση δεν μπορεί να συμβαίνει για πολλά καρέ του βίντεο. Εάν ένας στόχος που παρακολουθείται, δηλαδή, δεν συναντάται από ορθές ανιχνεύσεις για μερικά καρέ τότε είναι αρκετά πιθανό να πρόκειται απλά για μία λάθος απόκριση. Συνεπώς, εξετάζεται το ιστορικό παρακολούθησης του στόχου και υπολογίζεται η επικάλυψη των κουτιών περιγράμματος μεταξύ του στόχου και των αντίστοιχων ανιχνεύσεων για ένα πλήθος καρέ. Τέλος, υπολογίζεται η μέση επικάλυψη των κουτιών περιγράμματος για αυτό το πλήθος καρέ, η οποία εν τέλει

χρησιμοποιείται ως μία άλλη μετρική για τη λήψη της απόφασης. Αξίζει να σημειωθεί ότι η ανανέωση των templates γίνεται μόνο στην περίπτωση όπου το πρότυπο που χρησιμοποιείται μπορεί να παρακολουθεί ικανοποιητικά τον στόχο.

Στη περίπτωση που ένας **στόχος έχει χαθεί**, η ΜΔΑ αναλαμβάνει να αποφασίσει για το αν ο στόχος είτε θα συνεχίσει να θεωρείται ότι έχει χαθεί είτε θα μεταφερθεί στην ανενεργή κατάσταση είτε θα ξαναρχίσει να παρακολουθείται. Αν ένας στόχος έχει χαθεί από το βίντεο για παραπάνω από T_{lost} καρέ τότε μεταφέρεται στην ανενεργή κατάσταση. Η δύσκολη περίπτωση έγκειται στη λήψη απόφασης για το πόσο ένας στόχος πρέπει να θεωρείται ότι έχει χαθεί ή πρέπει να αρχίσει να παρακολουθείται ξανά. Προκειμένου ένας στόχος να μεταφερθεί στην κατάσταση παρακολούθησης από την κατάσταση που έχει χαθεί, λοιπόν, πρέπει αυτός να αντιστοιχίζεται με τις ανιχνεύσεις που έχουν προκύψει από τον αλγόριθμο ανίχνευσης. Η αντιστοίχιση των δεδομένων γίνεται με τη χρήση ενός δυαδικού ταξινομητή που χρησιμοποιεί μία πραγματική γραμμική συνάρτηση η οποία ορίζεται από ένα σύνολο παραμέτρων προς μάθηση και δέχεται ως είσοδο ένα διάνυσμα χαρακτηριστικών που περιγράφει την ομοιότητα μεταξύ του στόχου και της ανίχνευσης. Ο δυαδικός αυτός ταξινομητής εκπαιδεύεται με τη χρήση της τεχνικής της εξαναγκασμένης εκμάθησης κατά την οποία δημιουργούνται θετικά και αρνητικά παραδείγματα εκπαίδευσης. Το σετ δεδομένων εκπαίδευσης μεγαλώνει σταδιακά καθώς προστίθενται δεδομένα βίντεο στο σύστημα και ο ταξινομητής επανεκπαιδεύεται κάθε φορά στο καινούργιο σετ δεδομένων εκπαίδευσης. Ένα πλεονέκτημα αυτής της διαδικασίας εκμάθησης είναι η δυνατότητα που παρέχει στο σχεδιασμό χαρακτηριστικών που βασίζονται στην εκάστοτε κατάσταση και στο ιστορικό παρακολούθησης του εκάστοτε στόχου. Ειδικότερα, το ιστορικό παρακολούθησης αναπαρίσταται από k πρότυπα στα περασμένα k καρέ όπου ο στόχος παρακολουθούνταν πριν μεταφερθεί στη κατάσταση που έχει χαθεί. Επίσης, υπολογίζεται η οπτική ροή από το κάθε πρότυπο στην ανίχνευση αλλά περιορίζεται μέσα σε μία γειτονιά γύρω από το κουτί περιγράμματος της ανίχνευσης. Έπειτα, υπολογίζεται βάσει διάφορων μετρικών η ποιότητα της οπτικής ροής με αποτέλεσμα να συνίσταται ένα σύνολο χαρακτηριστικών. Τέλος, σε αυτό το διάνυσμα χαρακτηριστικών προστίθενται χαρακτηριστικά ομοιότητας μεταξύ των κουτιών περιγράμματος, του στόχου και της ανίχνευσης.

Παρακολούθηση πολλαπλών αντικειμένων με ΜΔΑ

Για την παρακολούθηση πολλαπλών αντικειμένων αφιερώνεται μία ΜΔΑ για κάθε αντικείμενο/στόχο, η οποία ακολουθεί την πολιτική την οποία έχει μάθει για να παρακολουθήσει αυτό το αντικείμενο. Ειδικότερα, δεδομένου ενός νέου καρέ εισόδου, οι στόχοι που βρίσκονται σε κατάσταση παρακολούθησης επεξεργάζονται πρώτοι και προσδιορίζεται το κατά πόσο πρέπει να παραμείνουν υπό παρακολούθηση ή να μεταφερθούν σε κατάσταση που έχουν χαθεί. Έπειτα, υπολογίζεται η κατά ζεύγη ομοιότητα

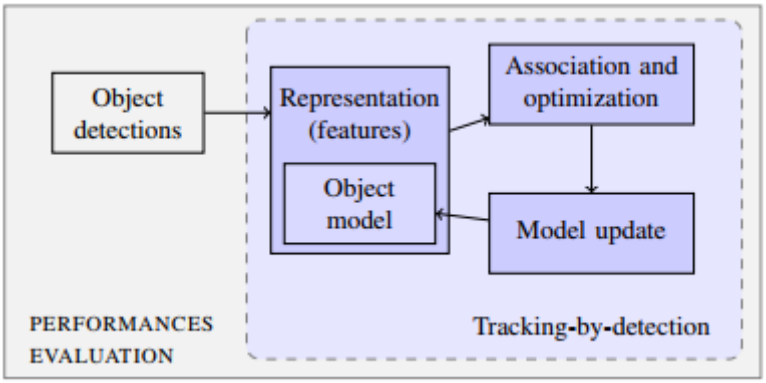
μεταξύ στόχων που έχουν χαθεί και ανιχνεύσεων που δεν ανήκουν σε υπό παρακολούθηση στόχους. Εν προκειμένω, χρησιμοποιείται η τεχνική καταστολής μη μεγίστων (non maximum suppression) βασιζόμενη στην επικάλυψη των κουτιών περιγράμματος των ανιχνεύσεων έτσι ώστε να αποφευχθούν επικαλυπτόμενες ανιχνεύσεις. Το τελικό κόστος ομοιότητας υπολογίζεται από τον δυαδικό ταξινομητή αντιστοίχισης δεδομένων και χρησιμοποιείται στον Hungarian Algorithm [31] για να προκύψουν οι τελικές αντιστοιχίσεις μεταξύ ανιχνεύσεων και χαμένων στόχων. Ανάλογα με τα αποτελέσματα του παραπάνω αλγορίθμου αποφασίζεται ποιοι χαμένοι στόχοι θα παραμείνουν σε αυτή την κατάσταση ή θα αρχίσουν να παρακολουθούνται ξανά. Τέλος, για κάθε ανίχνευση η οποία δεν έχει ανατεθεί σε κάποιο στόχο, αρχικοποιείται μία νέα ΜΔΑ και συνεπώς ένας νέος στόχος προς παρακολούθηση στα επόμενα καρέ. Αξίζει να σημειωθεί ότι οι στόχοι που παρακολουθούνται έχουν υψηλότερη προτεραιότητα σε σχέση με τους χαμένους στόχους, καθώς και οι ανιχνεύσεις που επικαλύπτονται από ανιχνεύσεις υπό παρακολούθηση στόχων αγνοούνται έτσι ώστε να μειωθούν οι αμφιβολίες κατά τη φάση αντιστοίχισης των δεδομένων.

ΚΕΦΑΛΑΙΟ 4 : ΠΕΡΙΓΡΑΦΗ ΜΕΘΟΔΟΛΟΓΙΑΣ ΠΑΡΑΚΟΛΟΥΘΗΣΗΣ ΠΟΛΛΑΠΛΩΝ ΑΝΤΙΚΕΙΜΕΝΩΝ

Σε αυτό το κεφάλαιο αναλύεται η μεθοδολογία παρακολούθησης πολλαπλών στόχων που σχεδιάστηκε και αναπτύχθηκε στο πλαίσιο της διπλωματικής εργασίας. Αρχικά θα πραγματοποιηθεί μία συνοπτική περιγραφή και έπειτα θα αναλυθούν περαιτέρω τα επιμέρους στάδια του αλγορίθμου. Σε πρώτο στάδιο θα σχολιαστούν τα βήματα προεπεξεργασίας που απαιτούνται και στη συνέχεια τον τρόπο υπολογισμού των κοστών αντιστοίχισης και της διαδικασίας αντιστοίχισης που χρησιμοποιείται από τον αλγόριθμο.

4.1. Κύρια δομή του αλγορίθμου

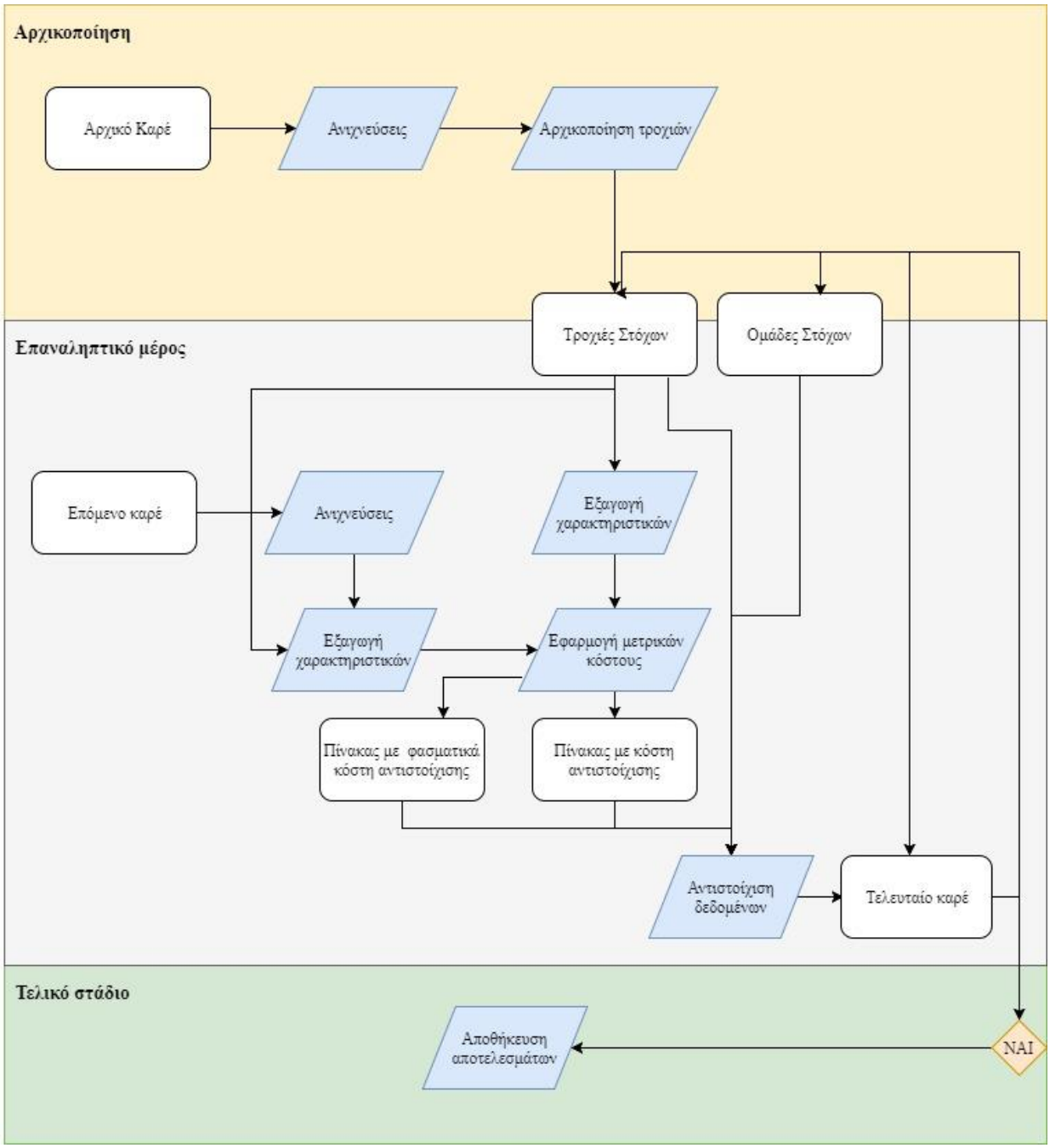
Ο σκοπός ενός αλγορίθμου παρακολούθησης πολλαπλών αντικειμένων είναι η ανίχνευση των κινούμενων αντικειμένων σε κάθε καρέ ενός βίντεο και η ορθή αντιστοίχιση των ταυτοτήτων τους σε διαφορετικά καρέ συμβάλλοντας με αυτό τον τρόπο στη δημιουργία ενός συνόλου τροχιών των στόχων του βίντεο. Η λογική πάνω στην οποία στηρίζεται ο αλγόριθμος που αναπτύχθηκε στη παρούσα εργασία είναι η "παρακολούθηση μέσω ανιχνεύσεων" (tracking-by-detection). Όπως παρουσιάζεται στην Εικόνα 4.1, αυτού του είδους οι αλγόριθμοι αρχικά αξιοποιούν τις ανιχνεύσεις των αντικειμένων που παρέχονται ως δεδομένα εισόδου ώστε να εξάγουν ορισμένα χαρακτηριστικά και να μοντελοποιήσουν τα αντικείμενα. Στη συνέχεια, πραγματοποιείται η αντιστοίχιση αυτών των αναπαραστάσεων των αντικειμένων με τις ανιχνεύσεις του επόμενου καρέ μέσω ενός αλγορίθμου αντιστοίχισης, ενώ έπειτα τα μοντέλα των αντικειμένων ανανεώνονται. Η διαδικασία είναι επαναληπτική για κάθε καρέ του βίντεο και σημειώνεται ότι σε αυτό το σύστημα πρέπει να περιλαμβάνεται ένας μηχανισμός γένεσης νέων τροχιών και τερματισμού ανενεργών τροχιών αντικειμένων.



Εικόνα 4.1 : Επισκόπηση της λογικής της "παρακολούθησης μέσω ανιχνεύσεων" [30]

Σε αυτή τη μεταπτυχιακή εργασία, λοιπόν, ακολουθείται μία τέτοια τύπου μεθοδολογία για την ανάπτυξη ενός αλγορίθμου παρακολούθησης πολλαπλών αντικειμένων, το διάγραμμα ροής του οποίου παρουσιάζεται στο Διάγραμμα 4.1 που ακολουθεί. Αναλυτικότερα, στο πρώτο καρέ του βίντεο πραγματοποιείται το στάδιο αρχικοποίησης του αλγορίθμου κατά το οποίο δημιουργούνται οι τροχιές των στόχων βάσει των ανιχνεύσεων που έχουν προκύψει. Έτσι, λοιπόν, υφίσταται ένα σύνολο τροχιών και ενδεχομένως δημιουργούνται ορισμένες «ομάδες» (groups) τροχιών που αφορούν στόχους που τα κουτιά περιγράμματός τους έχει κάποια επικάλυψη. Η λογική πάνω στην οποία βασίστηκε αυτή η ιδέα είναι ότι η κίνηση ενός συγκεκριμένου στόχου δεν εξαρτάται μόνο από τον ίδιο αλλά και από τη συμπεριφορά των όμορών του στόχων. Η υλοποίησή τους πρόκειται να σχολιαστεί αναλυτικότερα σε επόμενο εδάφιο αυτού του κεφαλαίου.

Στη συνέχεια, δεδομένων αυτών των τροχιών και των όποιων ομάδων τροχιών υφίστανται, ακολουθεί το επαναληπτικό στάδιο του αλγορίθμου. Αρχικά, εξάγονται ορισμένα χαρακτηριστικά για τις ανιχνεύσεις και στο αμέσως επόμενο καρέ εξάγονται τα ίδιου τύπου χαρακτηριστικά για τις προβλέψεις των νέων τροχιών. Σε αυτό το σημείο υπολογίζονται οι μετρικές κόστους με αποτέλεσμα να προκύπτει ένας πίνακας που εμπεριέχει τα κόστη αντιστοίχισης. Πρέπει να σημειωθεί ότι στη περίπτωση που σε ένα καρέ υπάρχουν ομάδες τροχιών τότε δημιουργείται επιπλέον ένας πίνακας που εμπεριέχει μόνο τα «φασματικά» κόστη αντιστοίχισης, δηλαδή τα κόστη αφορούν τις διαφορές μεταξύ των μοντέλων εμφάνισης των ανιχνεύσεων και των στόχων. Βάσει αυτών των πινάκων, λοιπόν, πραγματοποιείται η αντιστοίχιση των στόχων με τις ανιχνεύσεις μέσω του αλγορίθμου του Munkres [31]. Η διαδικασία επαναλαμβάνεται για όλα τα καρέ του βίντεο.



Διάγραμμα 4.1 : Διάγραμμα ροής του αλγορίθμου παρακολούθησης

4.2. Στάδια προεπεξεργασίας

Σε αυτό το εδάφιο πρόκειται να παρουσιαστούν αναλυτικά τα μοντέλα κίνησης και εμφάνισης των στόχων που χρησιμοποιήθηκαν για την ανάπτυξη του αλγορίθμου παρακολούθησης, καθώς και ένα σύνολο χαρακτηριστικών που προκύπτουν από αυτά.

4.2.1. Μοντέλο κίνησης

Τα μοντέλα κίνησης αποτελούν βασικό κομμάτι όλων των αλγορίθμων παρακολούθησης και αφορούν τη δυνατότητα πρόβλεψης της μελλοντικής κίνησης των στόχων βοηθώντας, έτσι, τον αλγόριθμο αντιστοίχισης. Στη βιβλιογραφία συναντώνται ποικίλα είδη μοντέλων κίνησης από πολύ απλοϊκά, όπως γραμμικές παλινδρομήσεις, έως πιο σύνθετα, όπως το φίλτρο Kalman ή προσεγγίσεις με LSTM νευρωνικά δίκτυα. Στη παρούσα εργασία για τη δυνατότητα πρόβλεψης της ταχύτητας του στόχου στο κάθε καρέ ώστε να είναι δυνατή η πρόβλεψη της μελλοντικής του θέσης, χρησιμοποιήθηκε τύπος ακρίβειας $3^{ης}$ τάξης που υπολογίσθηκε μέσω του αναπτύγματος Taylor. Η ταχύτητα κίνησης εκτιμάται ως συνδυασμός των τριών τελευταίων θέσεων του στόχου ως εξής :

$$V_n = \frac{3}{2}\binom{x_n}{y_n} - 2\binom{x_{n-1}}{y_{n-1}} + \frac{1}{2}\binom{x_{n-2}}{y_{n-2}}$$

Όπως αναφέρθηκε ήδη, ένα μοντέλο κίνησης επιτρέπει την πρόβλεψη της μελλοντικής θέσης του κάθε στόχου πριν την αντιστοίχισή του. Από τη πρόβλεψη αυτή προκύπτουν στα πλαίσια της εργασίας ένα σύνολο "κινηματικών" χαρακτηριστικών τα οποία είναι τα εξής :

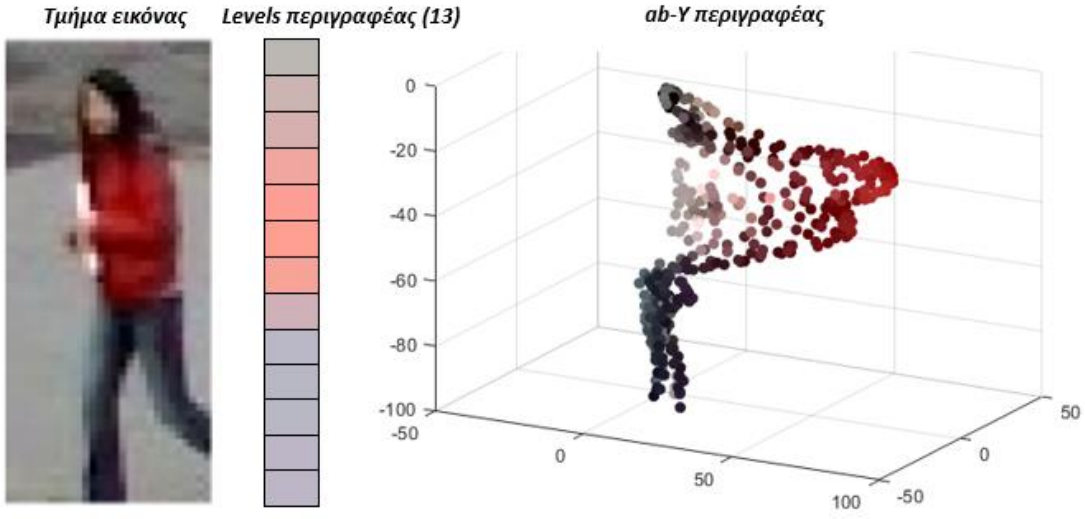
- **Μετάθεση στόχου ανάθεσης** : Το μήκος του διανύσματος μετάθεσης μεταξύ του προβλεπόμενου κεντροειδούς του κάθε στόχου και των κεντροειδών των ανιχνεύσεων.
- **Διαφορά εμβαδού** : Η απόλυτη τιμή της διαφοράς εμβαδού του κουτιού περιγράμματος κάθε στόχου από τα αντίστοιχα εμβαδά των ανιχνεύσεων διηρημένο από το εμβαδόν του στόχου.
- **Δείκτης προβλεπόμενης επικάλυψης στόχων** : Η μέγιστη % επικάλυψη της πρόβλεψης του κάθε στόχου με το σύνολο των λοιπών προβλέψεων των στόχων.
- **Δείκτης σχετικού προβλεπόμενου λόγου επικάλυψης** : Δίνεται από τον τύπο για κάθε d ανίχνευση :

$$RPOR(d) = \max_{i}\left(\frac{BBT_j \cap BBD_d}{BBT_i \cap BBD_d}\right), i \neq j$$

όπου BBT_i το προβλεπόμενο κουτί περιγράμματος του i-οστού στόχου και BBD_d το κουτί περιγράμματος της d-οστής ανίχνευσης. Να σημειωθεί ότι ο δείκτης προβλεπόμενης επικάλυψης στόχων ρυθμίζει και το ρυθμό μάθησης του μοντέλου εμφάνισης για το συγκεκριμένο καρέ και στόχο.

4.2.2. Μοντέλο εμφάνισης

Το μοντέλο εμφάνισης που χρησιμοποιήθηκε ονομάζεται “Levels” και αναπτύχθηκε στη διπλωματική του Β. Τσιρώνη [47]. Το μοντέλο αυτό στηρίζεται σε ένα άλλο μοντέλο εμφάνισης που αναπτύχθηκε σε εκείνη την εργασία που ονομάζεται “ab-Y”, βάση του οποίου αποτελεί ο χώρος χρώματος CIELAB. Συνοπτικά, το μοντέλο εμφάνισης “Levels” υπολογίζει τον μέσο όρο των χρωματικοτήτων a, b για δεδομένο εύρος της χωρικής συνιστώσας Y (του συστήματος της εικόνας) για το σύνολο των εικονοστοιχείων που εμπεριέχονται στο κουτί περιγράμματος του στόχου και της ανίχνευσης. Μετά από κατάλληλη διαμέριση του συνολικού εύρους της διεύθυνσης Y που ορίζεται από τον χρήστη προκύπτει ένα διάνυσμα περιγραφής που αποτελεί το τελικό προϊόν του μοντέλου. Μάλιστα, προκειμένου να αποφευχθούν τυχόν σφάλματα κατά τη διεύθυνση Y, χρησιμοποιείται κατάλληλη επικάλυψη μεταξύ των σταθμών. Αξίζει να σημειωθεί ότι η διαδικασία αυτή δεν εφαρμόζεται ακριβώς σε κάθε εικονοστοιχείο του κουτιού περιγράμματος, αλλά στα υπερεικονοστοιχεία (superpixels) που έχουν προκύψει από την εφαρμογή του αλγορίθμου SLIC [32]. Στην παρακάτω εικόνα παρουσιάζεται ένα παράδειγμα οπτικοποίησης των χαρακτηριστικών “Levels” και του χώρου “ab-Y” στον οποίο βασίζεται.



Εικόνα 4.2 : Οπτικοποίηση του μοντέλου εμφάνισης “Levels”

4.3. Κόστη αντιστοίχισης

Σε αυτό το εδάφιο πρόκειται να αναλυθεί η διαδικασία αντιστοίχισης στόχων και ανιχνεύσεων, η οποία σαφώς παίζει πολύ σημαντικό ρόλο για την ορθή λειτουργία ενός αλγορίθμου παρακολούθησης. Το πρόβλημα αυτό συνοψίζεται στην «ένα-προς-ένα» διαδικασία αντιστοίχισης του συνόλου των στόχων με το σύνολο των ανιχνεύσεων σε ένα καρέ του βίντεο ελαχιστοποιώντας το συνολικό κόστος αντιστοιχίσεων.

Σε πρώτη φάση θα ορίσουμε τον πίνακα κόστους. Έστω $c_{i,j}$ το κόστος αντιστοίχισης του i-οστού στόχου και της j-οστής ανίχνευσης, τότε ο πίνακας κόστους ή αλλιώς ο πίνακας αντιστοίχισης CM_{txd} ορίζεται ως εξής :

$$CM = \begin{pmatrix} c_{1,1} & \cdots & c_{1,d} \\ \vdots & \ddots & \vdots \\ c_{t,1} & \cdots & c_{t,d} \end{pmatrix}$$

όπου t το σύνολο των στόχων και d το σύνολο των ανιχνεύσεων. Ως αντιστοίχιση θεωρείται το σύνολο των t-στοιχείων του πίνακα κόστους, τα οποία ανήκουν το καθένα σε διαφορετική σειρά και στήλη. Η βέλτιστη αντιστοίχιση προκύπτει ως η αντιστοίχιση με το ελάχιστο κόστος, δηλαδή το άθροισμα των t-στοιχείων της κάθε αντιστοίχισης.

Στη παρούσα εργασία ο πίνακας κόστους που δημιουργείται βασίζεται σε παραμέτρους που περιγράφουν το μοντέλο εμφάνισης, το μοντέλο κίνησης και τις χωρικές αλληλεπιδράσεις που υφίστανται μεταξύ των στόχων και των ανιχνεύσεων. Κάθε στοιχείο του πίνακα κόστους προκύπτει ως ένας συνδυασμός των εξής 5 συναρτήσεων υποκόστους στόχων – ανιχνεύσεων :

Συνάρτηση (υπο-)κόστους	**Περιγραφή**
$c_1(t,d) = \ln(\|AM_t - AM_d\|_2)$	Φασματικό κόστος
$c_2(t,d) = -\ln(BB_t \cap BB_d)$	Κόστος επικάλυψης
$c_3(t,d) = \max\left(ln\left(RPOR(d)\right), -1\right)$	Κόστος σχετικής επικάλυψης
$c_4(t,d) = -\ln\left(1 - \delta E(t,d)\right)$	Κόστος αλλαγής εμβαδού
$c_5(t,d) = \min\left(max\left(-1, ln\left(\frac{Tr(t,d)}{10 * (FL_t + 1)}\right)\right)\right)$	Κόστος μετάθεσης

όπου **t** το σύνολο των στόχων, **d** το σύνολο των ανιχνεύσεων, $\mathbf{AM_t}$ και $\mathbf{AM_d}$ το μοντέλο εμφάνισης του t στόχου και της d ανίχνευσης αντίστοιχα, $\mathbf{BB_t}$ και $\mathbf{BB_d}$ το εμ-

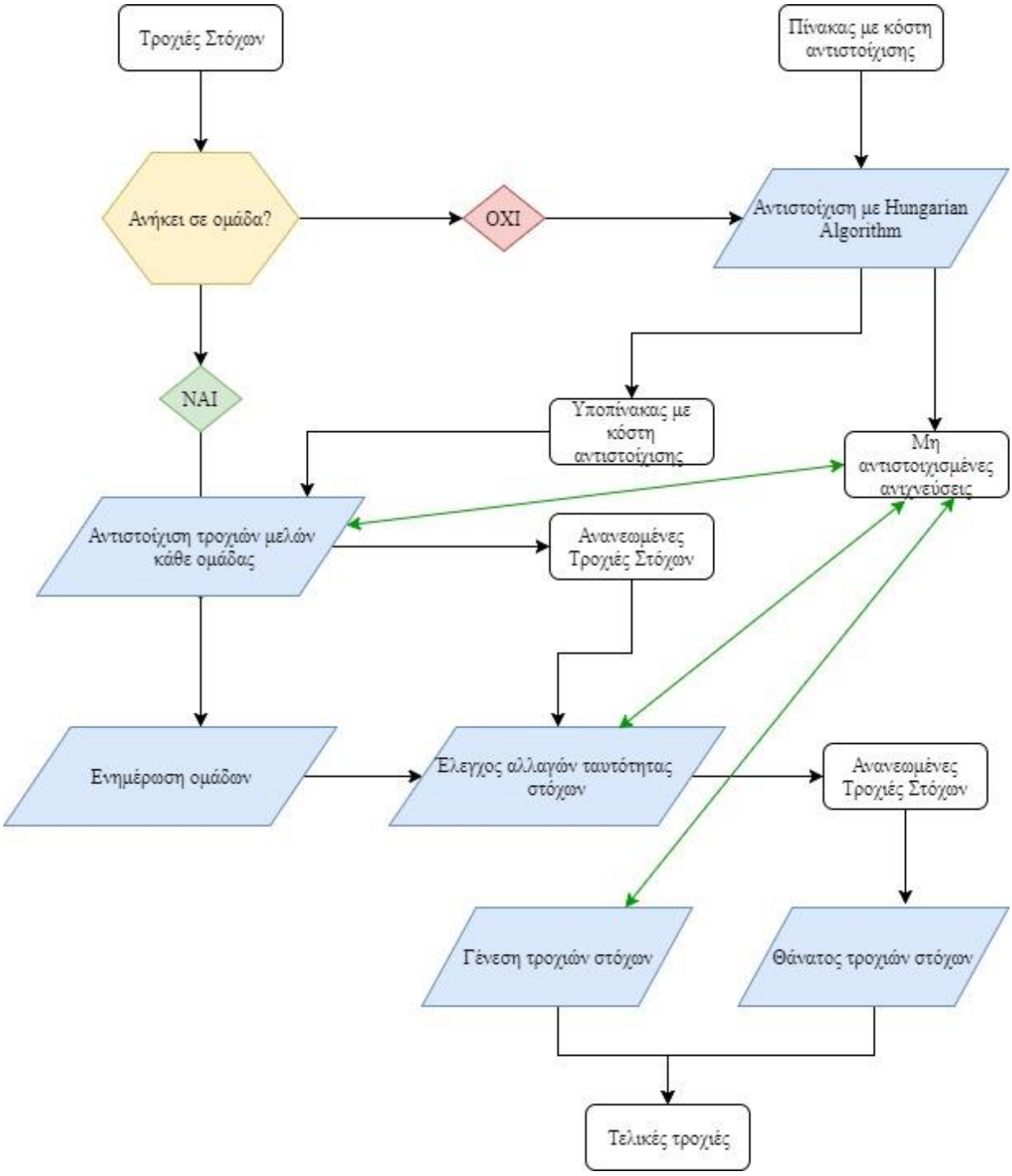
βαδόν του κουτιού περιγράμματος του t στόχου και της d ανίχνευσης, **RPOR(d)** ο δείκτης προβλεπόμενου λόγου επικάλυψης που αναφέρθηκε στο μοντέλο κίνησης, δE(t,d) η διαφορά εμβαδού που αναφέρθηκε στο μοντέλο κίνησης, **Tr(t,d)** η μετάθεση στόχου ανάθεσης που αναφέρθηκε επίσης στο μοντέλο κίνησης και $\mathbf{FL_t}$ το πλήθος των καρέ που ο t στόχος έχει «χαθεί». Αξίζει να σημειωθεί ότι χρησιμοποιήθηκαν ευριστικές μέθοδοι για τη δημιουργία των παραπάνω συναρτήσεων μετά από αρκετές δοκιμές σε διαφορετικά βίντεο. Ακόμη, αναφέρεται ότι όσον αφορά την μετρική c_1 η τιμή της απειρίζεται όταν υπερβεί την τιμή 4 ώστε να αποτρέπονται αντιστοιχίσεις υψηλού φασματικού κόστους.

Ο πίνακας κόστους υπολογίζεται ως εξής:

$$\boldsymbol{CM(t,d)} = c_1(t,d) + \min(\max(2c_2(t,d), -1), 1) + c_3(t,d) + \min(\max(c_4(t,d), -1), 1) + c_5(t,d)$$

4.4. Αντιστοίχιση Δεδομένων (Data Association)

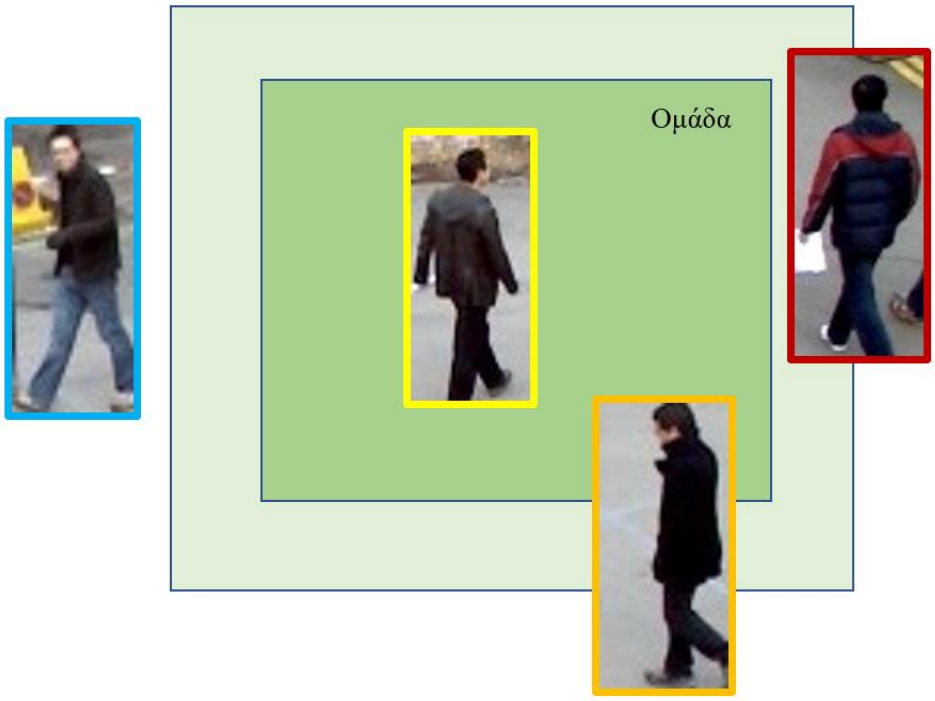
Σε αυτό το σημείο θα αναλυθεί εκ βάθους η διαδικασία αντιστοίχισης (data association) μεταξύ των τροχιών των στόχων και των ανιχνεύσεων που χρησιμοποιείται από τον αλγόριθμο παρακολούθησης που αναπτύχθηκε. Για το σκοπό αυτό υιοθετήθηκε μία σύνθετη διαδικασία αντιστοίχισης, η οποία επηρεάζεται από τις χωρικές αλληλεπιδράσεις των στόχων και πραγματοποιείται σε πολλαπλά βήματα. Το διάγραμμα ροής της διαδικασίας αντιστοίχισης περιγράφεται στο Διάγραμμα 4.2 και το τελικό της αποτέλεσμα είναι το ανανεωμένο σύνολο τροχιών και η ανανεωμένη πληροφορία ομαδοποίησής τους. Η διαδικασία πραγματοποιείται σε κάθε καρέ του βίντεο και για την έναρξή της απαιτείται η γνώση των τροχιών των στόχων και του πίνακα αντιστοίχισης.



Διάγραμμα 4.2 : Διάγραμμα ροής για τη διαδικασία αντιστοίχισης (data association) μεταξύ των ανιχνεύσεων και των κινούμενων στόχων

Αρχικά, διαχωρίζονται οι τροχιές σε αυτές που κατά το προηγούμενο καρέ ανήκουν σε κάποια ομάδα και σε αυτές που δεν ανήκουν σε κάποια ομάδα. Για τις τελευταίες, δηλαδή όσες για **όσες τροχιές δεν ανήκουν σε κάποια ομάδα**, η διαδικασία αντιστοίχισης πραγματοποιείται με τον αλγόριθμο του Munkres [31] οπότε προκύπτει το σύνολο των ανιχνεύσεων που έχουν αντιστοιχηθεί με κάποια τροχιά και απομένουν κάποιες ανιχνεύσεις που δεν έχουν αντιστοιχηθεί. Στη συνέχεια, **για κάθε ομάδα οι τροχιές που ανήκουν σε αυτή** αντιστοιχίζονται ξανά με τον ίδιο αλγόριθμο αντιστοίχισης στις εναπομείναντες ανιχνεύσεις.

Επόμενο στάδιο αποτελεί η **ανανέωση των ομάδων**. Σε πρώτη φάση υπολογίζονται οι επικαλυπτόμενες δυάδες τροχιών και είτε δημιουργούνται νέες ομάδες στη περίπτωση που καμία τροχιά της δυάδας δεν ανήκει σε κάποια ομάδα είτε προστίθενται σε υπάρχουσα ομάδα νέες τροχιές. Στη περίπτωση που και οι δύο τροχιές της δυάδας ανήκουν σε διαφορετικές ομάδες, τότε οι δύο ομάδες ενώνονται σε μία ενιαία.



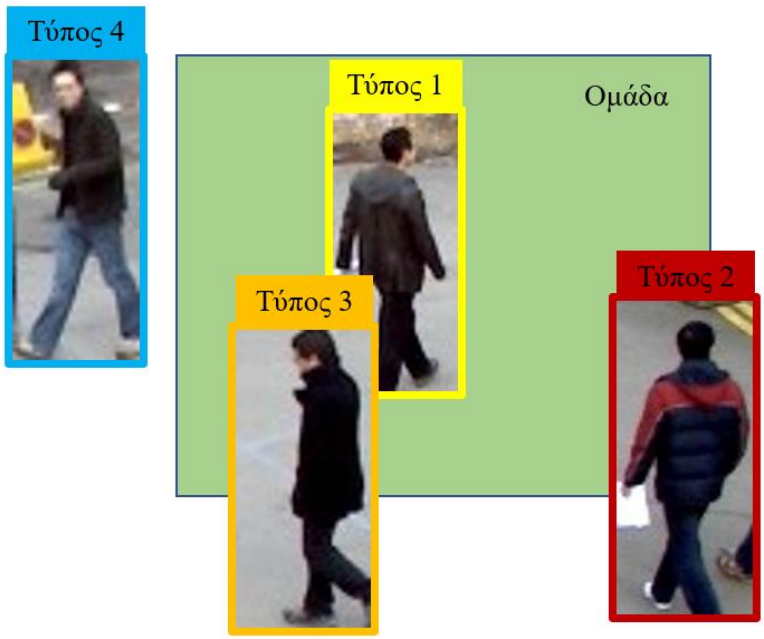
Εικόνα 4.3 : Οπτικοποίηση επαυξημένου κουτιού περιγράμματος ομάδας στόχων

Η διαδικασία που ακολουθεί είναι ο **έλεγχος αλλαγής ταυτότητας των τροχιών**, το διάγραμμα ροής της οποίας παρουσιάζεται στο Διάγραμμα 4.3. Ως πρώτο βήμα αποτελεί η εύρεση για κάθε ομάδα του αριθμού, έστω Κ, των τροχιών που έχουν «χαθεί». Έπειτα, για τις εναπομείναντες ανεξάρτητες ανιχνεύσεις υπολογίζονται ποιες τέμνουν το κουτί περιγράμματος της ομάδας επαυξημένο κατά μία σταθερά (Εικόνα 4.3). Για το παραπάνω σύνολο τροχιών-ανιχνεύσεων επιτελείται μία διαδικασία αντιστοίχισης με τον αλγόριθμο του Munkres. Ως δεύτερο βήμα αποτελεί η επιλογή των

Κ-καλύτερων αντιστοιχίσεων. Για το ανανεωμένο πλήθος τροχιών της κάθε ομάδας, πλέον, επαναντιστοιχίζονται οι τροχιές της ομάδας στις αντιστοιχισμένες ανιχνεύσεις με τη χρήση **μόνο** με φασματικά κριτήρια, δηλαδή του μοντέλου εμφάνισης τόσο των τροχιών όσο και των ανιχνεύσεων.

Ως τρίτο και τελευταίο βήμα εντός της κάθε ομάδας προσδιορίζεται ο «τύπος» της κάθε τροχιάς σύμφωνα με την Εικόνα 4.4. Οι τύποι τροχιάς είναι οι εξής τέσσερις:

i. Τύπος 1 : Οι τροχιές των οποίων το κουτί περιγράμματος βρίσκεται πλήρως εντός του κουτιού περιγράμματος της ομάδας.
ii. Τύπος 2 : Οι τροχιές των οποίων το κουτί περιγράμματος βρίσκεται μερικώς εντός του κουτιού περιγράμματος της ομάδας και δεν επικαλύπτονται με τροχιές τύπου 1.
iii. Τύπος 3 : Οι τροχιές των οποίων το κουτί περιγράμματος βρίσκεται μερικώς εντός του κουτιού περιγράμματος της ομάδας και επικαλύπτονται με τροχιές τύπου 1.
iv. Τύπος 4 : Οι τροχιές των οποίων το κουτί περιγράμματος δεν βρίσκεται εντός του κουτιού περιγράμματος της ομάδας.



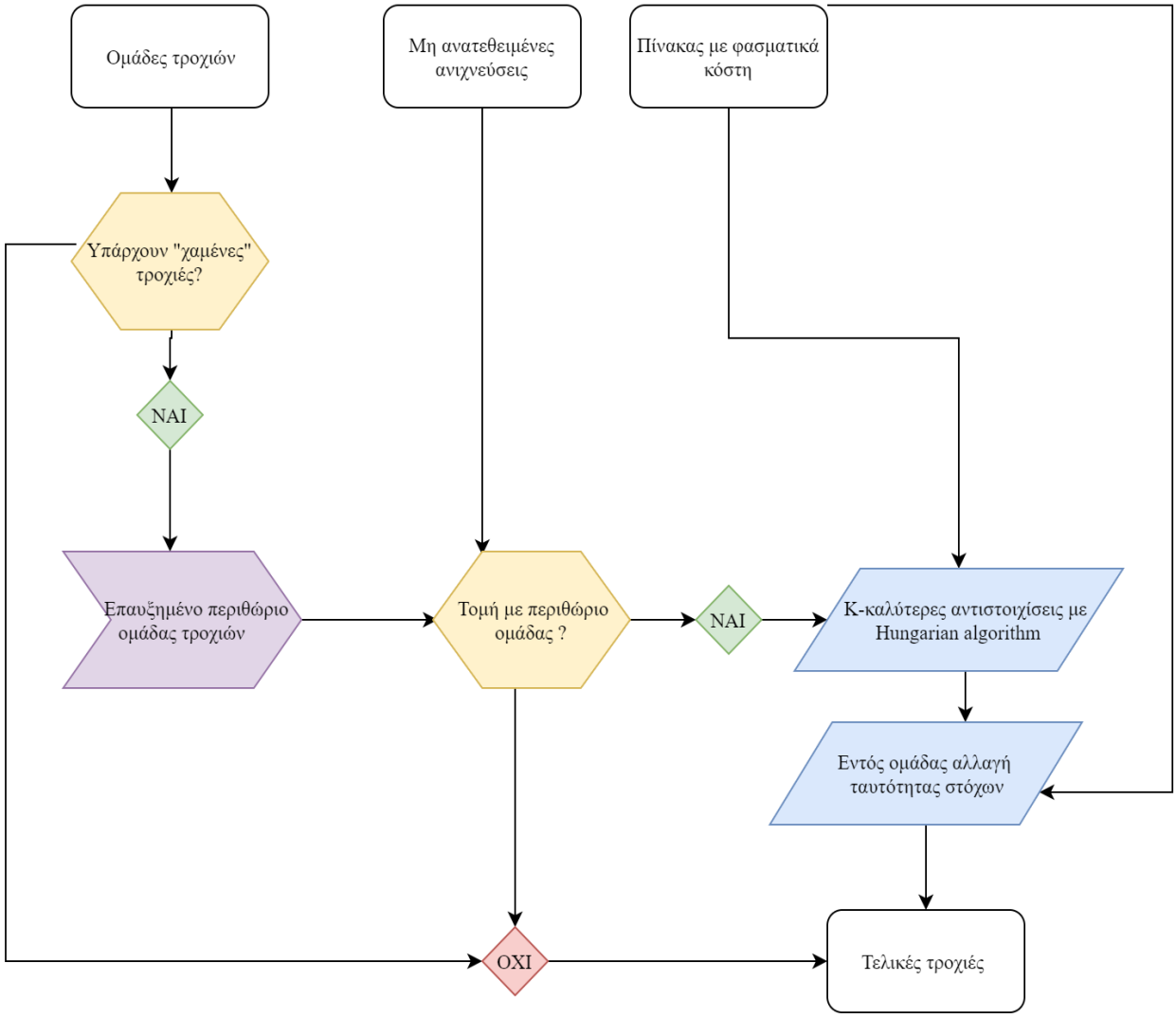
Εικόνα 4.4 : Τύποι τροχιών ομάδας

Το νέο κουτί περιγράμματος της ομάδας καθορίζεται από τις μη «χαμένες» τροχιές τύπου 1 και 3 άμεσα υποχρεώνοντάς τες να βρίσκονται πλήρως εντός του. Επίσης, για τις τροχιές τύπου 2 επιβάλλεται το νέο κουτί περιγράμματος της ομάδας να τις επικαλύπτει τουλάχιστον κατά 50%. Οι τροχιές τύπου 4 δεν συμμετέχουν στον καθορισμό του νέου κουτιού περιγράμματος της ομάδας και αφαιρούνται από μέλη της ομάδας.

Για αυτές τις τροχιές επανεξετάζεται το ενδεχόμενο αλλαγής ταυτότητας μέσω φασματικών κριτηρίων αποκλειστικά και αν προκύψει ανάγκη αλλαγής ταυτότητας αυτή γίνεται με αμοιβαία εναλλαγή ταυτότητας με το αντίστοιχο μέλος της ομάδας.

Σε αυτό το σημείο θα αναλυθεί ο μηχανισμός γένεσης νέων τροχιών και τερματισμός ανενεργών τροχιών που χρησιμοποιήθηκε σε αυτό τον αλγόριθμο. Όσον αφορά τη γένεση, για όσες εναπομείναντες ανιχνεύσεις δεν αντιστοιχήθηκαν με κάποια τροχιά και δεν επικαλύπτονται με κάποια άλλη τροχιά δημιουργείται ένας νέος στόχος. Ακολουθεί ένας έλεγχος ως προς τα φασματικά χαρακτηριστικά αυτού του στόχου σε σχέση με τα μοντέλα εμφάνισης των ανενεργών τροχιών και στη περίπτωση που αυτά συμφωνούν ανακτάται η ανενεργή τροχιά. Σε διαφορετική περίπτωση η νέα τροχιά διατηρείται και αν συνεχίσει την παρακολούθησή της για τουλάχιστον 4 καρέ τότε εντάσσεται πλήρως στο σύνολο των υπό παρακολούθηση τροχιών. Από την άλλη μεριά, ο τερματισμός των ανενεργών τροχιών επιτελείται αν ισχύει τουλάχιστον μία από τις παρακάτω συνθήκες :

- Η τροχιά παρακολουθείται για το πολύ 2 καρέ και θεωρείται «χαμένη» για τουλάχιστον 2 καρέ.
- Η τροχιά βρίσκεται στο περιθώριο της εικόνας και έχει «χαθεί» για τουλάχιστον 3 καρέ.
- Η τροχιά θεωρείται «χαμένη» για περισσότερα από 5 καρέ.



Διάγραμμα 4.3 : Διάγραμμα ροής για τη διαδικασία ελέγχου αλλαγής ταυτότητας (id switch) των στόχων

ΚΕΦΑΛΑΙΟ 5 : ΠΕΙΡΑΜΑΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ ΚΑΙ ΑΞΙΟΛΟΓΗΣΗ

Σε αυτό το κεφάλαιο παρουσιάζονται τα αποτελέσματα από την εφαρμογή του αλγορίθμου και η ποσοτική και ποιοτική αξιολόγηση. Αρχικά περιγράφονται τα σετ δεδομένων που χρησιμοποιήθηκαν για την πειραματική διαδικασία και στη συνέχεια τα αποτελέσματα για συγκεκριμένους αλγόριθμους και παραλλαγές τους που επιλέχτηκαν. Τέλος, περιγράφονται τα αποτελέσματα της ποσοτικής αξιολόγησης με χρήση καθιερωμένων δεικτών της βιβλιογραφίας και της ποιοτικής αξιολόγησης με αναφορά και περιγραφή συγκεκριμένων καρέ που έχουν ενδιαφέρον και αποτυπώνουν την τυπική συμπεριφορά και τα προβλήματα της μεθοδολογίας.

5.1. Σύνολο Δεδομένων Αξιολόγησης

Η ποσοτική αξιολόγηση μεθόδων παρακολούθησης πολλαπλών αντικειμένων/ στόχων έχει σημαντικό βαθμό δυσκολίας, κυρίως λόγω της δυσκολίας παραγωγής δεδομένων αναφοράς/ αληθείας (reference data), ήτοι χειροκίνητη ψηφιοποίηση ανά καρέ του βίντεο πολυάριθμων κινούμενων αντικειμένων. Επίσης, η χρονοβόρα και κοστοβόρα διαδικασία αυτή, είναι και σε πολλές περιπτώσεις ασαφής ως προς τον ακριβή προσδιορισμό των αντικειμένων (πχ. μερική απόκρυψη των στόχων, σκιάσεις, κλπ.) και σε αρκετές περιπτώσεις μπορεί διαφορετικοί φωτοερμηνευτές να παράξουν διαφορετικά, έως ένα βαθμό, δεδομένα αναφοράς. Επιπρόσθετα, η ύπαρξη διαφορετικών μέτρων σύγκρισης με ενδεχομένως ελεύθερες παραμέτρους μπορεί να οδηγήσει σε μη συγκρίσιμα αποτελέσματα που προκύπτουν από διαφορετικούς αλγορίθμους της βιβλιογραφίας.

Για τους παραπάνω λόγους, λοιπόν, την τελευταία δεκαετία η επιστημονική κοινότητα βασίζεται εκτεταμένα σε αξιολογήσεις που έχουν προκύψει από την εφαρμογή αλγορίθμων σε δεδομένα βίντεο που συνοδεύονται από ελεγμένα δεδομένα αναφοράς τα οποία ελευθέρα διατίθενται στην κοινότητα. Πιο συγκεκριμένα, ειδικά για το πρόβλημα της παρακολούθησης πολλαπλών αντικειμένων/ στόχων έχει δημιουργηθεί ένα πλήθος διαδικτυακών διαγωνισμών (benchmarks), δηλαδή ενιαίων πλατφορμών διάθεσης δεδομένων και αξιολόγησης, οι οποίοι επιτρέπουν στους συμμετέχοντες να υποβάλλουν όχι μόνο τους δικούς τους αλγορίθμους, αλλά επίσης τα δικά τους δεδομένα και νέες μεθοδολογίες αξιολόγησης που προτείνουν να προστεθούν στις ήδη υπάρχουσες ώστε να επωφεληθεί όλη η κοινότητα.

Ένας από τους πιο δημοφιλείς διαγωνισμούς είναι ο MOTC (Multiple Object Tracking Challenge) [1], ο οποίος διαθέτει ελεύθερα δεδομένα βίντεο, τα οποία και ανανεώνονται κάθε χρονιά, καθώς και μία συγκεκριμένη και ενιαία για όλους μεθοδολογία αξιολόγησης των αποτελεσμάτων. Στα δεδομένα αυτού του διαγωνισμού περι-

λαμβάνονται πολλαπλά βίντεο, τα αντίστοιχα δεδομένα αναφοράς και τα αποτελέσματα του αλγορίθμου εντοπισμού των στόχων που έχει εφαρμοστεί. Πρόκειται για βίντεο πεζών που έχουν ληφθεί σε διαφορετικά μέρη και το καθένα από αυτά έχει διαφορετικό επίπεδο δυσκολίας, ενώ κάθε χρονιά ο διαγωνισμός ανανεώνεται με νέα βίντεο. Αξίζει να σημειωθεί ότι ο προκαθορισμένος αλγόριθμος εντοπισμού των στόχων που χρησιμοποιείται είναι ο ACF (Aggregated Channel Features) [7], ωστόσο ο διαγωνισμός δίνει τη δυνατότητα στους συμμετέχοντες να υποβάλλουν τα αποτελέσματα των αλγορίθμων τους για τα οποία έχει χρησιμοποιηθεί ένας άλλος αλγόριθμος εντοπισμού. Αυτός ο διαγωνισμός είναι ο πλέον δημοφιλής και θεωρείται ως ένας από τους πιο δύσκολους στην επιστημονική κοινότητα λόγω της πληθώρας των βίντεο με διαφορετική γωνία λήψης και επίπεδα πυκνότητας των πεζών και συνθηκών φωτισμού που διαθέτει.

Σε αυτή τη διπλωματική τα πειράματα που πραγματοποιήθηκαν αφορούν τα δεδομένα εκπαίδευσης του 2DMOT15, τα οποία αφορούν τον διαγωνισμό MOTC για το 2015. Τα αναλυτικά στοιχεία του κάθε βίντεο που περιλαμβάνεται παρουσιάζονται στον Πίνακα 5.1 :

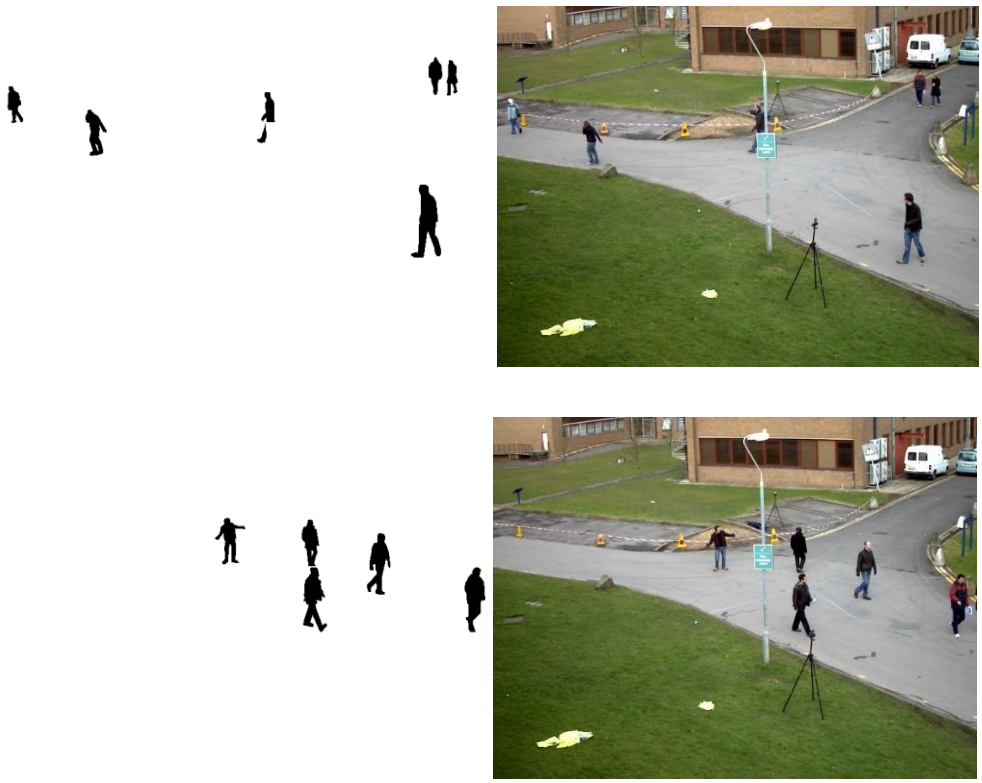
Βίντεο εκπαίδευσης											
Όνομα	**FPS**	**Ανάλυση**	**Διάρκεια (sec)**	**Στόχοι**	**Boxes**	**Πυκνότητα**	**3D**	**Είδος κάμερας**	**Viewpoint**	**Συνθήκες καιρού**	**Πηγή**
TUD-Stadmitte	25	640x480	179 (00:07)	10	1156	6.5	ναι	στατική	μέσο	συννεφώδης	[2]
TUD-Campus	25	640x480	71 (00:03)	8	359	5.1	όχι	στατική	μέσο	συννεφώδης	[3]
PETS09-S2L1	7	768x576	795 (01:54)	19	4476	5.6	ναι	στατική	υψηλό	συννεφώδης	[4]
ETH-Bahnhof	14	640x480	1000 (01:11)	171	5415	5.4	ναι	κινούμενη	χαμηλό	συννεφώδης	[5]
ETH-Sunnyday	14	640x480	354 (00:25)	30	1858	5.2	ναι	κινούμενη	χαμηλό	ηλιόλουστος	[5]
ETH-Pedcross2	14	640x480	840 (01:00)	133	6263	7.5	όχι	κινούμενη	χαμηλό	ηλιόλουστος	[5]
ADL-Rundle-6	30	1920x1080	525 (00:18)	24	5009	9.5	όχι	στατική	χαμηλό	συννεφώδης	νέο
ADL-Rundle-8	30	1920x1080	654 (00:22)	28	6783	10.4	όχι	κινούμενη	μέσο	νυχτερινός	νέο
KITTI-13	10	1242x375	340 (00:34)	42	762	2.2	όχι	κινούμενη	μέσο	ηλιόλουστος	[6]
KITTI-17	10	1242x370	145 (00:15)	9	683	4.7	όχι	στατική	μέσο	ηλιόλουστος	[6]
Venice-2	30	1920x1080	600 (00:20	26	7141	11.9	όχι	στατική	μέσο	ηλιόλουστος	νέο
Συνολικά			**5503 (06:29)**	**500**	**39905**	**7.3**					

Πίνακας 5.1 : Αναλυτικά στοιχεία των δεδομένων εκπαίδευσης του 2DMOT15

5.2. Πειραματικά αποτελέσματα με εφαρμογή του αναπτυγμένου αλγορίθμου

Στα πλαίσια της εργασίας πραγματοποιήθηκαν ποικίλα πειράματα προκειμένου να αξιολογηθεί ο ανεπτυγμένος αλγόριθμος *(Αλγόριθμος Α, Αλ-Α)* παρακολούθησης πολλαπλών αντικειμένων/ στόχων. Αρχικά, ο αλγόριθμος αυτός εφαρμόστηκε σε όλα τα βίντεο εκπαίδευσης του MOTC που αναφέρθηκε προηγουμένως. Ακόμη, προκειμένου να εξεταστεί η σημασία τους στη βελτίωση της διαδικασίας της παρακολούθησης, αποφασίστηκε να διενεργηθεί μία σημασιολογική κατάτμηση κάθε καρέ σε ένα από τα διαθέσιμα βίντεο του σετ δεδομένων. Συγκεκριμένα, δημιουργήθηκαν 400 μάσκες για

το βίντεο PETS09-S2L1 που διαχωρίζει τους πεζούς από το παρασκήνιο (background) της εικόνας και εφαρμόστηκε ο ίδιος αλγόριθμος χωρίς καμία μετατροπή στις παραμέτρους του.



Εικόνα 5.1 : Ενδεικτικές μάσκες πεζών για κάποια καρέ του βίντεο PETS09-S2L1

Ωστόσο, προκειμένου να αξιολογηθεί ορθότερα η αποδοτικότητα του αλγορίθμου, αποφασίστηκε να εφαρμοστούν για τα ίδια δεδομένα ο state-of-the-art αλγόριθμος παρακολούθησης MDP [8] που έχει αναλυθεί σε προηγούμενο κεφάλαιο και να αναπτυχθεί ένας ακόμη αλγόριθμος παρακολούθησης. Όσον αφορά τον MDP *(Αλγόριθμος Β, Αλ-Β),* αξίζει να σημειωθεί ότι η λειτουργία του επιτάσσει τον καθορισμό ενός σετ εκπαίδευσης και ενός σετ εφαρμογής. Στα πειράματα που πραγματοποιήθηκαν ως σετ εκπαίδευσης ορίστηκαν τα βίντεο “ETH-Bahnhof”, “ETH-Sunnyday”, “ADL-Rundle-6”, “TUD-Campus”, “KITTI-17”, “Venice-2” και ως σετ δοκιμής τα “TUD-Stadmitte”, “ETH-Pedcross2”, “PETS09-S2L1”, “ADL-Rundle-8”, “KITTI-13”. Σημειώνεται ότι στη συνέχεια τα σετ δοκιμής και εκπαίδευσης εναλλάχθηκαν ώστε να προκύψουν αποτελέσματα για όλα τα βίντεο.

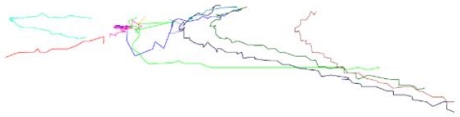
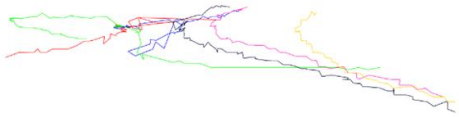
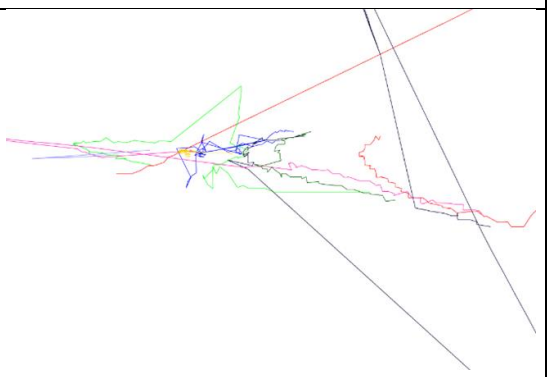
Όσον αφορά τον τρίτο αλγόριθμο (*Αλγόριθμος Γ, Αλ-Γ)* που αναπτύχθηκε στο πλαίσιο της αξιολόγησης, πρόκειται για μία απλή υλοποίηση αυτόματου εντοπισμού και παρακολούθησης πολλαπλών στόχων που βασίζεται στην κίνηση των αντικειμένων

(motion-based), η οποία διατίθεται από στις βιβλιοθήκες του λογισμικό της MATLAB (https://www.mathworks.com/help/vision/examples/motion-based-multiple-object-tracking.html). Στον πρωτότυπο αλγόριθμο ο εντοπισμός των κινούμενων αντικειμένων πραγματοποιείται μέσω της εξαγωγής του παρασκηνίου της εικόνας (background subtraction) που βασίζεται σε Γκαουσιανά μοντέλα ανάμειξης (Gaussian Mixture Models). Αφού εφαρμοστεί μία σειρά μορφολογικών φίλτρων στο προσκήνιο (foreground) της εικόνας για την αποφυγή θορύβου, πραγματοποιείται μία διαδικασία εξαγωγής blobs ώστε να εντοπιστούν σύνολα εικονοστοιχείων που είναι πιθανόν να αντιστοιχούν σε κινούμενα αντικείμενα. Ωστόσο, προκειμένου τα αποτελέσματα του αλγορίθμου να είναι άμεσα συγκρίσιμα με τους υπόλοιπους αλγορίθμους, αποφασίστηκε η αντικατάσταση της ανίχνευσης μέσω αφαίρεσης υποβάθρου από τον αλγόριθμο ανίχνευσης ACF. Η αντιστοίχιση των στόχων με τα εντοπισμένα αντικείμενα (detections) βασίζεται αποκλειστικά στην κίνηση. Για κάθε στόχο εκτιμάται η κίνησή του από ένα φίλτρο Kalman, το οποίο χρησιμοποιείται ώστε να προβλέψει την τοποθεσία του κάθε στόχου σε κάθε καρέ και να καθορίσει την πιθανότητα του κάθε εντοπισμένου αντικειμένου να αντιστοιχεί με κάποιο στόχο.

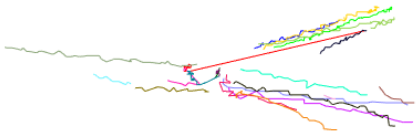
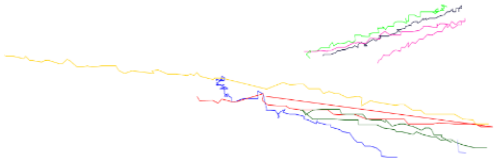
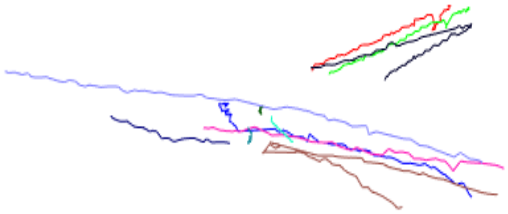
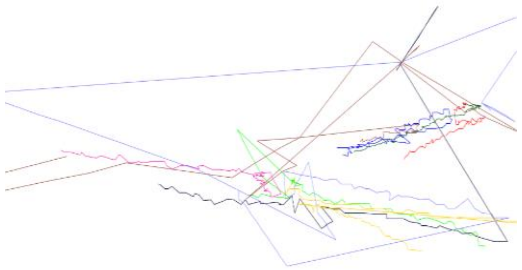
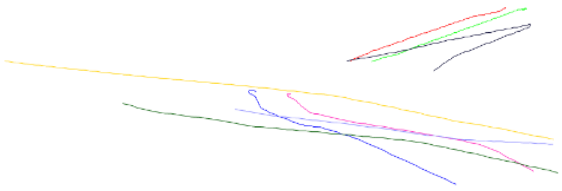
Στη συνέχεια ακολουθούν ορισμένα διαγράμματα των τροχιών που καταγράφουν οι στόχοι για κάθε αλγόριθμο που εφαρμόστηκε και εκείνο που προκύπτει από τα δεδομένα αληθείας. Συγκεντρωτικά, οι αλγόριθμοι που εφαρμόστηκαν είναι :

i. Ο αλγόριθμος που αναπτύχθηκε σε αυτή την εργασία με την προσθήκη σημασιολογικών μασκών *[Αλ-Am]*
ii. Ο αλγόριθμος που αναπτύχθηκε σε αυτή την εργασία χωρίς την προσθήκη σημασιολογικών μασκών *[Αλ-Α]*
iii. Ο state-of-the-art αλγόριθμος MDP της διεθνούς βιβλιογραφίας *[Αλ-Β]*
iv. Ο απλοϊκός αλγόριθμος παρακολούθησης που βασίζεται μόνο στην κίνηση των αντικειμένων *[Αλ-Γ]*

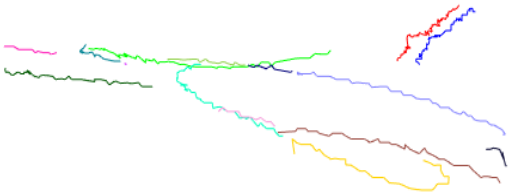
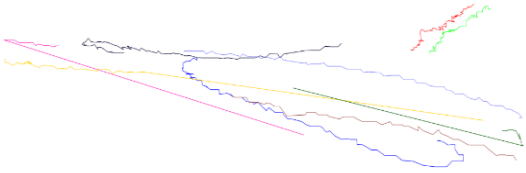
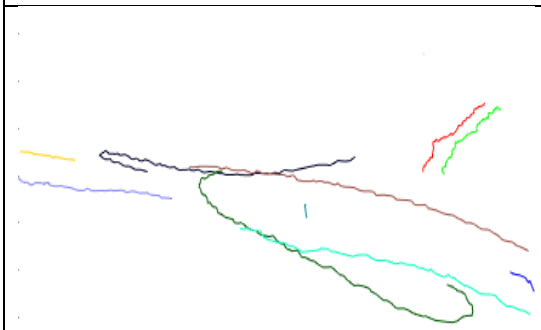
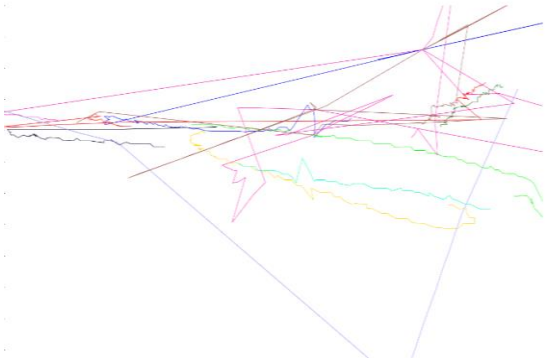
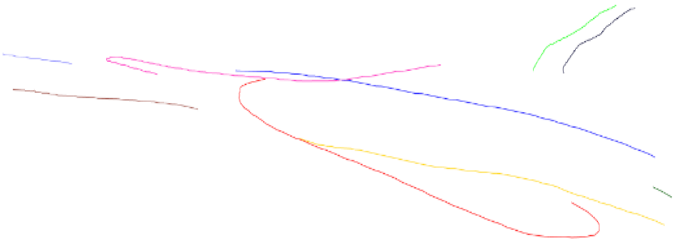
Τα τέσσερα σετ διαγραμμάτων που ακολουθούν αναφέρονται στο διάγραμμα που προκύπτει από τα δεδομένα αληθείας και από τους τέσσερις αλγορίθμους που εφαρμόστηκαν για το ίδιο χρονικό διάστημα του ίδιου βίντεο ώστε να είναι άμεσα συγκρίσιμα τα αποτελέσματά τους.

[Αλ-Am]	***[Αλ-Α]***
	
[Αλ-Β]	***[Αλ-Γ]***
	
Δεδομένα (τροχιές) αναφοράς	
	

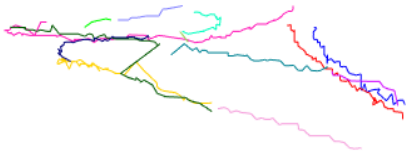
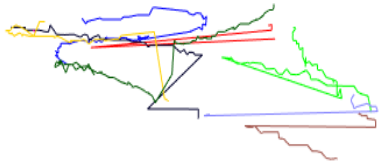
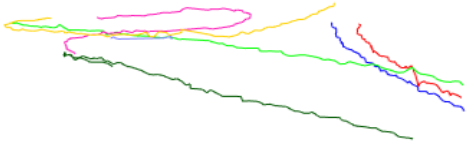
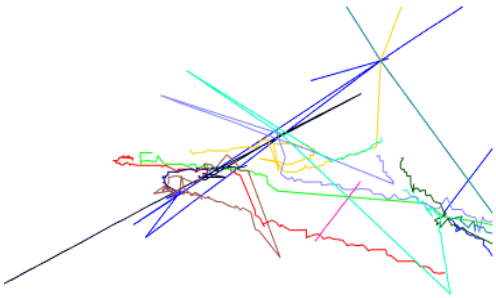
*Διάγραμμα 5.1 : Τα διαγράμματα τροχιών για τα καρέ **1 έως 100** για το βίντεο PETS09-S2L1 που προέκυψαν από τους Αλγορίθμους Αλ-Α, Αλ-Α2, Αλ-Β, Αλ-Γ, καθώς επίσης και οι αληθείς τροχιές αναφοράς (κάτω).*

[Αλ-Am]	***[Αλ-Α]***
	
[Αλ-Β]	***[Αλ-Γ]***
	
Δεδομένα (τροχιές) αναφοράς	
	

*Διάγραμμα 5.2 : Τα διαγράμματα τροχιών για τα καρέ **101 έως 200** για το βίντεο PETS09-S2L1 που προέκυψαν από τους Αλγορίθμους Αλ-Α, Αλ-Α2, Αλ-Β, Αλ-Γ, καθώς επίσης και οι αληθείς τροχιές αναφοράς (κάτω).*

[Αλ-Am]	***[Αλ-Α]***
	
[Αλ-Β]	***[Αλ-Γ]***
	
Δεδομένα (τροχιές) αναφοράς	
	

*Διάγραμμα 5.3 : Τα διαγράμματα τροχιών για τα καρέ **201 έως 300** για το βίντεο PETS09-S2L1 που προέκυψαν από τους Αλγορίθμους Αλ-Α, Αλ-Α2, Αλ-Β, Αλ-Γ, καθώς επίσης και οι αληθείς τροχιές αναφοράς (κάτω).*

[Αλ-Am]	***[Αλ-Α]***
	
[Αλ-Β]	***[Αλ-Γ]***
	
Δεδομένα (τροχιές) αναφοράς	
	

*Διάγραμμα 5.4 : Τα διαγράμματα τροχιών για τα καρέ **301 έως 400** για το βίντεο PETS09-S2L1 που προέκυψαν από τους Αλγορίθμους Αλ-Α, Αλ-Α2, Αλ-Β, Αλ-Γ, καθώς επίσης και οι αληθείς τροχιές αναφοράς (κάτω).*

Σύμφωνα με τα παραπάνω διαγράμματα, παρατηρείται ποιοτικά ότι ο αλγόριθμος που αναπτύχθηκε με τη προσθήκη ή μη δεδομένων κατάτμησης (*Αλ-Α*) ανταποκρίνεται ικανοποιητικά στις ιδιαιτερότητες του προβλήματος, του συγκεκριμένου βίντεο και της διάταξης των κινούμενων αντικειμένων. Η μορφή των τροχιών των στόχων στις περισσότερες περιπτώσεις ομοιάζει με αυτές των δεδομένων αληθείας με τον αριθμό των αλλαγών ταυτότητας (id-switches) μεταξύ των στόχων να είναι περιορισμένος. Ακόμη, το μήκος των τροχιών είναι σε γενικές γραμμές μεγάλο και άμεσα συγκρίσιμο με το αληθές μήκος, γεγονός που υποδεικνύει μία χρονική συνέχεια και συνέπεια στη παρακολούθηση του ίδιου στόχου χωρίς διακοπές ή εναλλαγές.

Υπάρχουν, όμως, και καρέ όπου αλλάζει η ομοτοπία των τροχιών σε σχέση με τις αληθείς. Οι περιοχές αυτές αποτελούνται από αλλαγές ταυτότητας των στόχων και είναι σφάλματα του αλγορίθμου, τα οποία σε μεγάλο βαθμό οφείλονται σε αποκρύψεις των στόχων και έλλειψη ποιοτικών ανιχνεύσεων. Μία σαφής βελτίωση των αποτελεσμάτων παρατηρείται με την ενσωμάτωση των δεδομένων κατάτμησης διότι ισχυροποιείται η περιγραφική ικανότητα του μοντέλου εμφάνισης της κάθε τροχιάς. Οι κύριες διαφοροποιήσεις παρατηρήθηκαν σε λιγότερες λανθασμένες αλλαγές ταυτότητας και σε αυξημένη δυνατότητα του αλγορίθμου για επαναφορά μιας «χαμένης» τροχιάς (κινούμενου αντικειμένου). Αντιθέτως, στη περίπτωση της μη ενσωμάτωσης δεδομένων κατάτμησης (*Αλ-Α2*) παρατηρήθηκε έντονα το φαινόμενο απώλειας της ταυτότητας μιας τροχιάς και γένεσης στη θέση της μιας νέας, απόρροια της ελλιπούς πληροφορίας.

Εν συγκρίσει με τους άλλους δύο αλγορίθμους, δηλαδή την απλή υλοποίηση (*Αλ-Γ*) και τον MDP (*Αλ-Β*), παρατηρείται ότι ο *Αλ-Γ* αποτυγχάνει πλήρως στη δημιουργία ορθών τροχιών, ενώ ο *Αλ-Β* σε γενικές γραμμές παρουσιάζει χειρότερη «εικόνα» σε σχέση με τις υλοποιήσεις της παρούσας εργασίας. Το γεγονός αυτό πιθανώς οφείλεται στα μειωμένα δεδομένα εκπαίδευσης για τον αλγόριθμο του MDP αφού πρόκειται για έναν από τους state-of-the-art αλγορίθμους στην παρακολούθηση πολλαπλών αντικειμένων. Στο επόμενο εδάφιο, περιγράφονται τα αντίστοιχα ποσοτικά αποτελέσματα της αξιολόγησης για το συγκεκριμένο βίντεο και ελέγχεται η επίδοση των αλγορίθμων στα υπόλοιπα βίντεο του σετ δεδομένων.

5.3. Ποσοτική Αξιολόγηση

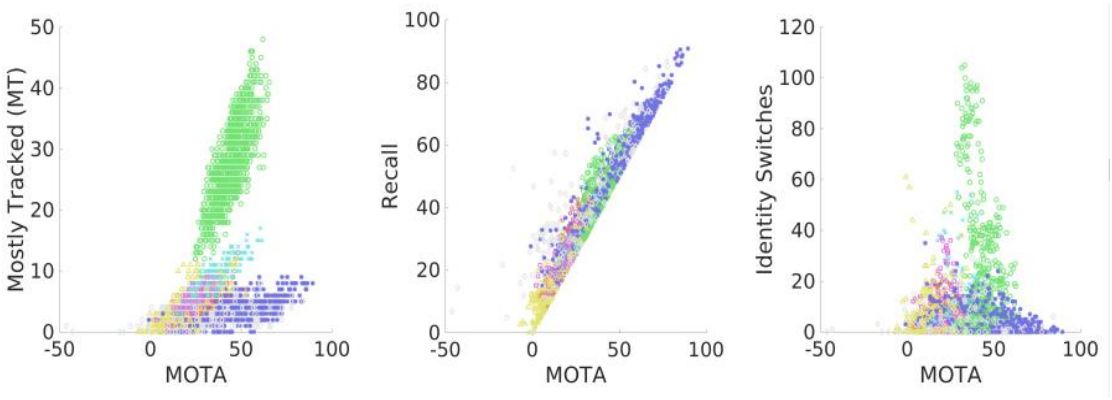
Η ποσοτική αξιολόγηση των αλγορίθμων πραγματοποιήθηκε με βάση το MOTC και τις μετρικές CLEAR [9]. Ως πιο αντιπροσωπευτικοί δείκτες για τη συνολική επίδοση των αλγορίθμων παρακολούθησης θεωρούνται οι MOTA (MOT Accuracy), MOTP (MOT Precision) και MOTAL (MOT Accuracy with Log_{10} (ID Switches)). Πιο συγκεκριμένα, οι μαθηματικοί τύποι για τους δείκτες MOTA και MOTP ακολουθούν στη συνέχεια :

$$\mathrm{MOTA} = 1 - \frac{\sum_t (\mathrm{FN}_t + \mathrm{FP}_t + \mathrm{IDSW}_t)}{\sum_t \mathrm{GT}_t}$$

$$\mathrm{MOTP} = \frac{\sum_{t,i} d_{t,i}}{\sum_t c_t},$$

όπου t είναι ο αριθμός των καρέ του βίντεο, GT ο αριθμός των δεδομένων αληθείας, c_t ο αριθμός των αντιστοιχίσεων στο t καρέ και $d_{t,i}$ η ποσοστιαία επικάλυψη του κουτιού περιγράμματος (bounding box) με το GT αντικείμενο, $IDSW_t$ ο αριθμός των αλλαγών ταυτότητας για το t καρέ, FN_t το πλήθος των τροχιών που δεν βρέθηκαν, FP_t το πλήθος των λάθος αντιστοιχίσεων μεταξύ ανιχνεύσεων και στόχων.

Περιγραφικά, ο δείκτης MOTA αναφέρεται στην ακρίβεια ενός αλγορίθμου παρακολούθησης και αποτελεί μία καλή ένδειξη για την συνολική επίδοσή του. Σημειώνεται ότι αυτός ο δείκτης μπορεί να βγει αρνητικός σε περίπτωση που τα λάθη του αλγορίθμου υπερβαίνουν τον αριθμό των αντικειμένων στη σκηνή. Όσον αφορά τον δείκτη MOTP, εκφράζει τη μέση ανομοιότητα μεταξύ όλων των True Positives (TP) και των αντίστοιχων GT αντικειμένων και επηρεάζεται κυρίως από τις ανιχνεύσεις και λιγότερο από τα πραγματικά αποτελέσματα του αλγορίθμου παρακολούθησης. Από διάφορα πειράματα ερευνητών έχει αποφανθεί ότι ο δείκτης MOTA συμφωνεί κατά ένα μεγάλο βαθμό με την ανθρώπινη γνώμη που προκύπτει από τα οπτικά αποτελέσματα ενός αλγορίθμου παρακολούθησης. Ακολουθούν ορισμένα διαγράμματα που παρουσιάζουν τις συσχετίσεις του δείκτη MOTA με τους δείκτες MT (Mostly Tracked), Recall και ID Switches.



Διάγραμμα 5.5 : Συσχετίσεις μεταξύ του δείκτη MOTA και των δεικτών MT, Accuracy και ID Switches

Στους παρακάτω πίνακες παρουσιάζονται οι μετρικές MOTA, MOTP και MOTAL του αλγορίθμου παρακολούθησης που αναπτύχθηκε στη παρούσα εργασία *[Αλ-Α]*, του απλοϊκού αλγορίθμου που βασίζεται αποκλειστικά στην κίνηση των αντικειμένων *[Αλ-Γ]* και του αλγορίθμου MDP *[Αλ-Β]* για όλο το σετ δεδομένων.

Αποτελέσματα Αλγόριθμου Α [Αλ-Α]				
#	**Βίντεο**	**MOTA**	**MOTP**	**MOTAL**
1	TUD-Stadmitte	51,1	64,8	53
2	TUD-Campus	14,5	71	18,6
3	PETS09-S2L1 **(με μάσκα)**	60,4	71,3	61,6
4	PETS09-S2L1	55,7	70,8	58,9
5	ETH-Bahnhof	4,7	71,7	5,6
6	ETH-Sunnyday	28,6	74,6	30,4
7	ETH-Pedcross2	2	68,9	3,2
8	ADL-Rundle-6	2,4	70,9	6
9	ADL-Rundle-8	7,3	71,2	8,3
10	KITTI-13	1,3	70,9	1,9
11	KITTI-17	30,9	70,1	34,5
12	Venice-2	4,8	72,4	6,1
	Συνολικά	**18,5**	**70,7**	**20,6**

Αποτελέσματα Αλγόριθμου Β [Αλ-Β]				
#	**Βίντεο**	**MOTA**	**MOTP**	**MOTAL**
1	TUD-Stadmitte	62,3	65,7	62,8
2	TUD-Campus	53,5	71,5	55,7
3	PETS09-S2L1	-28,6	72,3	-27,8
4	ETH-Bahnhof	20,5	74	22,3
5	ETH-Sunnyday	46,7	76,5	47,4
6	ETH-Pedcross2	11,3	70,8	11,6
7	ADL-Rundle-6	31,7	73,0	32,4
8	ADL-Rundle-8	15,1	72,6	15,5
9	KITTI-13	3,4	71,2	4,1
10	KITTI-17	60,0	71,6	60,8
11	Venice-2	12,9	74,2	13,5
	Συνολικά	**26,3**	**72,1**	**27,1**

Αποτελέσματα Αλγόριθμου Γ [Αλ-Γ]				
#	**Βίντεο**	**MOTA**	**MOTP**	**MOTAL**
1	TUD-Stadmitte	24,3	65,1	26,6
2	TUD-Campus	10,6	70,5	14,2
3	PETS09-S2L1	40,9	70,9	43,2
4	ETH-Bahnhof	-20,6	72,1	-16
5	ETH-Sunnyday	24,6	74,2	27
6	ETH-Pedcross2	-5,8	69,3	-5,1
7	ADL-Rundle-6	-29,7	70,9	-26,4
8	ADL-Rundle-8	-52,3	71,5	-50,3
9	KITTI-13	-163,8	70,8	-162,1
10	KITTI-17	10,1	70,9	11,7
11	Venice-2	-17,7	72,3	-15,4
	Συνολικά	**-16,3**	**70,8**	**-13,9**

Η ποσοτική αξιολόγηση αναδεικνύει τα πλεονεκτήματα και τα μειονεκτήματα της κάθε υλοποίησης. Τα αποτελέσματα του αλγορίθμου που αναπτύχθηκε σε αυτή την εργασία ως σύνολο μπορούν να θεωρηθούν ικανοποιητικά, όμως παρουσιάζουν υψηλή μεταβλητότητα με το αν η κάμερα είναι σταθερή ή κινούμενη. Συγκεκριμένα, στη περίπτωση της σταθερής κάμερας η επίδοση του αλγορίθμου είναι αρκετά καλή, ιδιαίτερα στο βίντεο "PETS09-S2L1" που χαρακτηρίζεται από την υψηλή τοποθέτηση της κάμερας. Κάτι τέτοιο είναι αναμενόμενο αφού ο αλγόριθμος αναπτύχθηκε για την παρακολούθηση πολλαπλών στόχων με παρόμοια γεωμετρία λήψης. Αξίζει να σημειωθεί ότι η προσθήκη δεδομένων κατάτμησης βελτιώνει τις μετρικές κατά περίπου 10% επαληθεύοντας το ποιοτικό συμπέρασμα που εξήχθη από τα διαγράμματα τροχιών που προηγήθηκαν. Όσον αφορά τις περιπτώσεις βίντεο με κινούμενη κάμερα, τα αποτελέσματα δεν είναι ικανοποιητικά και εξηγούνται εν μέρει από την απλοϊκότητα του μοντέλου κίνησης που χρησιμοποιήθηκε. Σχετικά με την μετρική MOTP που υποδεικνύει την ακρίβεια χωρικού προσδιορισμού της τροχιάς, παρατηρούμε ότι τα αποτελέσματα βρίσκονται σε πολύ καλό επίπεδο περί το 70% και άμεσα συγκρίσιμα με του state-of-the-art αλγορίθμου MDP.

Όσον αφορά τον πιο απλοϊκό αλγόριθμο *[Αλ-Γ]* τα αποτελέσματα, όπως ήταν αναμενόμενο, υπολείπονται σημαντικά των άλλων δυο μεθοδολογιών. Αντιθέτως, ο αλγόριθμος MDP *[Αλ-Γ]* παρουσιάζει πολύ καλά αποτελέσματα σε όλα τα βίντεο, με εξαίρεση το βίντεο #3: "PETS09-S2L1" όπου αποτυγχάνει σημαντικά να εντοπίσει και να παρακολουθήσει αποτελεσματικά τα κινούμενα αντικείμενα. Αυτό οφείλεται πιθανόν στην αδυναμία του αλγορίθμου να διατηρήσει τις τροχιές προχωρώντας σε πολλές αλλαγές ταυτοτήτων των στόχων, τομέας στον οποίο είναι ιδιαίτερα ευαίσθητη η μετρική MOTA. Αξίζει να αναφερθεί ότι τα αποτελέσματα του MDP μπορούν να βελτιωθούν ραγδαία για το συγκεκριμένο βίντεο αν εκπαιδευτεί σε μεγαλύτερο σετ δεδομένων που θα εμπεριέχει αρκετά βίντεο με παρόμοια γεωμετρία λήψης.

	[Αλ-Α]			[Αλ-Β]			[Αλ-Γ]		
	MOTA	**MOTP**	**MOTAL**	**MOTA**	**MOTP**	**MOTAL**	**MOTA**	**MOTP**	**MOTAL**
1	51,1	64,8	53,0	62,3	65,7	62,8	24,3	65,1	26,6
2	14,5	71,0	18,6	53,5	71,5	55,7	10,6	70,5	14,2
3	55,7	70,8	58,9	-28,6	72,3	-27,8	40,9	70,9	43,2
4	4,7	71,7	5,6	20,5	74,0	22,3	-20,6	72,1	-16,0
5	28,6	74,6	30,4	46,7	76,5	47,4	24,6	74,2	27,0
6	2,0	68,9	3,2	11,3	70,8	11,6	-5,8	69,3	-5,1
7	2,4	70,9	6,0	31,7	73,0	32,4	-29,7	70,9	-26,4
8	7,3	71,2	8,3	15,1	72,6	15,5	-52,3	71,5	-50,3
9	1,3	70,9	1,9	3,4	71,2	4,1	-163,8	70,8	-162,1
10	30,9	70,1	34,5	60,0	71,6	60,8	10,1	70,9	11,7
11	4,8	72,4	6,1	12,9	74,2	13,5	-17,7	72,3	-15,4
Μ.Ο.	**18,5**	**70,7**	**20,6**	**26,3**	**72,1**	**27,1**	**-16,3**	**70,8**	**-13,9**

Πίνακας 5.2 : Συγκεντρωτικά αποτελέσματα αλγορίθμων

Σε γενικές γραμμές, τα αποτελέσματα της αναπτυγμένης μεθοδολογίας *[Αλ-Α]* υπολείπονται αλλά είναι σχετικά κοντά με αυτά του αλγορίθμου MDP *[Αλ-Β]*. Τόσο ποσοτικά, όσο και ποιοτικά, τα αποτελέσματα μπορούν να θεωρηθούν ικανοποιητικά μιας και διαφαίνεται η υψηλή αποδοτικότητα του αλγορίθμου στη διατήρηση των τροχιών και στην ελαχιστοποίηση των αλλαγών ταυτότητας. Μεγάλη συνεισφορά στο τελευταίο έχει η εκμετάλλευση πολλαπλών ενδείξεων για την επίλυση του προβλήματος, όπως η μοντελοποίηση των χωρικών αλληλεπιδράσεων των στόχων.

Ένα παράδειγμα τέτοιας ορθής συμπεριφοράς του αλγορίθμου παρουσιάζεται στην Εικόνα 5.2. Λόγω των επικαλύψεων των στόχων (α' τρόπος χωρικής αλληλεπίδρασης) οι τροχιές 15 και 16 ομαδοποιούνται (β' τρόπος χωρικής αλληλεπίδρασης) και παρακολουθούνται πλέον από κοινού. Όταν πλέον δημιουργείται ξανά σαφής χωρικός διαχωρισμός μεταξύ αυτών των δύο στόχων, η ομάδα (group) διαλύεται και ο κάθε στόχος παρακολουθείται ξεχωριστά διατηρώντας την ταυτότητά του. Η δυσκολία της συγκεκριμένης περίπτωσης έγκειται στην ύπαρξη φυσικού εμποδίου έμπροσθεν των στόχων (αποκρύψεις-έλλειψη ανιχνεύσεων) και στην κίνηση του στόχου 12 προς την ομάδα των στόχων 15 και 16. Τελικά, η τροχιά του στόχου 16 επανακτάται όταν υπάρχει ανίχνευση εξαιτίας τόσο της ομαδοποίησης όσο και του ισχυρού μοντέλου εμφάνισης που χρησιμοποιείται.

Εικόνα 5.2 : Ενδεικτικά αποτελέσματα του αλγορίθμου για το βίντεο PETS09-S2L1

Όπως γίνεται φανερό από τα παραπάνω, πέραν της μοντελοποίησης των χωρικών αλληλεπιδράσεων μεταξύ των στόχων εξαιρετικά σημαντικό ρόλο παίζει η ισχυρή μοντελοποίηση της εμφάνισης των στόχων. Ένα τέτοιο παράδειγμα παρουσιάζεται στην Εικόνα 5.3. Παρά το γεγονός ότι η ομάδα των στόχων 29 και 30 έχει διαλυθεί, η ύπαρξη φυσικού εμποδίου και η έλλειψη ανιχνεύσεων δεν είναι αρκετή για να «χαθεί»

η τροχιά του στόχου 30, ο οποίος άμεσα επανασυσχετίζεται στην αμέσως επόμενη σωστή ανίχνευση βασιζόμενος σε μεγάλο βαθμό στο μοντέλο εμφάνισής του.

Εικόνα 5.3 : Ενδεικτικά αποτελέσματα του αλγορίθμου για το βίντεο PETS09-S2L1

Επίσης, το πρόβλημα της έλλειψης ανιχνεύσεων αποτελεί την κύρια πηγή απώλειας τροχιών για τον αναπτυγμένο αλγόριθμο, όπως φαίνεται στην Εικόνα 5.4. Στην συγκεκριμένη περίπτωση δεν παρέχεται κάποια ανίχνευση για τον στόχο #2 για αρκετά καρέ του βίντεο (τουλάχιστον 15 καρέ) λόγω ισχυρών αποκρύψεων από φυσικά εμπόδια. Ως αποτέλεσμα η τροχιά του στόχου #2 τερματίζεται, ενώ εξαιτίας του μοντέλου κίνησης όταν στα επόμενα καρέ είναι πλέον διαθέσιμη μια νέα ορθή ανίχνευση, ο αλγόριθμος αρχικοποιεί ένα νέο στόχο, δημιουργεί μία νέα τροχιά και προκύπτει επομένως μία λανθασμένη αλλαγή στην ταυτότητα ενός στόχου.

Εικόνα 5.4 : Ενδεικτικά αποτελέσματα του αλγορίθμου για το βίντεο PETS09-S2L1 μαζί με τις αντίστοιχες ανιχνεύσεις

Ένας ακόμη παράγοντας που επηρεάζει την επίδοση, όχι μόνο του αλγορίθμου που αναπτύχθηκε στα πλαίσια αυτής της εργασίας, αλλά και του συνόλου των αλγορίθμων που βασίζονται στη λογική της παρακολούθησης μέσω ανιχνεύσεων (tracking-by-detection) είναι η παρουσία λάθος ανιχνεύσεων. Ένα χαρακτηριστικό παράδειγμα παρουσιάζεται στην Εικόνα 1.5, όπου η πινακίδα στο πάνω μέρος της εικόνας θεωρείται άνθρωπος. Ο αλγόριθμος που αναπτύχθηκε αντιμετωπίζει το παραπάνω πρόβλημα απαιτώντας τη συνεχή παρακολούθηση ενός στόχου για τουλάχιστον 4 καρέ έτσι ώστε να επικυρωθεί η νέα τροχιά.

Εικόνα 5.5 : Παράδειγμα λανθασμένων ανιχνεύσεων που επηρεάζει την αποτελεσματικότητα του αλγορίθμου

Τέλος, στην Εικόνα 5.6 απεικονίζεται μία περίπτωση αλλαγής ταυτότητας για τον στόχο #20. Η αιτία της αποτυχίας έγκειται σε ένα συνδυασμό παραγόντων. Συγκεκριμένα, κανείς μπορεί να παρατηρήσει ότι μεταξύ των στόχων 1 και 18 υπάρχει ένας πεζός κρυμμένος από φυσικό εμπόδιο με παντελή έλλειψη ανιχνεύσεων. Στα επόμενα καρέ ο στόχος 18 απομακρύνεται και ο πεζός αυτός συνεχίζει να κινείται παράλληλα με τον στόχο 1, ενώ ταυτόχρονα ο στόχος 20 κινείται με φορά προς αυτούς. Εξαιτίας της υστέρησης του αλγορίθμου να επικυρώσει νέα τροχιά στον πεζό προκύπτει μία ενιαία ανίχνευση για τον πεζό και τον στόχο 20. Κάτι τέτοιο καταστρέφει την δημιουργία νέας τροχιάς για τον πεζό. Στη συνέχεια, προκύπτουν δύο ανιχνεύσεις, μία για τον πεζό και μία για τους στόχους 1 και 20. Τότε ο αλγόριθμος αντιστοίχισης επιλέγει να ταιριάξει το στόχο 20 στον πεζό λόγω των χωρικών χαρακτηριστικών των ανιχνεύσεων, δηλαδή των επικαλύψεων μεταξύ στόχων και ανιχνεύσεων. Για το λόγο αυτό, δημιουργείται λανθασμένη αλλαγή ταυτότητας. Σημειώνεται ότι η μοντελοποίηση της εμφάνισης των στόχων 1 και 20 δεν είναι ικανή να τους διαχωρίσει επαρκώς, καθώς με εξαίρεση το χρώμα του δέρματος η εμφάνισή τους είναι πανομοιότυπη χρωματικά.

Εικόνα 5.6 : Παράδειγμα αλλαγής ταυτότητας ενός στόχου για το βίντεο PETS09-S2L1

Σε αυτό το σημείο θα σχολιαστούν ποιοτικά τα αποτελέσματα των αλγορίθμων (α) Αλγ-Α, (β) MDP και (γ) του απλοϊκού αλγορίθμου σε ορισμένα καρέ άλλων βίντεο του σετ δεδομένων 2DMOT15. Αναφέρεται ότι, εκτός του βίντεο PETS09-S2L1, δεν δημιουργήθηκαν σημασιολογικές μάσκες λόγω της χρονοβόρας διαδικασίας που απαιτείται, συνεπώς δεν είναι δυνατή η σύγκριση του αλγορίθμου που αναπτύχθηκε με τη χρήση σημασιολογικών μασκών. Στην Εικόνα 5.7 παρουσιάζονται τα αποτελέσματα των προαναφερθέντων αλγορίθμων για το βίντεο *ADL-Rundle6*. Πρόκειται για ένα βίντεο που έχει υψηλή ανάλυση (HD), αν και το γεγονός αυτό φαίνεται ότι δεν επηρεάζει τον αλγόριθμο παρακολούθησης αλλά περισσότερο τον αλγόριθμο ανίχνευσης. Παρατηρούμε ότι η διαφορετική γεωμετρία λήψης επηρεάζει σημαντικά τα αποτελέσματα του Αλγ-Α, κάτι το οποίο είναι αναμενόμενο αναλογιζόμενοι το γεγονός ότι ο αλγόριθμος σχεδιάστηκε να λειτουργεί κατ' αρχήν σε βίντεο με υψηλή τοποθέτηση κάμερας, όπως του βίντεο PETS09-S2L1. Όσον αφορά τους άλλους δύο αλγορίθμους, παρατηρείται ότι ο αλγόριθμος MDP συμπεριφέρεται καλύτερα σε σχέση με τους άλλους δύο, όμως αξίζει να αναφερθεί ότι σημαντικό ρόλο στη συμπεριφορά όλων των αλγορίθμων επιτελεί η απουσία αξιόπιστων ανιχνεύσεων.

(α)

(β)

(γ)

Εικόνα 5.7 : Αποτελέσματα των αλγορίθμων (α) Αλγ-Α, (β) MDP και (γ) της απλοϊκής υλοποίησης για το βίντεο ADL-Rundle6

Ένα ακόμη παράδειγμα διαφορετικής γεωμετρίας λήψης παρουσιάζεται στην Εικόνα 5.8, όπου οι στόχοι κινούνται παράλληλα στο επίπεδο της εικόνας με την τελευταία να χαρακτηρίζεται επίσης από τη χαμηλή της τοποθέτηση. Όπως στη προηγούμενη περίπτωση, τα αποτελέσματα είναι παρόμοια με τον αλγόριθμο MDP να είναι ο καλύτερος.

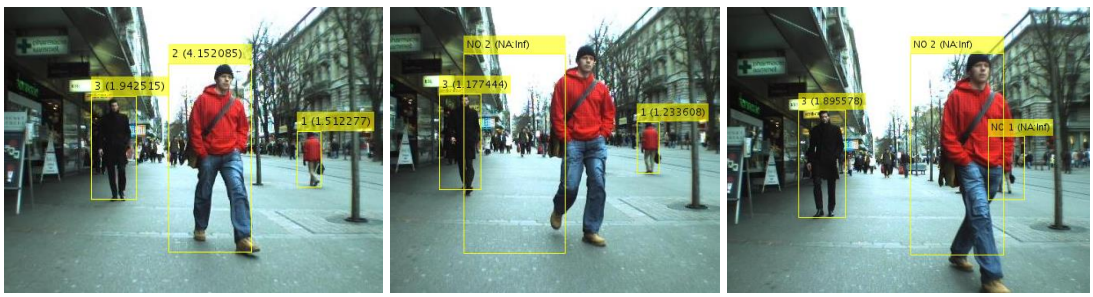
(α)

(β)

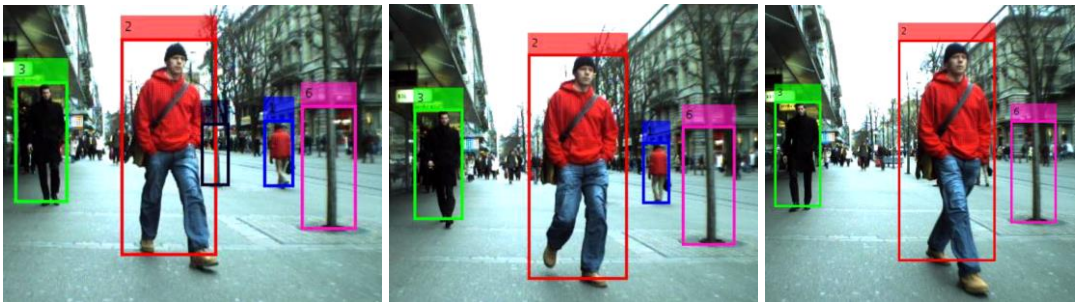
(γ)

Εικόνα 5.8 : Αποτελέσματα των αλγορίθμων (α) Αλγ-Α, (β) MDP και (γ) της απλοϊκής υλοποίησης για το βίντεο TUD-Campus

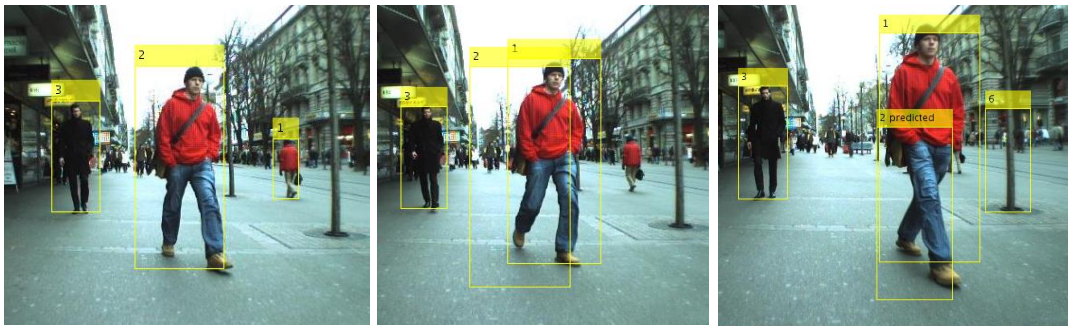
Τέλος, στην Εικόνα 5.9 παρουσιάζεται μια περίπτωση βίντεο με κινούμενη κάμερα. Τα αποτελέσματα του Αλγ-Α που αναπτύχθηκε σε αυτή την εργασία δεν είναι ιδιαίτερα ικανοποιητικά, γεγονός που πιθανότατα να οφείλεται στο γεγονός ότι δεν έχει ληφθεί υπόψη η σχετική κίνηση της κάμερας με τη σκηνή λόγω του σχετικά απλοϊκού μοντέλου κίνησης που χρησιμοποιείται. Επίσης, το πρόβλημα αυτό συναντάται και στους άλλους δύο αλγορίθμους με αποτέλεσμα κανένας να μην εμφανίζει ιδιαίτερα καλά αποτελέσματα, κάτι που επιβεβαιώνεται από τα ποσοτικά αποτελέσματα που προαναφέρθηκαν. Παρ' όλα αυτά, ο αλγόριθμος MDP δείχνει σχετικά καλύτερος σε αυτή τη περίπτωση βίντεο, αν και δεν λείπουν λανθασμένες αναθέσεις.



(α)



(β)



(γ)

Εικόνα 5.9 : Αποτελέσματα των αλγορίθμων (α) Αλγ-Α, (β) MDP και (γ) της απλοϊκής υλοποίησης για το βίντεο ETH-Bahnhof

ΚΕΦΑΛΑΙΟ 6 : ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΠΡΟΕΚΤΑΣΕΙΣ

6.1. Συμπεράσματα

Το πρόβλημα της ανίχνευσης και παρακολούθησης πολλαπλών αντικειμένων από ακολουθίες εικόνων/ βίντεο αποτελεί ένα ανοιχτό ερευνητικά αντικείμενο για την διεθνή επιστημονική κοινότητα. Τα τελευταία χρόνια παρατηρείται έντονο ερευνητικό ενδιαφέρον και αύξηση των διαθέσιμων σετ δεδομένων και αντίστοιχων διαγωνισμών για την συστηματική αξιολόγηση νέων μεθοδολογιών. Σε αυτή τη μεταπτυχιακή εργασία αναπτύχθηκε ένας αλγόριθμος παρακολούθησης πολλαπλών αντικειμένων που εκμεταλλεύεται και συνδυάζει μια σειρά από παραμέτρους για το σκοπό αυτό. Ειδικότερα, χρησιμοποιούνται ενδείξεις όπως το μοντέλο κίνησης των αντικειμένων, η περιγραφή της εμφάνισής τους μέσω μοντελοποίησης του χρώματος, καθώς και οι χωρικές αλληλεπιδράσεις των στόχων είτε υπό τη μορφή επικαλύψεων είτε υπό τη μορφή ομάδων (groups) που σχηματίζονται από αυτούς.

Τα ποσοτικά και ποιοτικά αποτελέσματα της αξιολόγησης καταδεικνύουν μια ικανοποιητική συμπεριφορά και επίδοση, σε σχέση πάντα με τις αντίστοιχες επιδόσεις αλγορίθμων στην αιχμή της τεχνολογίας (state-of-the-art). Παρόλα αυτά, οι επιδόσεις αυτές δεν επαρκούν για την ανάπτυξη ενός ολοκληρωμένου συστήματος παρακολούθησης, αυτόνομου, γενικευμένου και επιχειρησιακού. Το γεγονός αυτό οφείλεται στις δυσκολίες του προβλήματος της παρακολούθησης πολλαπλών αντικειμένων, καθώς και στην ποικιλία τόσο των βίντεο (διαφορετικές γωνίες λήψης, συνθήκες φωτισμού, κινούμενη/σταθερή κάμερα κλπ.) όσο και των μοτίβων κίνησης και αλληλεπίδρασης των παρακολουθούμενων στόχων. Αντιθέτως, στη περίπτωση μιας συγκεκριμένης εφαρμογής στην οποία το σύνολο των βίντεο έχουν μια συγκεκριμένη δομή (π.χ. παρακολούθηση αυτοκινήτων σε ένα συγκεκριμένο αυτοκινητόδρομο) το πρόβλημα της παρακολούθησης πιθανώς να επιλύεται σε μεγαλύτερο βαθμό από τους υπάρχοντες αλγορίθμους με τις κατάλληλες τροποποιήσεις.

Πιο αναλυτικά, όσον αφορά τα αποτελέσματα του αλγορίθμου που αναπτύχθηκε σε αυτή την εργασία, αρχικά αποτελούν μία σαφή και μεγάλη βελτίωση έναντι του απλοϊκού αλγορίθμου. Ωστόσο, παρατηρούνται έντονες διακυμάνσεις στις μετρικές αξιολόγησης μεταξύ των βίντεο με σταθερή και κινούμενη κάμερα. Οι διακυμάνσεις αυτές δεν αποτελούν χαρακτηριστικό μόνο της αναπτυγμένης, στο πλαίσιο της παρούσας εργασίας, μεθοδολογίας αλλά και του συνόλου των αλγορίθμων αιχμής. Το γεγονός αυτό οφείλεται κατά κύριο λόγο στην σχετικά απλή διατύπωση του μοντέλου κίνησης που χρησιμοποιείται, το οποίο συνήθως δεν λαμβάνει υπόψη τη σχετική κίνηση κάμερας-σκηνής με αποτέλεσμα να προκύπτουν λανθασμένες προβλέψεις των μελλοντικών εμφανίσεων του στόχου σε επόμενα καρέ. Κάτι τέτοιο επηρεάζει άμεσα την εκάστοτε μετρική κόστους που χρησιμοποιείται για την αντιστοίχιση στόχων και ανιχνεύσεων.

Μία από τις σημαντικές διαφορές του αλγορίθμου που αναπτύχθηκε *[Αλ-Α]* και της πιο απλοϊκής *[Αλ-Γ]* μεθοδολογίας αποτελεί η ενσωμάτωση ενδείξεων (cues) χωρικής αλληλεπίδρασης μεταξύ των στόχων που παρακολουθούνται. Πράγματι, η αξιοποίηση αυτής της πληροφορίας μέσω υπολογισμού των επικαλύψεων μεταξύ των στόχων και της ομαδοποίησης αυτών βελτιώνει θεαματικά τα αποτελέσματα των μετρικών αξιολόγησης και οδηγεί σε ποιοτικά ανώτερα αποτελέσματα. Η παραπάνω πληροφορία πέραν του άμεσου αντίκτυπού της στα αποτελέσματα υποβοηθάει έμμεσα και λοιπά υποσυστήματα του αλγορίθμου. Ειδικότερα, το μοντέλο εμφάνισης (appearance model) των στόχων μεταβάλλει το ρυθμό εκμάθησής του (learning rate) βάσει της υπολογισθείσας επικάλυψης μεταξύ των στόχων, ενώ το μοντέλο κίνησης (motion model) προσαρμόζει τις προβλέψεις του βάσει της χωρικής ομαδοποίησης των στόχων.

Επιπλέον, αξίζει να σημειωθεί ότι η πληροφορία αυτή αξιοποιείται και στο σύστημα γένεσης/θανάτου (birth/death) τροχιών μεταβάλλοντας τις παραμέτρους του ανάλογα με το αν ο στόχος ανήκει σε κάποιο γκρουπ και αποτρέποντας τη γένεση νέων τροχιών που επικαλύπτονται με υπάρχοντες στόχους. Η χρήση ενδείξεων χωρικής αλληλεπίδρασης των στόχων παρατηρείται έντονα τα τελευταία χρόνια στη διεθνή βιβλιογραφία μέσω επικαλύψεων, μέσω πλεγμάτων πληρότητας (occupancy grids) και άλλων μεθόδων. Συμπερασματικά, οι αλγόριθμοι που εφαρμόζουν σχετικά πιο πολύπλοκα μοντέλα αλληλεπιδράσεων των στόχων τείνουν να λαμβάνουν υψηλότερες θέσεις στους διαδικτυακούς διαγωνισμούς.

Οι διαδικασίες του εντοπισμού του αντικειμένου και της αντιστοίχισης των στόχων σε διαδοχικά καρέ μπορεί να πραγματοποιούνται είτε από κοινού είτε να είναι πλήρως ανεξάρτητες. Η μοντέρνα θεώρηση συνηθίζει να είναι η πρώτη, δηλαδή η μέθοδος παρακολούθησης καλείται να αντιστοιχίσει ανιχνεύσεις του τρέχοντος καρέ με τις υπάρχουσες τροχιές του προηγούμενου καρέ. Αυτή η διαδικασία δίνει σημαντικό βάρος στον αλγόριθμο εντοπισμού και στη διεθνή βιβλιογραφία [3] είναι γνωστή ως παρακολούθηση μέσω εντοπισμού (tracking-by-detection).

Στη περίπτωση της από κοινού λειτουργίας των δύο συστημάτων, οι ανιχνεύσεις υποβοηθούνται από τις προβλέψεις του αλγορίθμου παρακολούθησης. Αυτή η λογική καθιστά φανερή την αδυναμία των σύγχρονων αλγορίθμων ανίχνευσης να υποστηρίξει πλήρως ένα ολοκληρωμένο σύστημα παρακολούθησης πολλαπλών στόχων. Σε αυτή την εργασία ο αλγόριθμος που αναπτύχθηκε ακολουθεί καθαρά μία λογική παρακολούθησης μέσω εντοπισμού, κάτι το οποίο περιορίζει άμεσα τη μέγιστη δυνατή ακρίβεια που μπορεί να επιτύχει. Συγκεκριμένα, ο ανιχνευτής ACF που χρησιμοποιήθηκε σε πολλές περιπτώσεις αδυνατεί να παράγει τις απαιτούμενες ανιχνεύσεις όλων των υπό παρακολούθηση στόχων. Κάτι τέτοιο δεν συναντάται μόνο σε αυτό τον αλγόριθμο, αλλά και σε άλλους αλγορίθμους ανίχνευσης όπως για παράδειγμα διάφορες προσεγγίσεις που κάνουν χρήση συνελικτικών νευρωνικών δικτύων [29]. Επίσης, τα σφάλματα των παραπάνω αλγορίθμων (λάθη ή/και πολλαπλές ανιχνεύσεις του ίδιου αντικειμένου) μεταφέρονται στο σύστημα παρακολούθησης μειώνοντας την τελική του ακρίβεια. Παρατηρείται ότι βελτιωμένα αποτελέσματα προκύπτουν από τη συνένωση

των πολλαπλών ανιχνεύσεων του ίδιου αντικειμένου αφού παράγονται «καθαρότερες» ανιχνεύσεις. Το πρόβλημα των μη ανιχνεύσεων, ωστόσο, παραμένει ένα σημαντικό πρόβλημα τόσο για τον αλγόριθμο που αναπτύχθηκε όσο και για τους state-of-the-art αλγορίθμους που ακολουθούν αυτή τη λογική.

Ένας σημαντικός παράγοντας στη βελτίωση των αποτελεσμάτων αποτέλεσε η σημασιολογική κατάτμηση (semantic segmentation) του κάθε καρέ του βίντεο. Παρά το γεγονός ότι στα πλαίσια αυτής της εργασίας η κατάτμηση πραγματοποιήθηκε χειροκίνητα και αποτέλεσε μια εξαιρετικά χρονοβόρα διαδικασία, παρατηρήθηκε για το βίντεο στο οποίο εφαρμόστηκε μια βελτίωση της τάξεως του 10%. Αυτό οφείλεται στον ορθότερο υπολογισμό του μοντέλου εμφάνισης των στόχων αφού φιλτράρεται όλη η πληροφορία του παρασκηνίου και αξιοποιείται ατόφια η φασματική πληροφορία του στόχου. Αντίστοιχες, προσεγγίσεις αιχμής στη σημασιολογική κατάτμηση [26, 27] απαιτούν τεράστιο όγκο δεδομένων εκπαίδευσης τα οποία δεν ήταν διαθέσιμα. Συνεπώς, δεν διερευνήθηκε η επίδοση τέτοιων αλγορίθμων και η εργασία βασίστηκε σε δεδομένα από χειροκίνητη κατάτμηση.

Συμπερασματικά, ο αλγόριθμος που αναπτύχθηκε εκπληρώνει στο σύνολό του τον στόχο της παρούσας εργασίας, δηλαδή τη διερεύνηση του προβλήματος της παρακολούθησης πολλαπλών στόχων και της ενσωμάτωσης διάφορων ενδείξεων για την επίλυση αυτού. Η μεγαλύτερη αδυναμία της μεθοδολογίας που αναπτύχθηκε αποτελεί ο χειροκίνητος καθορισμός των παραμέτρων του αλγορίθμου σε αντίθεση με τη συντριπτική πλειοψηφία των state-of-the-art αλγορίθμων που χρησιμοποιούν συστήματα μάθησης γι’ αυτό το σκοπό.

6.2. Μελλοντικές κατευθύνσεις

Το πρόβλημα της παρακολούθησης πολλαπλών αντικειμένων αποτελεί ένα από τα πιο ενεργά πεδία έρευνας στην επιστήμη της Όρασης Υπολογιστών. Πρόκειται για ένα ανοιχτό ζήτημα το οποίο απαιτεί την επίλυση αρκετών επιμέρους προβλημάτων όπως η αναγνώριση αντικειμένων, η αντιστοίχιση δεδομένων, η μαθηματική προτυποποίηση μοντέλων εμφάνισης και κίνησης. Τα τελευταία χρόνια παρατηρείται μία στροφή της κοινότητας στη χρήση τεχνικών Μηχανικής Μάθησης (Machine Learning) και Βαθιάς Μάθησης (Deep Learning) που αξιοποιούν δομές όπως τα τεχνητά νευρωνικά δίκτυα. Στην κατεύθυνση αυτή έχει συμβάλλει αποφασιστικά τόσο η αποδοτικότητα αυτών των τεχνικών σε όμορες εφαρμογές όπως η ταξινόμηση εικόνων και η αναγνώριση αντικειμένων, όσο και το ολοένα και αυξανόμενο πλήθος δεδομένων αληθείας για το συγκεκριμένο ζήτημα. Στον παρακάτω πίνακα παρουσιάζονται οι επιδόσεις των state-of-the-art αλγορίθμων παρακολούθησης πολλαπλών αντικειμένων στο MOTC για το σετ δεδομένων 2DMOT16.

Το σύνολο των παραπάνω αλγορίθμων έχουν πέτυχει τις καλύτερες επιδόσεις διότι αξιοποιούν πολύπλοκες αρχιτεκτονικές νευρωνικών δικτύων και κυρίως χαρα-

κτηρίζονται από την σύνθετη μοντελοποίηση των επιμέρους υποσυστημάτων ενός ολοκληρωμένου συστήματος παρακολούθησης που λειτουργεί με τη λογική της παρακολούθησης μέσω εντοπισμού (tracking-by-detection). Συγκεκριμένα, οι αρχιτεκτονικές που είναι ιδιαίτερα δημοφιλείς στον κλάδο αφορούν τα Συνελικτικά Νευρωνικά Δίκτυα (Convolutional NN) για την εξαγωγή χαρακτηριστικών εμφάνισης, καθώς και τα Αναδρομικά Νευρωνικά Δίκτυα (Recurrent NN) και τα Νευρωνικά Δίκτυα Μακρο-Βραχυπρόθεσμης Μνήμης (Long-Short Term Memory) για την αντιστοίχιση δεδομένων σε μία ακολουθία εικόνων. Τα τελευταία έχουν την ιδιότητα αποθήκευσης της πληροφορίας εν είδη μνήμης για πολλά καρέ ενός βίντεο επιτρέποντας με αυτό τον τρόπο τη δημιουργία ενός χρονικά εύρωστου συνόλου χαρακτηριστικών για το κάθε αντικείμενο που παρακολουθείται.

Αλγόριθμος	MOTA	MOTP	FAF	MT	ML	FP	FN	ID Switches	Frag
NOMT [10]	46,4 ± 9,9	76,6	1,6	18,3	41,4	9753	87565	359	504
JMC [11]	46,3 ± 9,0	75,7	1,1	15,5	39,7	6373	90914	657	1114
MDPNN16 [12]	43,8 ± 7,3	75,5	0,6	12,4	40,7	3501	98193	723	2036
oICF [13]	43,2 ± 10,2	74,3	1,1	11,3	48,5	6651	96515	381	1404
MHT_DAM [14]	42,9 ± 8,9	76,6	1	13,6	46,9	5668	97919	499	659
LINF1 [15]	41,0 ± 9,5	74,8	1,3	11,6	51,3	7896	99224	430	963
EAMTT_pub [16]	38,8 ± 8,5	75,1	1,4	7,9	49,1	8114	102452	965	1657
OVBT [17]	38,4 ± 8,8	75,4	1,9	7,5	47,3	11517	99463	1321	2140
LTTSC-CRF [18]	37,6 ± 9,9	75,9	2	9,6	55,2	11969	101343	481	1012
LP2D [19]	35,7 ± 10,1	75,8	0,9	8,7	50,7	5084	111163	915	1264
TBD [20]	33,7 ± 9,2	76,5	1	7,2	54,2	5804	112587	2418	2252
CEM [21]	33,2 ± 7,9	75,8	1,2	7,8	54,4	6837	114322	642	731
DP_NMS [22]	32,2 ± 9,8	76,4	0,2	5,4	62,1	1123	121579	972	944
GMPHD_HDA [23]	30,5 ± 6,9	75,4	0,9	4,6	59,7	5169	120970	539	731
SMOT [24]	29,7 ± 7,3	75,2	2,9	5,3	47,7	17426	107552	3108	4483
JPDA_m [25]	26,2 ± 6,1	76,3	0,6	4,1	67,5	3689	130549	365	638

Πίνακας 6.1 : Επιδόσεις των καλύτερων αλγορίθμων στο MOTC για το σετ δεδομένων MOT16 [46]

Στη βιβλιογραφία εντοπίζεται, επίσης, η τάση ενσωμάτωσης παράπλευρων διαδικασιών που απαιτούνται για τη διαδικασία παρακολούθησης, όπως η ανίχνευση αντικειμένων, σε ένα ολοκληρωμένο σύστημα. Για παράδειγμα, αρκετοί από τους παραπάνω αλγορίθμους χρησιμοποιούν μία ενιαία αρχιτεκτονική ενός νευρωνικού δικτύου ώστε η ανίχνευση αντικειμένων και η παρακολούθηση των στόχων να επιτυγχάνεται σε ένα χρονικό βήμα με τα δύο συστήματα να υποβοηθούν το ένα το άλλο. Αντίστοιχες προσεγγίσεις περιλαμβάνουν τη παράλληλη σημασιολογική κατάτμηση ώστε να επιτύχουν ακόμη καλύτερα αποτελέσματα.

Ωστόσο, παρά την προσπάθεια της ερευνητικής κοινότητας να λύσει το πρόβλημα της παρακολούθησης πολλαπλών αντικειμένων, τα καλύτερα αποτελέσματα σε καμία περίπτωση δεν μπορούν να χαρακτηριστούν ως αποδεκτά για ένα τέτοιο ζήτημα. Αυτό το γεγονός γεννά ερωτήματα για τις μελλοντικές κατευθύνσεις της έρευνας και για το κατά πόσο η έκφραση του προβλήματος με τη λογική της παρακολούθησης μέσω

εντοπισμού επαρκεί για την επίλυσή του. Κατά τη διάρκεια εκπόνησης της παρούσας εργασίας παρατηρήθηκε η ανάγκη εξαγωγής ποιοτικής πληροφορίας υψηλού επιπέδου. Χαρακτηριστικό παράδειγμα αποτελεί το μοντέλο εμφάνισης που χρησιμοποιεί ο αλγόριθμος παρακολούθησης που αναπτύχθηκε. Το προϊόν αυτού του μοντέλου δεν είναι παρά ένα διάνυσμα περιγραφής στο οποίο περικλείεται το σύνολο της σημασιολογικής πληροφορίας για την εμφάνιση ενός ανθρώπου. Κάτι τέτοιο είναι μόνο έμμεσα συνυφασμένο με ποιοτικά χαρακτηριστικά εμφάνισης, όπως το χρώμα της μπλούζας, το χρώμα του δέρματος, των μαλλιών κλπ. Στη περίπτωση που τέτοιου είδους ποιοτικά χαρακτηριστικά ήταν στη διάθεση του αλγορίθμου, τότε το πρόβλημα αντιστοίχισης θα επιλύονταν σε σημασιολογικό επίπεδο και όχι σε επίπεδο μέτρησης μιας απόστασης σε έναν πεπερασμένης διάστασης χώρο. Ακόμη, η σημασιολογική πληροφορία για την κίνηση ενός ανθρώπου, δηλαδή ο χαρακτηρισμός του τύπου κίνησης του πεζού (βάδην, τρέξιμο κλπ.), σαφώς θα ενίσχυε τον προσδιορισμό των στόχων στις ορθές ανιχνεύσεις σε αντίθεση με την απλή σύγκριση διανυσμάτων ταχύτητας. Η κοινότητα τα τελευταία χρόνια δείχνει την τάση προς εξαγωγή πληροφορίας υψηλότερου σημασιολογικού επιπέδου, γεγονός το οποίο θα μπορούσε σημαντικά τα αποτελέσματα ενός συστήματος παρακολούθησης.

Εικόνα 6.1 : Παραδείγματα εικόνων με σημασιολογική κατάτμηση [28]

Εν κατακλείδι, το πρόβλημα της παρακολούθησης πολλαπλών αντικειμένων αποτελεί ένα ανοιχτό ζήτημα για τη διεθνή κοινότητα. Οι state-of-the-art αλγόριθμοι

ακόμη δεν βρίσκονται σε επίπεδο επαρκούς λύσης του, συνεπώς το μέλλον προβλέπεται δραστήριο σε αυτό το επιστημονικό πεδίο. Η επικείμενη στροφή της κοινότητας προς την αξιοποίηση ολοένα και λεπτομερέστερης σημασιολογικής αναπαράστασης μοιάζει ως αναπόφευκτη. Παρόλα αυτά, η συνεχόμενη ανάπτυξη τέτοιου είδους αλγορίθμων προμηνύει σημαντικές εξελίξεις στον κλάδο, καθώς και στο σύνολο της επιστήμης της Όρασης Υπολογιστών.

ΠΑΡΑΡΤΗΜΑ

Hungarian Algorithm

Ο Αλγόριθμος του Munkres ή αλλιώς Hungarian Algorithm [31] είναι ένας συνδυαστικός αλγόριθμος βελτιστοποίησης που λύνει το πρόβλημα της αντιστοίχισης των δεδομένων σε πολυωνυμικό χρόνο. Η μεθοδολογία αναπτύχθηκε και δημοσιεύθηκε από τον Harold Kuhn, ο οποίος την ονομάτισε "Hungarian Method" διότι ο αλγόριθμος βασίστηκε σε μεγάλο βαθμό από τη προγενέστερη έρευνα δύο Ούγγρων μαθηματικών, τους Dénes Kőnig και Jenő Egerváry. Ο James Munkres ασχολήθηκε περαιτέρω με αυτό τον αλγόριθμο το 1957 και παρατήρησε ότι είναι ισχυρώς πολυωνυμικός. Από εκείνη την εποχή έγινε γνωστός, λοιπόν, ως ο αλγόριθμος του Munkres με τη χρονική πολυπλοκότητα του οποίου να είναι της τάξεως $O(n^3)$.

Το **πρόβλημα της αντιστοίχισης** που λύνει αυτός ο αλγόριθμος συνοψίζεται ως εξής : Δεδομένων m στοιχείων θέλουμε να αντιστοιχίσουμε "1-προς-1" n στοιχεία με τον βέλτιστο δυνατό τρόπο. Στη περίπτωση που m > n ή το αντίθετο δεν θα πραγματοποιηθούν κάποιες αντιστοιχίσεις.

Προκειμένου να αναλυθούν τα βήματα του αλγορίθμου, ας ορίσουμε πρώτα το μαθηματικό μοντέλο. Έστω c_{ij} τα κόστη αντιστοίχισης της i-οστής τιμής στην j-οστή τιμή. Ο **πίνακας κόστους** ή αλλιώς ο πίνακας αντιστοίχισης ορίζεται να είναι ο πίνακας μεγέθους (m x n) ως εξής :

$$\begin{pmatrix} c_{1,1} & \cdots & c_{1,n} \\ \vdots & \ddots & \vdots \\ c_{m,1} & \cdots & c_{m,n} \end{pmatrix}$$

Ως **αντιστοίχιση** θεωρείται το σύνολο των min(m,n) στοιχείων του πίνακα όπου κανένα από τα οποία δεν βρίσκονται στην ίδια σειρά ή στήλη. Το σύνολο των τιμών των n στοιχείων του πίνακα αποτελεί το κόστος αντιστοίχισης, ενώ η αντιστοίχιση με το μικρότερο δυνατό κόστος καλείται **βέλτιστη αντιστοίχιση**.

Ο αλγόριθμος του Munkres, λοιπόν, υπολογίζει τη βέλτιστη δυνατή αντιστοίχιση ακολουθώντας τα παρακάτω βήματα :

1. Αφαίρεση της μικρότερης τιμής κάθε σειράς από όλα τα στοιχεία της αντίστοιχης σειράς.
2. Αφαίρεση της μικρότερης τιμής κάθε στήλης από όλα τα στοιχεία της αντίστοιχης στήλης.
3. Εύρεση του συνδυασμού γραμμών και στηλών έτσι ώστε όλα τα μηδενικά στοιχεία του πίνακα κόστους να εμπεριέχονται σε αυτές με τέτοιο τρόπο ώστε να χρησιμοποιείται ο ελάχιστος αριθμός γραμμών και στηλών.

4. *Τεστ βέλτιστης λύσης* : (α) Αν ο ελάχιστος αριθμός γραμμών και στηλών είναι min(m,n) τότε η διαδικασία τερματίζεται. (β) Αν ο ελάχιστος αριθμός γραμμών και στηλών είναι μικρότερος από min(m,n) τότε ακολουθεί το Βήμα 5.
5. Καθορισμός του μικρότερου στοιχείου του πίνακα που δεν ανήκει στο συνδυασμό γραμμών και στηλών που επιλέχθηκε τελικά. Αφαίρεση του στοιχείου από κάθε μη επιλεχθείσα σειρά και πρόσθεσή του σε κάθε επιλεχθείσα στήλη. Επιστροφή στο Βήμα 3.

Ακολουθεί ένα παράδειγμα εφαρμογής του αλγορίθμου του Munkres :

Βήμα 1

$$\begin{bmatrix} 90 & 75 & 75 & 80 \\ 35 & 85 & 55 & 65 \\ 125 & 95 & 90 & 105 \\ 45 & 110 & 95 & 115 \end{bmatrix} \sim \begin{bmatrix} 15 & 0 & 0 & 5 \\ 0 & 50 & 20 & 30 \\ 35 & 5 & 0 & 15 \\ 0 & 65 & 50 & 70 \end{bmatrix}$$

Βήμα 2

$$\begin{bmatrix} 15 & 0 & 0 & 5 \\ 0 & 50 & 20 & 30 \\ 35 & 5 & 0 & 15 \\ 0 & 65 & 50 & 70 \end{bmatrix} \sim \begin{bmatrix} 15 & 0 & 0 & 0 \\ 0 & 50 & 20 & 25 \\ 35 & 5 & 0 & 10 \\ 0 & 65 & 50 & 65 \end{bmatrix}$$

Βήμα 3

$$\begin{bmatrix} \cancel{15} & \cancel{0} & \cancel{0} & \cancel{0} \\ \cancel{0} & 50 & \cancel{20} & 25 \\ \cancel{35} & 5 & \cancel{0} & 10 \\ \cancel{0} & 65 & \cancel{50} & 65 \end{bmatrix}$$

Βήμα 4 : Επειδή ο ελάχιστος αριθμός γραμμών είναι μικρότερος από 4, συνεχίζουμε στο επόμενο βήμα.

Βήμα 5

$$\begin{bmatrix} 15 & 0 & 0 & 0 \\ 0 & 50 & 20 & 25 \\ 35 & 5 & 0 & 10 \\ 0 & 65 & 50 & 65 \end{bmatrix} \sim \begin{bmatrix} 15 & 0 & 0 & 0 \\ -5 & 45 & 15 & 20 \\ 30 & 0 & -5 & 5 \\ -5 & 60 & 45 & 60 \end{bmatrix}$$

$$\begin{bmatrix} 15 & 0 & 0 & 0 \\ -5 & 45 & 15 & 20 \\ 30 & 0 & -5 & 5 \\ -5 & 60 & 45 & 60 \end{bmatrix} \sim \begin{bmatrix} 20 & 0 & 5 & 0 \\ 0 & 45 & 20 & 20 \\ 35 & 0 & 0 & 5 \\ 0 & 60 & 50 & 60 \end{bmatrix}$$

Επιστροφή στο Βήμα 3

$$\begin{bmatrix} \cancel{40} & \cancel{0} & \cancel{5} & \cancel{0} \\ \cancel{0} & \cancel{25} & \cancel{0} & \cancel{0} \\ \cancel{55} & \cancel{0} & \cancel{0} & \cancel{5} \\ \cancel{0} & \cancel{40} & \cancel{30} & \cancel{40} \end{bmatrix}$$

Βήμα 4 : Επειδή ο ελάχιστος αριθμός γραμμών είναι 4 τότε η βέλτιστη αντιστοίχιση είναι δυνατή και η διαδικασία τερματίζεται. Ο πίνακας αντιστοίχισης που προκύπτει, λοιπόν, είναι ο εξής :

$$\begin{bmatrix} 40 & 0 & 5 & \boxed{0} \\ 0 & 25 & \boxed{0} & 0 \\ 55 & \boxed{0} & 0 & 5 \\ \boxed{0} & 40 & 30 & 40 \end{bmatrix} \begin{bmatrix} 90 & 75 & 75 & \boxed{80} \\ 35 & 85 & \boxed{55} & 65 \\ 125 & \boxed{95} & 90 & 105 \\ \boxed{45} & 110 & 95 & 115 \end{bmatrix}$$

ΒΙΒΛΙΟΓΡΑΦΙΑ

[1] Leal-Taixé, Laura, et al. "Motchallenge 2015: Towards a benchmark for multi-target tracking." *arXiv preprint arXiv:1504.01942* (2015).

[2] Andriluka, Mykhaylo, Stefan Roth, and Bernt Schiele. "Monocular 3d pose estimation and tracking by detection." *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 2010.

[3] Andriluka, Mykhaylo, Stefan Roth, and Bernt Schiele. "People-tracking-by-detection and people-detection-by-tracking." *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. IEEE, 2008.

[4] Ellis, Anna, and James Ferryman. "PETS2010 and PETS2009 evaluation of results using individual ground truthed single views." *Advanced Video and Signal Based Surveillance (AVSS), 2010 Seventh IEEE International Conference on*. IEEE, 2010.

[5] Ess, Andreas, et al. "A mobile vision system for robust multi-person tracking." *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. IEEE, 2008.

[6] Geiger, Andreas, Philip Lenz, and Raquel Urtasun. "Are we ready for autonomous driving? the kitti vision benchmark suite." *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012.

[7] Dollár, Piotr, et al. "Fast feature pyramids for object detection." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 36.8 (2014): 1532-1545.

[8] Xiang, Yu, Alexandre Alahi, and Silvio Savarese. "Learning to track: Online multi-object tracking by decision making." *Proceedings of the IEEE International Conference on Computer Vision*. 2015.

[9] Kasturi, Rangachar, et al. "Framework for performance evaluation of face, text, and vehicle detection and tracking in video: Data, metrics, and protocol." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31.2 (2009): 319-336.

[10] Choi, Wongun. "Near-online multi-target tracking with aggregated local flow descriptor." *Proceedings of the IEEE International Conference on Computer Vision*. 2015.

[11] Tang, Siyu, et al. "Multi-person tracking by multicut and deep matching." *European Conference on Computer Vision*. Springer International Publishing, 2016.

[12] Sadeghian, Amir, Alexandre Alahi, and Silvio Savarese. "Tracking the untrackable: Learning to track multiple cues with long-term dependencies." *arXiv preprint arXiv:1701.01909* (2017).

[13] Kieritz, Hilke, et al. "Online multi-person tracking using Integral Channel Features." *Advanced Video and Signal Based Surveillance (AVSS), 2016 13th IEEE International Conference on*. IEEE, 2016.

[14] Kim, Chanho, et al. "Multiple hypothesis tracking revisited." *Proceedings of the IEEE International Conference on Computer Vision*. 2015.

[15] Fagot-Bouquet, Loïc, et al. "Improving multi-frame data association with sparse representations for robust near-online multi-object tracking." *European Conference on Computer Vision*. Springer International Publishing, 2016.

[16] Sanchez-Matilla, Ricardo, Fabio Poiesi, and Andrea Cavallaro. "Online multi-target tracking with strong and weak detections." *European Conference on Computer Vision*. Springer International Publishing, 2016.

[17] Ban, Yutong, et al. "Tracking multiple persons based on a variational bayesian model." *European Conference on Computer Vision*. Springer International Publishing, 2016.

[18] Le, Nam, Alexander Heili, and Jean-Marc Odobez. "Long-term time-sensitive costs for crf-based tracking by detection." *European Conference on Computer Vision*. Springer International Publishing, 2016.

[19] Leal-Taixé, Laura, Gerard Pons-Moll, and Bodo Rosenhahn. "Everybody needs somebody: Modeling social and grouping behavior on a linear programming multiple people tracker." *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*. IEEE, 2011.

[20] Geiger, Andreas, et al. "3d traffic scene understanding from movable platforms." *IEEE transactions on pattern analysis and machine intelligence* 36.5 (2014): 1012-1025.

[21] Milan, Anton, Stefan Roth, and Konrad Schindler. "Continuous energy minimization for multitarget tracking." *IEEE transactions on pattern analysis and machine intelligence* 36.1 (2014): 58-72.

[22] Pirsiavash, Hamed, Deva Ramanan, and Charless C. Fowlkes. "Globally-optimal greedy algorithms for tracking a variable number of objects." *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011.

[23] Song, Young-min, and Moongu Jeon. "Online multiple object tracking with the hierarchically adopted gm-phd filter using motion and appearance." *Consumer Electronics-Asia (ICCE-Asia), IEEE International Conference on*. IEEE, 2016.

[24] Dicle, Caglayan, Octavia I. Camps, and Mario Sznaier. "The way they move: Tracking multiple targets with similar appearance." *Proceedings of the IEEE International Conference on Computer Vision*. 2013.

[25] Hamid Rezatofighi, Seyed, et al. "Joint probabilistic data association revisited." *Proceedings of the IEEE International Conference on Computer Vision*. 2015.

[26] Badrinarayanan, Vijay, Alex Kendall, and Roberto Cipolla. "Segnet: A deep convolutional encoder-decoder architecture for image segmentation." *arXiv preprint arXiv:1511.00561* (2015).

[27] Uijlings, Jasper RR, et al. "Selective search for object recognition." *International journal of computer vision*104.2 (2013): 154-171.

[28] Mottaghi, Roozbeh, et al. "The role of context for object detection and semantic segmentation in the wild." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2014.

[29] Ren, Shaoqing, et al. "Faster R-CNN: Towards real-time object detection with region proposal networks." *Advances in neural information processing systems*. 2015.

[30] Solera, Francesco, Simone Calderara, and Rita Cucchiara. "Towards the evaluation of reproducible robustness in tracking-by-detection." *Advanced Video and Signal Based Surveillance (AVSS), 2015 12th IEEE International Conference on*. IEEE, 2015.

[31] Munkres, James. "Algorithms for the assignment and transportation problems." *Journal of the society for industrial and applied mathematics* 5.1 (1957): 32-38.

[32] Achanta, Radhakrishna, et al. "SLIC superpixels compared to state-of-the-art superpixel methods." *IEEE transactions on pattern analysis and machine intelligence* 34.11 (2012): 2274-2282.

[33] Yilmaz, Alper, Omar Javed, and Mubarak Shah. "Object tracking: A survey." *Acm computing surveys (CSUR)* 38.4 (2006): 13.

[34] Li, Yuan, Chang Huang, and Ram Nevatia. "Learning to associate: Hybrid-boosted multi-target tracker for crowded scene." *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 2009.

[35] Reid, Donald. "An algorithm for tracking multiple targets." *IEEE transactions on Automatic Control*24.6 (1979): 843-854.

[36] Bar-Shalom, Y., T. Fortmann, and M. Scheffe. "Joint probabilistic data association for multiple targets in clutter." *Proc. Conf. on Information Sciences and Systems*. 1980.

[37] Kalal, Zdenek, Krystian Mikolajczyk, and Jiri Matas. "Tracking-learning-detection." *IEEE transactions on pattern analysis and machine intelligence* 34.7 (2012): 1409-1422.

[38] Pérez, Patrick, et al. "Color-based probabilistic tracking." *Computer vision—ECCV 2002* (2002): 661-675.

[39] Oron, Shaul, et al. "Locally orderless tracking." *International Journal of Computer Vision* 111.2 (2015): 213-228.

[40] Ross, David A., et al. "Incremental learning for robust visual tracking." *International Journal of Computer Vision* 77.1 (2008): 125-141.

[41] Kwon, Junseok, and Kyoung Mu Lee. "Tracking by sampling trackers." *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011.

[42] Wu, Yi, Bin Shen, and Haibin Ling. "Online robust image alignment via iterative convex optimization." *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012.

[43] Grabner, Helmut, Michael Grabner, and Horst Bischof. "Real-time tracking via on-line boosting." *Bmvc*. Vol. 1. No. 5. 2006.

[44] Wu, Yi, Jongwoo Lim, and Ming-Hsuan Yang. "Online object tracking: A benchmark." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2013.

[45] Pellegrini, Stefano, et al. "You'll never walk alone: Modeling social behavior for multi-target tracking." *Computer Vision, 2009 IEEE 12th International Conference on*. IEEE, 2009.

[46] Yu, Zhonghao, et al. "Tracking the trackers." *Proceedings of the 25th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 2016.

[47] Τσιρώνης Β. "Σχεδιασμός, Ανάπτυξη και Αξιολόγηση Μοντέλων εμφάνισης με Εφαρμογή στην Οπτική Παρακολούθηση", Μεταπτυχιακή εργασία, ΕΜΠ, 2017.