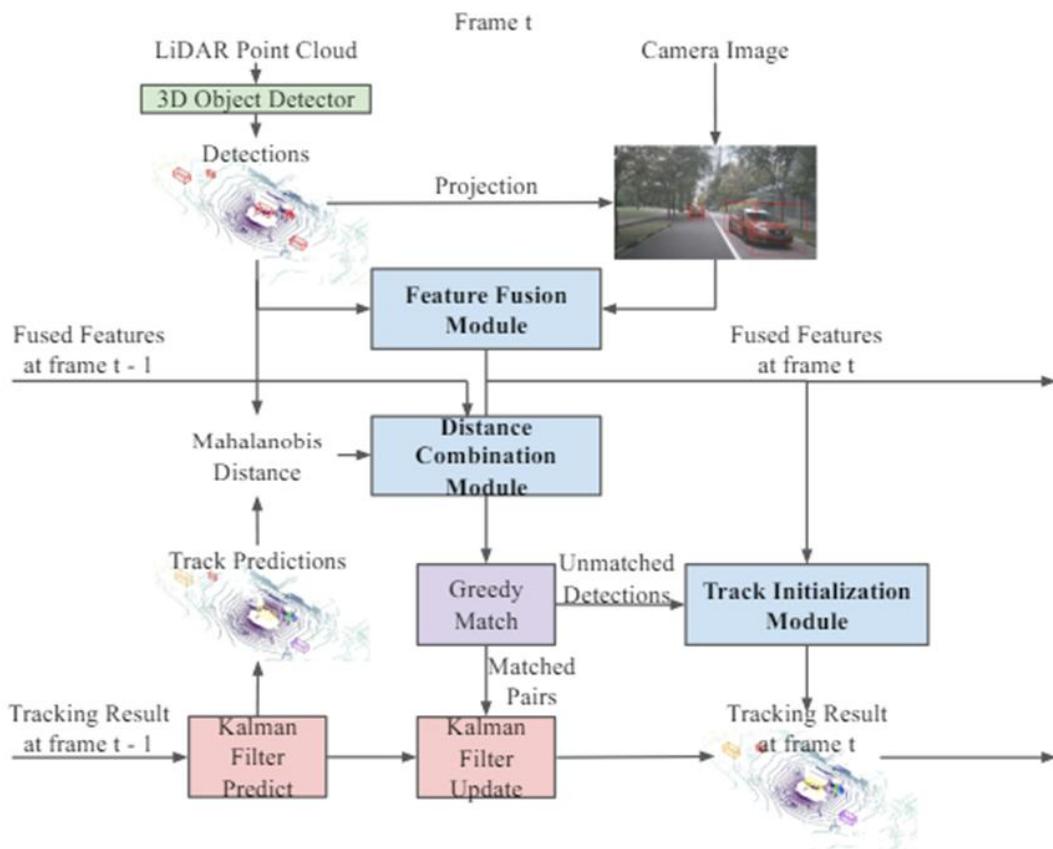


## Διπλωματική εργασία: Σπαθούλας Δημήτριος

Επιβλέπων: Γαστεράτος Αντώνιος, Καθηγητής, Κοσμήτορας Πολυτεχνικής Σχολής

# Τρισδιάστατη Πολυτροπική Παρακολούθηση Πολλαπλών Στόχων με Μηχανική Μάθηση και Αναλυτικός Υπολογισμός Κινδύνου Σύγκρουσης για την Αυτόνομη Πλούγηση Οχημάτων



Σπαθούλας Δημήτριος 92206

Ξάνθη Ιανουάριος 2025





Δημοκρίτειο Πανεπιστήμιο Θράκης

Πολυτεχνική Σχολή

Τμήμα: Μηχανικών Παραγωγής και Διοίκησης

## Διπλωματική εργασία: Σπαθούλας Δημήτριος

---

Επιβλέπων: Γαστεράτος Αντώνιος, Καθηγητής, Κοσμήτορας Πολυτεχνικής Σχολής

### Στοιχεία Διπλωματικής Εργασίας

Σχολή: Πολυτεχνική

Τμήμα: Μηχανικών Παραγωγής και Διοίκησης

Τομέας/Εργαστήριο: Εργαστήριο Ρομποτικής και Αυτοματισμών

Όνοματεπώνυμο (ελληνικά): Σπαθούλας Δημήτριος

Όνοματεπώνυμο (αγγλικά): Spathoulas Dimitrios

Όνομα Πατρός: Κωνσταντίνος

Τίτλος Εργασίας (ελληνικά): Τρισδιάστατη Πολυτροπική Παρακολούθηση Πολλαπλών Στόχων με Μηχανική Μάθηση και Αναλυτικός Υπολογισμός Κινδύνου Σύγκρουσης για την Αυτόνομη Πλοήγηση Οχημάτων

Τίτλος Εργασίας (αγγλικά): 3d Multi-modal Multi-object Tracking with Machine Learning and Analytic Collision Risk Calculation for Autonomous Vehicles Navigation

Ημερομηνία υποστήριξης διπλωματικής εργασίας: 24/1/2025

Επιβλέπων καθηγητής: Καθηγητής, Κοσμήτορας Πολυτεχνικής Σχολής, Α. Γαστεράτος

Εξεταστική Επιτροπή:

1. Αντώνιος Γαστεράτος, Καθηγητής, ΤΜΠΔ
2. Στέφανος Κατσαβούνης, Καθηγητής, ΤΜΠΔ
3. Αθανάσιος Βαβάτσικος, Καθηγητής, ΤΜΠΔ



## **Copyright ©**

Οι διπλωματικές εργασίες προστατεύονται με το νόμο των πνευματικών δικαιωμάτων (Ν.2121/93). Η διπλωματική εργασία αποτελεί προϊόν συνεργασίας του φοιτητή και των μελών ΔΕΠ, ΕΔΙΠ, ΕΕΠ και ΕΤΕΠ που επιβλέπουν την εκπόνησή της στο ΤΜΠΔ. Τα φυσικά αυτά πρόσωπα έχουν και τα πνευματικά δικαιώματα στη δημοσίευση των αποτελεσμάτων της διπλωματικής εργασίας σε επιστημονικά περιοδικά καθώς και στις ανακοινώσεις τους σε επιστημονικά συνέδρια. Στις δημοσιεύσεις αυτές τηρείται η επιστημονική δεοντολογία. Η διπλωματική εργασία και ότι άλλο έχει αναπτυχθεί στα πλαίσια της εκπόνησης ή θα προκύψει από αυτήν όπως τα πιθανά δικαιώματα ευρεσιτεχνίας ή εμπορικής εκμετάλλευσης, προστατεύονται με τη νομοθεσία Ν.2121/93 περί πνευματικής ιδιοκτησίας και ανήκουν στο φοιτητή(-τρια), τα μέλη ΔΕΠ, ΕΔΙΠ, ΕΕΠ και ΕΤΕΠ που έχουν επιβλέψει και αναφέρονται στη δήλωση ανάθεσης της διπλωματικής εργασίας. Στη λίστα των ατόμων που έχουν πνευματικά δικαιώματα μπορεί να προστεθούν και διδάσκοντες ΠΔ407/80 ή άλλα άτομα που έχουν συμβάλλει δημιουργικά στην επιτυχή ολοκλήρωσή τους, ύστερα από έγκριση της Γενικής Συνέλευσης.



Cowards die many times before their deaths;  
The valiant never taste of death but once.

Shakespeare



## **ΕΥΧΑΡΙΣΤΙΕΣ**

Θα ήθελα να ευχαριστήσω τους καθηγητές μου δρ. Αντώνιο Γαστεράτο και δρ. Στέφανο Κατσαβούνη για την εμπιστοσύνη που μου έχουν δείξει και την βοήθεια που μου έχουν προσφέρει κατά τα χρόνια.

Να ευχαριστήσω τον Κοσμά Τσιάκα για την πολύτιμη του βοήθεια στην περάτωση της παρούσας διπλωματικής εργασίας.

Τέλος, θα ήθελα να ευχαριστήσω την οικογένεια μου για την υποστήριξη και την κατανόηση της όλα αυτά τα χρόνια.

## ΠΡΟΛΟΓΟΣ

---

Οδεύοντας στον κόσμο του μέλλοντος λέξεις όπως όραση, αντίληψη μηχανών και μηχανική μάθηση έχουν καταφέρει να γίνουν αναπόσπαστο κομμάτι στο παρασκήνιο του παρόντος, όχι μόνο στον ευρύτερο τομέα της ρομποτικής, αλλά σε όλες τις επιστήμες. Έννοιες όπως αναγνώριση, αίσθηση, αντίληψη και προσοχή ποτέ άλλοτε έως σήμερα δεν είχαν τόσο άμεση και βαθιά συσχέτιση με κάτι άλλο εκτός της ανθρώπινης νόησης. Αυτή η γραμμή αν και ακόμα νωρίς, αρχίζει να ξεθωριάζει καθώς η ανθρωπότητα οδεύει στην δημιουργία και κατασκευή πραγματικής νοημοσύνης. Η παρούσα εργασία εξετάζει και χρησιμοποιεί αφηρημένα μαθηματικά μοτίβα με σκοπό την αναπαράσταση τους σε νοητές για τον άνθρωπο διαστάσεις και ως μέσο αντίληψης για έναν αυτόνομο πράκτορα. Στόχος είναι μέσω μεθόδων τεχνητής νοημοσύνης ο πράκτορας να καταφέρει να "δει" και να "αντιληφθεί" τα αντικείμενα και το περιβάλλον γύρω του, αλλά ακόμα και να μπορέσει να "αισθανθεί" τον επικείμενο κίνδυνο σύγκρουσης, με τον ίδιο τρόπο που ένας άνθρωπος ασυνείδητα κάνει. Όλα αυτά σε ένα από τα πιο πολύπλοκα και απαιτητικά στην προσοχή, περιβάλλοντα που αντιμετωπίζει ο άνθρωπος: αυτό της οδήγησης.

Χρησιμοποιείται ένα από τα πιο απαιτητικά σύνολα πραγματικών δεδομένων στον τομέα της αυτόνομης πλοήγησης. Αξιοποιούνται επτά αισθητήρες, ένα LiDAR και έξι κάμερες, για την πλήρη κάλυψη του περιβάλλοντα χώρου του αυτόνομου οχήματος. Συνολικά, αναπτύσσονται τέσσερις διακριτές και αλληλένδετες μονάδες: η μονάδα εξαγωγής χαρακτηριστικών, η μονάδα μάθησης πολλαπλοτήτων, η μονάδα παρακολούθησης πολλαπλών στόχων και η μονάδα ανίχνευσης κινδύνου.

Η μονάδα εξαγωγής χαρακτηριστικών εξάγει γεωμετρικά χαρακτηριστικά από έναν καινοτόμο τρισδιάστατο ανιχνευτή βασισμένο σε νέφος σημείων, προβάλει τους τρισδιάστατους εντοπισμούς στην αντίστοιχη κάμερα και εξάγει εμφανισιακά χαρακτηριστικά μέσω ενός πασίγνωστου δισδιάστατου ανιχνευτή και τμηματοποιητή. Στην μονάδα μάθησης πολλαπλότητων γίνεται αναπαράσταση αυτών των υψηλοδιάστατων χαρακτηριστικών μέσω γραμμικών και μη γραμμικών μεθόδων. Η μονάδα παρακολούθησης πολλαπλών στόχων, μέσω μεθόδων μηχανικής μάθησης, χρησιμοποιεί αυτά τα χαρακτηριστικά ως μέσο παρακολούθησης. Τέλος, η μονάδα ανίχνευσης κινδύνου, χρησιμοποιεί τα δεδομένα παρακολούθησης για τον αναλυτικό υπολογισμού κινδύνου σύγκρουσης. Όλες οι μονάδες αναπτύχθηκαν με στόχο την χρήση τους σε πραγματικό χρόνο.

Δημήτριος Σπαθούλας

dimispat1@pme.duth.gr

Τμήμα Μηχανικών Παραγωγής & Διοίκησης

Δημοκρίτειο Πανεπιστήμιο Θράκης

Ιανουάριος 2025

## ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ

**Αυτόνομη πλοϊγηση οχημάτων, Νέφος σημείων, Τρισδιάστατη όραση, Μηχανική αντίληψη, Οπτικό-χωρικά χαρακτηριστικά, Μηχανική μάθηση, Βαθιά συγχώνευση, Παρακολούθηση πολλαπλών στόχων, Χωροχρονική ανάλυση, Πολυτροπικότητα, Προγνωστικά μοντέλα, Ανίχνευση σύγκρουσης, Πραγματικός χρόνος**

## Title

# **3d multi-modal multi-object tracking with machine learning and analytic collision risk calculation for autonomous vehicles navigation**

## **ABSTRACT**

---

Moving towards the world of the future, terms such as machine vision, perception, and machine learning have managed to become integral components of the present, not only in the broader field of robotics but across all sciences. Concepts like recognition, sensation, perception, and attention have never before had such a direct and profound association with anything other than human cognition. This boundary, though still in early stages, is beginning to blur as humanity advances toward the creation and development of true intelligence. The present work examines and employs abstract patterns to first represent them in dimensions comprehensible to humans and then as a means of perception for an autonomous agent. The goal is for the agent, through artificial intelligence methods, to "see" and "perceive" objects and the environment around it, and even to "sense" the imminent danger of collision, the same way a human does unconsciously. All of this is set in one of the most complex and attention-demanding environments humans face: that of driving.

One of the most challenging real-world datasets in the field of autonomous navigation is utilized. Seven sensors are employed, one LiDAR and six cameras, to provide complete coverage of the autonomous vehicle's surrounding environment. Overall, four distinct and interconnected modules are developed: the feature extraction module, the manifold learning module, the multi-object tracking module, and the risk detection module.

The feature extraction module extracts geometric features from an innovative 3D detector based on point clouds, projects the 3D detections onto the corresponding camera, and extracts appearance features using a renowned 2D detector and segmentor. In the manifold learning module, these high-dimensional features are represented through linear and non-linear methods. The multi-object tracking module employs these features as input for tracking using machine learning techniques. Finally, the risk detection module utilizes the tracking data to perform an analytical collision risk assessment. All modules were developed with a focus on

real-time application.

Dimitrios Spathoulas

dimispat1@pme.duth.gr

Department of Industrial Engineering and Management

Democritus University of Thrace

January 2025

## KEYWORDS

**Autonomous vehicle navigation, Point cloud, 3d computer vision, Machine perception, Visual-spatial features, Machine learning, Deep fusion, Multi-object tracking, Spatiotemporal analysis, Multimodality, Predictive models, Collision detection, Real-time**

# ΠΕΡΙΕΧΟΜΕΝΑ

<b>ΕΥΧΑΡΙΣΤΙΕΣ</b>	.....	VII
<b>ΠΡΟΛΟΓΟΣ</b>	.....	VIII
<b>ABSTRACT</b>	.....	X
<b>ΠΕΡΙΕΧΟΜΕΝΑ</b>	.....	XII
<b>ΚΑΤΑΛΟΓΟΣ ΣΧΗΜΑΤΩΝ</b>	.....	XVII
<b>1 ΕΙΣΑΓΩΓΗ</b>	.....	1
1.1 ΣΚΟΠΟΣ ΚΑΙ ΣΥΝΕΙΣΦΟΡΑ	.....	1
1.2 ΔΟΜΗ ΕΡΓΑΣΙΑΣ	.....	1
<b>2 ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΡΘΟ ΚΑΙ ΑΝΑΛΥΣΗ ΕΡΕΥΝΗΤΙΚΗΣ ΠΕΡΙΟΧΗΣ</b>	.....	3
2.1 ΜΗΧΑΝΙΚΗ ΑΝΤΙΛΗΨΗ	.....	3
2.1.1 ΠΟΛΥΤΡΟΠΙΚΟΤΗΤΑ	.....	3
2.1.2 ΠΑΡΑΤΗΡΗΣΙΜΟΤΗΤΑ	.....	3
2.1.3 ΑΙΣΘΗΤΗΡΕΣ	.....	4
2.1.3.1 ΟΠΤΙΚΗ ΚΑΜΕΡΑ	.....	4
2.1.3.2 ΜΗΧΑΝΙΚΟ LiDAR	.....	5
2.1.3.3 ΒΑΘΜΟΝΟΜΗΣΗ	.....	6
2.1.3.3.1 ΧΡΟΝΙΚΗ ΒΑΘΜΟΝΟΜΗΣΗ	.....	6
2.1.3.3.2 ΕΞΩΓΕΝΗΣ ΒΑΘΜΟΝΟΜΗΣΗ	.....	6
2.1.3.3.3 ΕΝΔΟΓΕΝΗΣ ΒΑΘΜΟΝΟΜΗΣΗ	.....	7
2.1.4 ΔΙΓΡΑΜΜΙΚΗ ΠΑΡΕΜΒΟΛΗ	.....	9
2.1.5 ΤΑΞΙΝΟΜΗΣΗ, ΕΝΤΟΠΙΣΜΟΣ ΚΑΙ ΚΑΤΑΤΜΗΣΗ	.....	9
2.1.6 ΣΥΝΔΙΑΚΥΜΑΝΣΗ	.....	10
2.1.7 ΨΕΥΔΟ-ΑΠΟΣΤΑΣΗ MAHALANOBIS	.....	12
2.1.8 ΠΑΡΑΤΗΡΗΤΕΣ ΚΑΤΑΣΤΑΣΗΣ	.....	12
2.1.8.1 ΘΟΡΥΒΟΣ ΠΑΡΑΤΗΡΗΣΗΣ	.....	12
2.1.8.2 ΘΟΡΥΒΟΣ ΔΙΑΔΙΚΑΣΙΑΣ	.....	12
2.1.8.3 ΘΟΡΥΒΟΣ ΚΑΤΑΣΤΑΣΗΣ	.....	13
2.1.8.4 ΠΑΡΑΚΟΛΟΥΘΗΣΗ	.....	13
2.1.8.5 ΦΙΛΤΡΟ KALMAN	.....	14
2.1.8.6 ΕΚΤΕΤΑΜΕΝΟ ΦΙΛΤΡΟ KALMAN	.....	15
2.2 ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ	.....	17
2.2.1 ΕΝΤΟΠΙΣΜΟΣ ΑΝΤΙΚΕΙΜΕΝΩΝ ΣΕ 2D ΔΕΔΟΜΕΝΑ ΚΑΜΕΡΑΣ	.....	17
2.2.1.1 Η ΑΡΧΙΤΕΚΤΟΝΙΚΗ ΤΟΥ MASK R-CNN	.....	17

2.2.1.2	ΡΑΧΟΚΟΚΑΛΙΑ.....	17
2.2.1.3	ΔΙΚΤΥΑ ΠΥΡΑΜΙΔΑΣ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ .....	17
2.2.1.4	ΔΙΚΤΥΑ ΠΡΟΤΑΣΗΣ ΠΕΡΙΟΧΩΝ .....	18
2.2.1.5	ΕΥΘΥΓΡΑΜΜΙΣΗ ΠΕΡΙΟΧΗΣ ΕΝΔΙΑΦΕΡΟΝΤΟΣ .....	19
2.2.1.6	ΚΕΦΑΛΗ MASK R-CNN .....	19
2.2.1.7	ΕΜΦΑΝΙΣΙΑΚΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ .....	20
2.2.2	<b>ΕΝΤΟΠΙΣΜΟΣ ΑΝΤΙΚΕΙΜΕΝΩΝ ΣΕ 3D ΔΕΔΟΜΕΝΑ LIDAR.....</b>	20
2.2.2.1	Ο ΤΡΙΣΔΙΑΣΤΑΤΟΣ ΕΝΤΟΠΙΣΜΟΣ.....	20
2.2.2.2	ΒΟΞΕΛΟΠΟΙΗΣΗ.....	22
2.2.2.3	3D ΡΑΧΟΚΟΚΑΛΙΑ .....	23
2.2.2.3.1	ΑΡΑΙΕΣ ΣΥΝΕΛΙΞΕΙΣ .....	23
2.2.2.3.2	ΥΠΟΠΟΛΛΑΠΛΕΣ ΑΡΑΙΕΣ ΜΟΝΑΔΕΣ.....	23
2.2.2.4	ΧΑΡΤΗΣ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ.....	24
2.2.2.5	ΚΕΦΑΛΗ ΕΝΤΟΠΙΣΜΟΥ .....	24
2.2.2.5.1	ΠΡΩΤΟ ΣΤΑΔΙΟ: ΧΑΡΤΕΣ ΘΕΡΜΟΤΗΤΑΣ ΚΑΙ ΚΕΦΑΛΙΑ ΠΑΛΙΝΔΡΟΜΗΣΗΣ .....	24
2.2.2.5.2	ΔΕΥΤΕΡΟ ΣΤΑΔΙΟ: ΤΑΞΙΝΟΜΗΣΗ ΚΑΙ ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ.....	25
2.2.2.6	ΤΑ ΜΟΤΙΒΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ.....	26
2.2.3	<b>ΜΕΘΟΔΟΙ ΜΕΙΩΣΗΣ ΔΙΑΣΤΑΣΙΜΟΤΗΤΑΣ .....</b>	26
2.2.3.1	ΓΡΑΜΜΙΚΕΣ ΜΕΘΟΔΟΙ .....	27
2.2.3.1.1	ΑΝΑΛΥΣΗ ΚΥΡΙΩΝ ΣΥΝΙΣΤΩΣΩΝ.....	27
2.2.3.1.2	ΑΠΟΣΥΝΘΕΣΗ ΜΟΝΑΔΙΑΙΩΝ ΤΙΜΩΝ.....	28
2.2.3.2	ΜΗ ΓΡΑΜΜΙΚΕΣ ΤΕΧΝΙΚΕΣ.....	29
2.2.3.2.1	t-ΚΑΤΑΝΕΜΗΜΕΝΗ ΣΤΟΧΑΣΤΙΚΗ ΕΝΣΩΜΑΤΩΣΗ ΓΕΙΤΟΝΩΝ .....	29
2.2.3.2.2	ΟΜΟΙΟΜΟΡΦΗ ΠΡΟΣΕΓΓΙΣΗ ΚΑΙ ΠΡΟΒΟΛΗ ΠΟΛΛΑΠΛΟΤΗΤΩΝ.....	30
2.3	ΒΑΣΗ ΑΝΑΦΟΡΑΣ ΓΙΑ ΤΗΝ ΤΡΙΣΔΙΑΣΤΑΤΗ ΠΑΡΑΚΟΛΟΥΘΗΣΗ ΠΟΛΛΑΠΛΩΝ ΣΤΟΧΩΝ.....	32
2.4	ΕΚΤΙΜΗΣΗ ΚΙΝΔΥΝΟΥ ΣΤΗΝ ΑΥΤΟΝΟΜΗ ΠΛΟΗΓΗΣΗ .....	33
2.4.1	<b>ΠΙΘΑΝΟΤΗΤΑ ΣΥΓΚΡΟΥΣΗΣ ΚΑΤΑΣΤΑΣΗΣ ΚΑΙ ΠΥΚΝΟΤΗΤΑΣ ΠΙΘΑΝΟΤΗΤΑΣ ΣΥΜΒΑΝΤΟΣ ΣΥΓΚΡΟΥΣΗΣ .....</b>	33
2.4.2	ΓΕΝΙΚΗ ΛΥΣΗ .....	34
2.4.3	ΕΠΙΛΥΣΗ ΜΕΣΩ ΠΡΟΣΟΜΟΙΩΣΗΣ MONTE CARLO .....	35
<b>3</b>	<b>ΕΡΓΑΛΕΙΑ ΠΟΥ ΧΡΗΣΙΜΟΠΟΙΗΘΗΚΑΝ .....</b>	<b>37</b>
3.1	PYTHON .....	37
3.2	GITHUB.....	37
3.3	WSL2 ΚΑΙ UBUNTU .....	38
3.4	ANACONDA.....	38
3.5	NUMPY.....	39
3.6	FILTERPY .....	39
3.7	SciPy.....	39
3.7.1	ΣΥΝΑΡΤΗΣΗ ΟΛΟΚΛΗΡΩΣΗΣ ΤΗΣ SciPy.....	39

3.8	TENSORFLOW.....	40
3.9	PYTORCH.....	40
3.9.1	ΠΑΚΕΤΟ ΝΕΥΡΩΝΙΚΩΝ ΔΙΚΤΥΩΝ ΤΗΣ PyTorch.....	41
3.9.2	ΣΥΝΑΡΤΗΣΗ ΠΛΕΓΜΑΤΟΣ ΤΗΣ PyTorch.....	41
3.9.3	ΒΕΛΤΙΣΤΟΠΟΙΗΤΗΣ Adam ΤΗΣ PyTorch .....	42
3.10	SCIKIT-LEARN .....	42
3.11	CUDA ΚΑΙ CUDNN .....	42
3.12	ΠΡΩΤΟΚΟΛΛΟ SSH .....	43
3.13	ΥΠΟΛΟΓΙΣΤΙΚΗ ΥΠΟΔΟΜΗ.....	44
3.14	ΑΠΟΘΕΤΗΡΙΑ .....	44
3.15	ΤΟ ΣΥΝΟΛΟ ΔΕΔΟΜΕΝΩΝ NUSCENES .....	45
<b>4</b>	<b>ΕΞΑΓΩΓΗ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ ΑΠΟ 3D ΚΑΙ 2D ΑΝΙΧΝΕΥΤΕΣ .....</b>	<b>50</b>
4.1	ΜΟΝΑΔΑ ΕΞΑΓΩΓΗΣ ΓΕΩΜΕΤΡΙΚΩΝ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ ΝΕΦΟΥΣ ΑΠΟ ΤΟΝ ΤΡΙΣΔΙΑΣΤΑΤΟ ΑΝΙΧΝΕΥΤΗ CENTERPOINT .....	50
4.2	ΜΟΝΑΔΑ ΕΞΑΓΩΓΗΣ ΕΜΦΑΝΙΣΙΑΚΩΝ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ ΕΙΚΟΝΟΣΤΟΙΧΕΙΩΝ ΑΠΟ ΤΟΝ ΔΙΣΔΙΑΣΤΑΤΟ ΑΝΙΧΝΕΥΤΗ MASK R-CNN .....	51
4.2.1	ΠΡΟΒΟΛΗ 3D ΑΝΤΙΚΕΙΜΕΝΩΝ ΣΤΗΝ ΑΝΤΙΣΤΟΙΧΗ ΚΑΜΕΡΑ .....	51
4.2.2	ΕΞΑΓΩΓΗ ΔΙΑΝΥΣΜΑΤΟΣ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ ΑΠΟ MASK R-CNN .....	52
4.3	ΤΕΛΙΚΟ ΔΙΑΝΥΣΜΑ ΠΛΗΡΟΦΟΡΙΑΣ .....	53
<b>5</b>	<b>ΜΑΘΗΣΗ ΠΟΛΛΑΠΛΩΝ ΣΥΛΛΟΓΩΝ ΣΤΑ ΓΕΩΜΕΤΡΙΚΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΤΟΥ ΤΡΙΣΔΙΑΣΤΑΤΟΥ ΑΝΙΧΝΕΥΤΗ CENTERPOINT ΚΑΙ ΣΤΑ ΕΜΦΑΝΙΣΙΑΚΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΤΟΥ ΔΙΣΔΙΑΣΤΑΤΟΥ ΑΝΙΧΝΕΥΤΗ MASK R-CNN .....</b>	<b>54</b>
5.1	ΓΡΑΜΜΙΚΗ ΑΝΑΛΥΣΗ ΔΙΑΚΡΙΣΗΣ.....	54
5.2	ΑΠΟΜΟΝΩΜΕΝΕΣ ΠΟΛΛΑΠΛΟΤΗΤΕΣ .....	55
5.2.1	ΚΑΤΩΦΛΙ ΑΠΟΔΟΧΗΣ .....	55
5.2.1.1	ΑΝΑΛΥΣΗ ΚΥΡΙΩΝ ΣΥΝΙΣΤΩΣΩΝ .....	56
5.2.1.1.1	ΕΜΦΑΝΙΣΙΑΚΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ .....	56
5.2.1.1.2	ΓΕΩΜΕΤΡΙΚΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ .....	57
5.2.1.2	UMAP .....	57
5.2.2	ΑΠΟΘΟΡΥΒΟΠΟΙΗΣΗ .....	59
5.2.2.1	ΑΝΑΛΥΣΗ ΚΥΡΙΩΝ ΣΥΝΙΣΤΩΣΩΝ .....	60
5.2.2.1.1	ΓΕΩΜΕΤΡΙΚΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ .....	60
5.2.2.1.2	ΕΜΦΑΝΙΣΙΑΚΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ .....	61
5.2.2.2	UMAP .....	61
5.2.3	ΣΥΓΚΡΙΣΗ ΜΕΤΑΞΥ ΚΑΤΩΦΛΙΟΥ ΚΑΙ ΑΠΟΘΟΡΥΒΟΠΟΙΗΣΗΣ .....	63
5.2.4	ΣΥΓΚΡΙΣΗ ΜΕΤΑΞΥ ΚΥΨΕΛΩΝ .....	64

5.2.5	<i>ΧΡΗΣΗ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ</i>	66
<b>6</b>	<b>ΒΑΘΙΑ ΣΥΓΧΩΝΕΥΣΗ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ ΓΙΑ ΤΗΝ ΤΡΙΣΔΙΑΣΤΑΣΗ ΠΟΛΥΤΡΟΠΙΚΗ ΠΑΡΑΚΟΛΟΥΘΗΣΗ ΠΟΛΛΑΠΛΩΝ ΣΤΟΧΩΝ</b>	<b>67</b>
6.1	ΠΙΘΑΝΟΚΡΑΤΙΚΗ ΤΡΙΣΔΙΑΣΤΑΣΗ ΠΑΡΑΚΟΛΟΥΘΗΣΗ ΠΟΛΛΑΠΛΩΝ ΣΤΟΧΩΝ ΓΙΑ ΑΥΤΟΝΟΜΗ ΠΛΟΗΓΗΣΗ	67
6.1.1	<i>ΜΟΝΤΕΛΟΠΟΙΗΣΗ ΦΙΛΤΡΟΥ KALMAN</i>	68
6.1.2	<i>ΚΑΤΑΣΚΕΥΗ ΠΙΝΑΚΩΝ ΣΥΝΔΙΑΚΥΜΑΝΣΗΣ ΜΕΣΩ ΤΟΥ ΣΥΝΟΛΟ ΔΕΔΟΜΕΝΩΝ</i>	69
6.1.3	<i>ΣΥΣΧΕΤΙΣΗ ΚΑΙ ΑΝΤΙΣΤΟΙΧΗΣΗ</i>	70
6.1.3.1	<i>ΣΥΣΧΕΤΙΣΗ ΔΕΔΟΜΕΝΩΝ</i>	70
6.1.3.2	<i>ΑΝΤΙΣΤΟΙΧΙΣΗ ΔΕΔΟΜΕΝΩΝ</i>	70
6.1.4	<i>ΓΕΝΕΣΗ ΚΑΙ ΘΑΝΑΤΟΣ ΙΧΝΩΝ</i>	71
6.2	ΤΡΙΣΔΙΑΣΤΑΣΗ ΠΟΛΥΤΡΟΠΙΚΗ ΠΑΡΑΚΟΛΟΥΘΗΣΗ ΠΟΛΛΑΠΛΩΝ ΣΤΟΧΩΝ ΜΕΣΩ ΓΕΩΜΕΤΡΙΚΩΝ ΚΑΙ ΕΜΦΑΝΙΣΙΑΚΩΝ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ	72
6.2.1	<i>ΕΞΩΤΕΡΙΚΗ ΑΡΧΙΤΕΚΤΟΝΙΚΗ ΜΟΝΑΔΑΣ ΠΑΡΑΚΟΛΟΥΘΗΣΗΣ</i>	73
6.2.2	<i>ΜΟΝΑΔΑ ΣΥΓΧΩΝΕΥΣΗΣ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ</i>	74
6.2.3	<i>ΧΑΡΤΗΣ ΣΥΝΕΝΩΜΕΝΩΝ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ</i>	75
6.2.4	<i>ΜΗΤΡΑ ΒΑΘΙΑΣ ΑΠΟΣΤΑΣΗΣ ΣΥΝΔΥΑΣΜΩΝ</i>	75
6.2.5	<i>ΟΛΙΚΟΣ ΣΧΕΔΙΑΣΜΟΣ ΕΚΠΑΙΔΕΥΣΗΣ ΚΑΙ ΣΥΜΠΕΡΑΣΜΟΥ</i>	77
6.2.6	<i>ΑΠΟΤΕΛΕΣΜΑΤΑ</i>	78
6.2.6.1	<i>ΜΕΤΡΙΚΕΣ</i>	78
6.2.6.2	<i>ΣΥΓΚΡΙΣΗ ΑΠΟΤΕΛΕΣΜΑΤΩΝ ΔΙΑΦΟΡΕΤΙΚΩΝ ΔΙΑΚΥΜΑΝΣΕΩΝ ΒΑΣΗ ΑΝΙΧΝΕΥΤΩΝ</i>	80
6.2.6.3	<i>ΣΥΓΚΡΙΣΗ ΜΕ ΜΕΘΟΔΟΥΣ ΒΑΣΗΣ</i>	81
6.2.6.4	<i>ΑΝΑΛΥΣΗ ΕΥΑΙΣΘΗΣΙΑΣ ΠΑΡΑΓΟΝΤΑ ΑΝΑΜΕΙΞΗΣ</i>	83
6.2.6.5	<i>ΑΝΑΛΥΣΗ ΕΥΑΙΣΘΗΣΙΑΣ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ</i>	83
<b>7</b>	<b>ΑΝΑΛΥΤΙΚΟΣ ΥΠΟΛΟΓΙΣΜΟΣ ΡΙΣΚΟΥ ΣΥΓΚΡΟΥΣΗΣ ΠΟΛΛΑΠΛΩΝ ΣΤΟΧΩΝ ΓΙΑ ΤΗΝ ΑΥΤΟΝΟΜΗ ΠΛΟΗΓΗΣΗ</b>	<b>85</b>
7.1	ΥΠΟΛΟΓΙΣΜΟΣ ΤΡΟΧΙΑΣ ΚΑΙ ΚΑΤΑΣΚΕΥΗ ΔΙΕΥΡΥΝΣΗΣ ΠΕΡΙΟΧΗΣ ΑΣΦΑΛΕΙΑΣ	85
7.2	ΑΝΑΛΥΤΙΚΟΣ ΥΠΟΛΟΓΙΣΜΟΣ ΧΩΡΙΚΗΣ ΕΠΙΚΑΛΥΨΗΣ	86
<b>8</b>	<b>ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΠΕΡΑΙΤΕΡΩ ΕΡΕΥΝΑ</b>	<b>90</b>
8.1	ΣΥΜΠΕΡΑΣΜΑΤΑ	90
8.1.1	<i>ΜΟΝΑΔΑ ΜΑΘΗΣΗΣ ΠΟΛΛΑΠΛΟΤΗΤΩΝ</i>	90
8.1.2	<i>ΜΟΝΑΔΑ ΠΑΡΑΚΟΛΟΥΘΗΣΗΣ ΠΟΛΛΑΠΛΩΝ ΣΤΟΧΩΝ</i>	90
8.1.3	<i>ΜΟΝΑΔΑ ΑΝΙΧΝΕΥΣΗΣ ΚΙΝΔΥΝΟΥ</i>	91
8.2	ΠΕΡΑΙΤΕΡΩ ΕΡΕΥΝΑ	91
<b>ΒΙΒΛΙΟΓΡΑΦΙΑ</b>		<b>93</b>



# ΚΑΤΑΛΟΓΟΣ ΣΧΗΜΑΤΩΝ

Εικόνα 2-1: Ενδογενής και εξωγενής μετασχηματισμοί.....	9
Εικόνα 2-2: Ταξινόμηση, Εντοπισμός και Κατάτμηση .....	10
Εικόνα 2-3: Κύριοι άξονες δισδιάστατης συνδιακύμανσης .....	11
Εικόνα 2-4: Φίλτρο Kalman .....	14
Εικόνα 2-5: Μεθοδολογία Φίλτρου Kalman.....	14
Εικόνα 2-6: Ραχοκοκαλιά ResNet .....	17
Εικόνα 2-7: Δίκτυο Πυραμίδας Χαρακτηριστικών του Mask R-CNN .....	18
Εικόνα 2-8: Δίκτυο Πρότασης Περιοχών του Mask R-CNN.....	19
Εικόνα 2-9: Διγραμμική Παρεμβολή .....	19
Εικόνα 2-10: Κεφαλή Mask R-CNN .....	20
Εικόνα 2-11: Αρχιτεκτονική Mask R-CNN .....	20
Εικόνα 2-12: Τρισδιάστατος Εντοπισμός.....	21
Εικόνα 2-13: Αρχιτεκτονική Τρισδιάστατου Ανιχνευτή CenterPoint.....	22
Εικόνα 2-14: Τεχνική Βοξελοποίησης.....	23
Εικόνα 2-15: Υποπολλαπλές Αραιές Μονάδες .....	24
Εικόνα 2-16: Χάρτης Χαρακτηριστικών .....	24
Εικόνα 2-17: Πρώτο στάδιο Κεφαλής.....	25
Εικόνα 2-18: Δεύτερο στάδιο Κεφαλής .....	26
Εικόνα 2-19: Γεωμετρικά Χαρακτηριστικά .....	26
Εικόνα 2-20: Βασική Αρχιτεκτονική Παρακολούθησης.....	32
Εικόνα 2-21: Γεγονότα Σύγκρουσης .....	34
Εικόνα 2-22: Γενική Λύση στο πρόβλημα Υπολογισμού Κίνδυνου Σύγκρουσης.....	35
Εικόνα 3-1: Python .....	37
Εικόνα 3-2: GitHub.....	38
Εικόνα 3-3: TensorFlow .....	40
Εικόνα 3-4: PyTorch.....	41
Εικόνα 3-5: Cuda.....	43
Εικόνα 3-6: CuDNN .....	43
Εικόνα 3-7: OpenMMLab.....	44
Εικόνα 3-8: Όψη και αισθητήρες του αυτόνομου οχήματος από το σύνολο δεδομένων .....	45
Εικόνα 3-9: Στιγμιότυπο σκηνής συνόλου δεδομένων .....	46
Εικόνα 3-10: Αισθητήρες που χρησιμοποιούνται .....	46
Εικόνα 3-11: Πεδίο όρασης αισθητήρων .....	47
Εικόνα 3-12: Αρχιτεκτονική Δεδομένων .....	48
Εικόνα 4-1: Προβαλλόμενα πλαίσια εντοπισμών CenterPoint στην αντίστοιχη κάμερα (κατώφλι αποδοχής 0.03)	

.....	51
Εικόνα 4-2: Προβαλλόμενα πλαισία εντοπισμών στην αντίστοιχη κάμερα (κατώφλι εντοπισμού 0.81) .....	52
Εικόνα 5-1: Γραμμική ανάλυση διάκρισης στα εμφανισιακά και γεωμετρικά χαρακτηριστικά εντοπισμών (κατώφλι αποδοχής 0.57).....	55
Εικόνα 5-2: Πλήθος εντοπισμών (κατώφλι αποδοχής 0.57) .....	56
Εικόνα 5-3: Τρισδιάστατη αναπαράσταση εμφανισιακών χαρακτηριστικών μέσω ανάλυσης κύριων συνιστωσών (αριστερά αυτοκίνητα, δεξιά ποδήλατα) .....	56
Εικόνα 5-4: Τρισδιάστατη αναπαράσταση εμφανισιακών χαρακτηριστικών μέσω ανάλυσης κύριων συνιστωσών (αριστερά φορτηγά, δεξιά λεωφορεία) .....	57
Εικόνα 5-5: Τρισδιάστατη αναπαράσταση γεωμετρικών χαρακτηριστικών μέσω ανάλυσης κύριων συνιστωσών (αριστερά μοτοσικλέτες, δεξιά ποδήλατα) .....	57
Εικόνα 5-6: Δισδιάστατή αποτύπωση εμφανισιακών χαρακτηριστικών μέσω UMAP (αριστερά ποδήλατα, δεξιά μοτοσικλέτες) .....	58
Εικόνα 5-7: Τρισδιάστατη αποτύπωση εμφανισιακών χαρακτηριστικών μέσω UMAP (αριστερά ποδήλατα, δεξιά μοτοσικλέτες) .....	58
Εικόνα 5-8: Δισδιάστατή αποτύπωση γεωμετρικών χαρακτηριστικών μέσω UMAP (αριστερά αυτοκίνητα, δεξιά ποδήλατα) .....	59
Εικόνα 5-9: Πλήθος αντικειμένων (κατώφλι αποδοχής 0.03) .....	60
Εικόνα 5-10: Τρισδιάστατη αναπαράσταση γεωμετρικών χαρακτηριστικών μέσω ανάλυσης κύριων συνιστωσών (αριστερά φορτηγά, δεξιά τροχοφόρα) .....	61
Εικόνα 5-11: Τρισδιάστατη αναπαράσταση εμφανισιακών χαρακτηριστικών ποδηλάτων μέσω ανάλυσης κύριων συνιστωσών .....	61
Εικόνα 5-12: Τρισδιάστατη αποτύπωση γεωμετρικών χαρακτηριστικών μέσω UMAP (αριστερά λεωφορεία, δεξιά φορτηγά).....	62
Εικόνα 5-13: Τρισδιάστατη αποτύπωση γεωμετρικών χαρακτηριστικών μέσω UMAP (αριστερά αυτοκίνητα, δεξιά πεζοί).....	62
Εικόνα 5-14: Τρισδιάστατη αποτύπωση γεωμετρικών χαρακτηριστικών μέσω UMAP (αριστερά ποδήλατα, δεξιά μοτοσικλέτες) .....	63
Εικόνα 5-15: Σύγκριση τρισδιάστατης αποτύπωσης εμφανισιακών χαρακτηριστικών μέσω UMAP για τα αυτοκίνητα .....	64
Εικόνα 5-16:Σύγκριση γεωμετρικών χαρακτηριστικών διαφορετικής κυψέλης μέσω PCA για τα αυτοκίνητα (αριστερά 0.1, δεξιά 0.075) .....	64
Εικόνα 5-17: Σύγκριση γεωμετρικών χαρακτηριστικών διαφορετικής κυψέλης μέσω UMAP (αριστερά 0.1 και φορτηγά, δεξιά 0.075 και τροχοφόρα) .....	65
Εικόνα 5-18: Έλεγχος και επαλήθευση μορφολογίας εμφανισιακών χαρακτηριστικών μέσω UMAP από διαφορετικά μεγέθη κυψέλης (φιλτράρισμα και κατώφλι ανίχνευσης 0.57) .....	65
Εικόνα 5-19: Τρισδιάστατη και δισδιάστατη αναπαράσταση εμφανισιακών χαρακτηριστικών μέσω UMAP	

τροχοφόρων διαφορετικών σεναρίων εντοπισμού .....	66
Εικόνα 6-1:Πιθανοκρατική Τρισδιάστατη Παρακολούθηση Πολλαπλών Στόχων .....	67
Εικόνα 6-2:Βαθιά Συγχώνευση Χαρακτηριστικών για την Τρισδιάστατη Πολυτροπική Παρακολούθηση Πολλαπλών Στόχων .....	73
Εικόνα 6-3: Μονάδα Συγχώνευσης Χαρακτηριστικών .....	75
Εικόνα 6-4: Χάρτης Συνενωμένων Χαρακτηριστικών και Μήτρα Βαθιάς Απόστασης Συνδυασμών .....	76
Εικόνα 6-5: Αποτελέσματα παρακολούθησης ανά κλάση συνδιακύμανσης ανιχνευτή MEGVII (κατώφλι εντοπισμών εκπαίδευσης και συμπερασμού: 0.01) .....	80
Εικόνα 6-6:Αποτελέσματα παρακολούθησης ανά κλάση συνδιακύμανσης ανιχνευτή CenterPoint (κατώφλι εντοπισμών εκπαίδευσης 0.57 και συμπερασμού 0.03) .....	80
Εικόνα 6-7: Σύγκριση ολικών μετρικών (αριστερά CenterPoint, δεξιά MEGVII) .....	80
Εικόνα 6-8: Αποτελέσματα ανά κλάση αναφοράς μεθοδολογίας με Mahalanobis (κατώφλι αποδοχής 0.57) ..	81
Εικόνα 6-9: Ολικά Αποτελέσματα μεθοδολογίας Mahalanobis .....	81
Εικόνα 6-10: Αποτελέσματα ανά κλάση με Πίνακα Βαθιάς Απόστασης Συνδυασμών (κατώφλι αποδοχής 0.57, παράγοντας ανάμειξης 0.45 και κατώφλι αποδοχής συσχέτισης 0.9) .....	82
Εικόνα 6-11: Ολικές Μετρικές Παρακολούθησης με Πίνακα Βαθιάς Απόστασης Συνδυασμών .....	82
Εικόνα 6-12: Αποτελέσματα βάσης αναφοράς μεθοδολογίας (κατώφλι αποδοχής 0.01) .....	83
Εικόνα 6-13: Ανάλυση Ευαισθησίας Παράγοντα Ανάμειξης (κατώφλι αποδοχής συσχέτισης 0.85) .....	83
Εικόνα 6-14: Ανάλυση Ευαισθησίας Χαρακτηριστικών .....	84
Εικόνα 7-1: Συνέλιξη μεταξύ Αυτόνομου Οχήματος και Εμποδίου-Ιχνους .....	86
Εικόνα 7-2: Περιοχή Τομής μεταξύ Οκταγώνου και Συνδιακύμανσης .....	86
Εικόνα 7-3: Υπολογισμός κινδύνου σύγκρουσης (οκτάγωνο κάτω από τον άξονα x, αντιστροφή συνεισφορών) .....	88
Εικόνα 7-4: Υπολογισμός κινδύνου σύγκρουσης (τμήματα οκταγώνου άνω και κάτω οριζόντιου άξονα) .....	89

# **1 ΕΙΣΑΓΩΓΗ**

---

## **1.1 ΣΚΟΠΟΣ ΚΑΙ ΣΥΝΕΙΣΦΟΡΑ**

Σκοπός της εργασίας είναι η κατανόηση και η ανάπτυξη ενός συστήματος παρακολούθησης πολλαπλών στόχων και της αξιολόγησης του κινδύνου της σύγκρουσης τους με ένα αυτόνομο όχημα (A.O) σε πραγματικό χρόνο με στόχο την ασφαλή αυτόνομη πλοιήγηση οχημάτων. Το σύστημα βασίζεται στον εντοπισμό τρισδιάστατων αντικειμένων γύρω του οχήματος, μέσω πολλαπλών (έξι, 6) καμερών και ενός LiDAR. Η κατανόηση των γεωμετρικών και εμφανισιακών χαρακτηριστικών των εντοπιστών μπορεί να ενισχύσει άμεσα την κατανόηση και επαγωγικά τον τρόπο αντιμετώπισης και υλοποίησης αντίστοιχων προβλημάτων και ερευνητικά αλλά και σε πραγματικές εφαρμογές. Η παρακολούθηση είναι κρίσιμο σημείο στον σχεδιασμό αυτονομίας για την λήψη αποφάσεων (decision making), ενισχύοντας τον σχεδιασμό τροχιάς πορείας (path planning) του A.O. και την αποφυγή εμποδίων (obstacle avoidance) του χώρου. Αποτέλεσμα της εργασίας είναι ότι τα γεωμετρικά και εμφανισιακά χαρακτηριστικά που εξάγονται μπορούν να πλησιάσουν μία από τις ταχείες και εύρωστες τεχνικές παρακολούθησης, αλλά και να ξεπεράσουν σημαντικά τις πιο βασικές και ευρέως γνωστές αντιμετωπίσεις. Με τον αναλυτικό υπολογισμό κινδύνου το σύστημα απόφασης μπορεί ταχύτερα να καταλήξει σε σημαντικές αποφάσεις για την ασφάλεια και του ίδιο του οδηγού αλλά και του γύρω περιβάλλοντος.

Τα τέσσερα βασικά συστήματα (ολική μονάδα εξαγωγής χαρακτηριστικών, μονάδα μάθησης πολλαπλοτήτων, μονάδα παρακολούθησης πολλαπλών στόχων και μονάδα ανίχνευσης κινδύνου) αποτελούν κομμάτι συνεισφοράς στον ευρύτερο τομέα της αυτονομίας και της πλοιήγησης οχημάτων, πιο συγκεκριμένα στον χώρο της τρισδιάστατης ρομποτικής όρασης και αντίληψης. Με μικρές αλλαγές τα μοντέλα μπορούν να λειτουργήσουν και για τον αέρα (π.χ. UAVs) και για το νερό (π.χ. υποβρύχια) καθώς και σε στρατιωτικές εφαρμογές (π.χ. αμυντικά συστήματα).

## **1.2 ΔΟΜΗ ΕΡΓΑΣΙΑΣ**

Στο δεύτερο κεφάλαιο γίνεται θεωρητική επισκόπηση της μηχανικής αντίληψης, των αισθητήρων που χρησιμοποιούνται, στην αβεβαιότητα καθώς και στον παρατηρητή

κατάστασης Kalman. Στην συνέχεια γίνεται μία λεπτομερής ανάλυση των δισδιάστατων και τρισδιάστατων ανιχνευτών Mask R-CNN και CenterPoint και της τρισδιάστατης παρακολούθησης πολλαπλών στόχων, ακολουθημένης μιας λεπτομερής διατύπωσης της ποσοτικοποιημένης ερμηνείας κινδύνου στην αυτόνομη πλοιήγηση. Στο τρίτο κεφάλαιο, αναλύονται οι βασικότερες συναρτήσεις, εργαλεία, δομές καθώς και το σύνολο δεδομένων που χρησιμοποιήθηκαν. Στο επόμενο κεφάλαιο, εξηγείται λεπτομερειακά η μεθοδολογία εξαγωγής γεωμετρικών και εμφανισιακών χαρακτηριστικών και στο πέμπτο ερμηνεύονται οι πολλαπλότητες τους. Στο έκτο γίνεται χρήση αυτών μέσω βαθιάς συγχώνευσης για την παρακολούθηση πολλαπλών στόχων, γίνεται σύγκριση αποτελεσμάτων με άλλες μεθόδους, ακολουθούμενες από διάφορες αναλύσεις ευαισθησίας. Στην συνέχεια, τα αποτελέσματα χρησιμοποιούνται για την ανίχνευση κινδύνου μέσω ενός αναλυτικού υπολογισμού χωρικής επικάλυψης σε πραγματικό χρόνο. Στο όγδοο και τελευταίο κεφάλαιο συζητιόνται τα συμπεράσματα των διαφόρων συστημάτων που απαρτίζουν την εργασία και προτείνεται περαιτέρω έρευνα.

## **2 ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΡΘΟ ΚΑΙ ΑΝΑΛΥΣΗ ΕΡΕΥΝΗΤΙΚΗΣ ΠΕΡΙΟΧΗΣ**

---

### **2.1 ΜΗΧΑΝΙΚΗ ΑΝΤΙΛΗΨΗ**

#### **2.1.1 ΠΟΛΥΤΡΟΠΙΚΟΤΗΤΑ**

Με τον όρο πολυτροπικότητα εννοείται γενικά η συγχώνευση πολλών αισθητήρων ή και πρακτόρων με σκοπό την μείωση των πιθανών σφαλμάτων μετρήσεων και επαγωγικά της μεγαλύτερης ακρίβειας αυτών. Ο κόσμος δεν είναι τέλειος και οι μετρήσεις των αισθητήρων είναι επιρρεπής σε σφάλματα ακρίβειας χαμένων αλλά ακόμα και λανθασμένων μετρήσεων. Η συγχώνευση αισθητήρων μειώνει την αβεβαιότητα των εκτιμήσεων για την πιθανή κατάσταση των αντικειμένων ενδιαφέροντος αφού υπάρχουν πάνω από μία πηγές επιβεβαίωσης των μετρήσεων με ποίκιλες μορφές εξόδου της ίδιας πληροφορίας. Η μοντελοποίηση αυτής της κατάστασης είναι στοχαστική και όχι ντετερμινιστική, έτσι η θέση ενός αντικειμένου δεν είναι μια πραγματική τιμή, αλλά μία εκτιμώμενη μέση μιας πυκνότητας πιθανότητας [1].

#### **2.1.2 ΠΑΡΑΤΗΡΗΣΙΜΟΤΗΤΑ**

Η παρατηρησιμότητα αναφέρεται στην ικανότητα ενός συστήματος αντίληψης (ή όρασης) να προσδιορίσει την πλήρη κατάσταση του μέσω των εξόδων του και των διαθέσιμων μετρήσεων από τους αισθητήρες. Δηλαδή, ένα σύστημα είναι παρατηρήσιμο εάν, μέσω των μετρήσεων που συλλέγονται από τους αισθητήρες, μπορεί να συναχθεί με ακρίβεια το διάνυσμα κατάστασης του συστήματος σε οποιαδήποτε χρονική στιγμή.

Κρίσιμο βήμα ώστε το σύστημα να είναι επαρκώς παρατηρήσιμο είναι η σωστή καθιέρωση του διανύσματος κατάστασης. Για παράδειγμα, δεν είναι αναγκαία η μέτρηση της θερμοκρασίας του κινητήρα εσωτερικής καύσης (MEK) του οχήματος για να γνωρίζει το σύστημα πού βρίσκεται στον χώρο, αλλά είναι εύλογη η μέτρηση του προσανατολισμού του με μία αδρανειακή μονάδα μέτρησης (Inertial Measurement Unit, IMU) [2].

Ακόμα ένα κρίσιμο σημείο είναι το πεδίο θέασης του συστήματος όρασης και η συμπληρωματικότητα μεταξύ των αισθητήρων. Για παράδειγμα, δεν είναι αναγκαίο να υπάρχει μια κάμερα για να εντοπίζει αντικείμενα εκτός του χώρου εργασίας ενός ρομποτικού

βραχίονα, αλλά είναι αναγκαία εντός του χώρου. Συμπερασματικά, το σύστημα αντίληψης λαμβάνει την μορφή που απαιτείται ανάλογα τους στόχους και σκοπούς της εργασίας. Υπάρχουν περιπτώσεις όπου χρειάζεται ενδελεχής σχεδιασμός και τοποθέτηση των αισθητήρων ώστε να εξασφαλιστεί ότι το επιθυμητό διάνυσμα κατάστασης μπορεί να παρατηρηθεί πλήρως.

### 2.1.3 ΑΙΣΘΗΤΗΡΕΣ

Γενικά, οι αισθητήρες ταξινομούνται σε δύο κατηγορίες με βάση την χρήση τους προς το σύστημα, στους ιδιοδεκτικούς και εξωδεκτικούς. Οι ιδιοδεκτικοί αισθητήρες (proprioceptive sensors) ή εσωτερικής κατάστασης, καταγράφουν και μετρούν τις εσωτερικές τιμές του συστήματος άμεσα ελεγχόμενες από το ίδιο. Οι εξωδεκτικοί αισθητήρες (exteroceptive sensors) ή εξωτερικής κατάστασης, ανιχνεύουν και συλλέγουν πληροφορίες σχετικά με το περιβάλλον γύρω από το σύστημα και εκτός των ορίων ελέγχου του.

Οι αισθητήρες μπορούν επίσης να χωριστούν σε ενεργητικούς και παθητικούς, ανάλογα με τις αρχές λειτουργίας τους. Οι παθητικοί λαμβάνουν πληροφορία από το περιβάλλον για να παράγουν εξόδους, χωρίς να εξάγουν κάποιο σήμα. Αντίθετα, οι ενεργητικοί εκπέμπουν πληροφορία στο περιβάλλον και μετρούν την αντίδραση του περιβάλλοντος σε αυτήν.

#### 2.1.3.1 ΟΠΤΙΚΗ ΚΑΜΕΡΑ

Η κάμερα λειτουργεί ως μετατροπέας φωτός σε ψηφιακά σήματα. Αυτό επιτυγχάνεται μέσω διαφόρων τεχνολογιών, όπως τα CCD (Charge-Coupled Devices) και CMOS (Complementary Metal-Oxide-Semiconductor)<sup>1</sup>. Η έξοδος είναι συνήθως ένας τανυστής τριών διαστάσεων (μήκος, πλάτος, χρώμα), που αντιστοιχεί στο μέγεθος της εικόνας και στον τρόπο με τον οποίο το χρώμα εκφράζεται σε αυτήν, συνήθως χρησιμοποιώντας το σύστημα RGB (Red, Green, Blue). Έτσι, κάθε εικονοστοιχείο (pixel) της εικόνας αποτελείται από έναν συνδυασμό των τριών βασικών χρωμάτων, με καθένα να έχει 256 επίπεδα έντασης (0-255) τα οποία διαμορφώνονται μέσω των εσωτερικών της φίλτρων.

Η συχνότητα δειγματοληψίας στην κάμερα είναι ο χρόνος μεταξύ της καταγραφής δύο διαδοχικών εικόνων. Κάθε κάμερα έχει συγκεκριμένο πεδίο θέασης που μπορεί να αποτυπώσει. Η εικόνα αποτυπώνεται την πληροφορία της υφής σε ένα δισδιάστατο επίπεδο, με την πληροφορία του βάθους ωστόσο να συμπιέζεται. Η απόδοση της επηρεάζεται άμεσα

---

<sup>1</sup> <https://www.nuscenes.org/>

από τον φωτισμό και τις καιρικές συνθήκες.

### 2.1.3.2 ΜΗΧΑΝΙΚΟ LiDAR

Ο αισθητήρας εντοπισμού φωτός και απόστασης (Light Detection And Ranging, LiDAR) εκπέμπει παλμούς ακτινοβολίας λέιζερ και υπολογίζει τον χρόνο ανάκλασης του σήματος (Time of Flight, ToF). Η βασική ιδιότητα που εκμεταλλεύονται είναι η μη διάχυση της ακτίνας laser λόγω υψηλής συνεκτικότητας, επιτρέποντας της έτσι να ταξιδέψει μέχρι τον στόχο και να επιστρέψει στον σαρωτή. Το τρισδιάστατο σύστημα συντεταγμένων του LiDAR συνήθως είναι κυλινδρικό. Η έξοδος απεικονίζει τον τρισδιάστατο χώρο γύρω του. Παρέχει πληροφορίες για τη μορφολογία των αντικειμένων στον χώρο. Για μία ακτίνα η απόσταση της από τον αισθητήρα υπολογίζεται ως:

$$d = \frac{c \cdot t}{2} \quad (2.1)$$

Με c ταχύτητα φωτός και t ο χρόνος πτήσης.

Κατά βάση δημιουργεί ένα σύνολο από τέτοια σημεία με συντεταγμένες στον χώρο. Το τελικό προϊόν της διαδικασίας ονομάζεται νέφος σημείων (point cloud), το οποίο μπορεί να χαρακτηριστεί ως αραιό ή πυκνό ανάλογα. Στα πλαίσια των εφαρμογών της αυτόνομης πλοήγησης το νέφος είναι συνήθως αραιό γιατί ο αισθητήρας βρίσκεται σε ανοιχτούς χώρους. Αντίθετα, (αν και όχι κατά κανόνα) η πυκνή διάταξη των σημείων στο νέφος συνιστά μια πυκνή δειγματοληψία της περιοχής που σαρώθηκε (π.χ. μικρό δωμάτιο, με πολλές επαναλήψεις δειγματοληψίας). Αν και μπορεί να είναι δύσκολο να αναγνωριστούν συγκεκριμένα σημεία εντός του νέφους, μπορεί να αποδοθεί με εξαιρετική λεπτομέρεια η τρισδιάστατη περιοχή σάρωσης, η οποία ωστόσο δεν ενσωματώνει πληροφορίες για την υφή των απεικονιζόμενων οντοτήτων ή σχέσεις σύνδεσης μεταξύ των σημείων, κάτι στο οποίο η κάμερα είναι ιδανική [3].

Το μηχανικό LiDAR περιστρέφεται και σαρώνει σε αρκετά μεγάλη εμβέλεια τον χώρο γύρω του κατά 360 μοίρες, περιορίζεται ωστόσο σε ένα σχετικά μικρό κατακόρυφο πεδίο θέασης. Ο ρυθμός δειγματοληψίας ή ρυθμός επανάληψης παλμών (Pulse Repetition Frequency, PRF) θεωρείται μία ολοκληρωμένη περιστροφή γύρω από το κέντρο του άξονα του. Η έξοδος αποτελείται από τα σημεία στον τρισδιάστατο χώρο, την έντασή τους και ενδεχομένως τον χρόνο πτήσης τους (TOF). Η λειτουργία του επηρεάζεται από καιρικές συνθήκες, όπως σύννεφα και ομίχλη, αλλά όχι έντονα από το φως. Συνήθης ρυθμός δειγματοληψίας είναι 20-

30 Hz.

### 2.1.3.3 ΒΑΘΜΟΝΟΜΗΣΗ

#### 2.1.3.3.1 ΧΡΟΝΙΚΗ ΒΑΘΜΟΝΟΜΗΣΗ

Για να μπορέσει να υπάρξει σωστός εντοπισμός και συγχώνευση μεταξύ αισθητήρων πρέπει να υπάρχει χρονική βαθμονόμηση, δηλαδή βηματικός συγχρονισμός διαθεσιμότητας. Για παράδειγμα, η χρήση IMU και η συγχώνευση με δεδομένα από GPS όταν είναι διαθέσιμα για τον ταυτόχρονο εντοπισμό και χαρτογράφηση (Simultaneous Localization And Mapping, SLAM) [4].

#### 2.1.3.3.2 ΕΞΩΓΕΝΗΣ ΒΑΘΜΟΝΟΜΗΣΗ

Η εξωγενής βαθμονόμηση αναφέρεται στον σχετικό μετασχηματισμό μεταξύ δύο συστημάτων αναφοράς, όπως από ένα πλαίσιο σε ένα άλλο. Αποτελείται από τον πίνακα περιστροφής και το διάνυσμα μετατόπισης. Ο πίνακας περιστροφής εξηγεί την αλλαγή της κατεύθυνσης των διανυσμάτων βάσης, ενώ το διάνυσμα μετατόπισης την σχετική θέση μεταξύ του αρχικού και του τελικού πλαισίου [5]. Για ένα τρισδιάστατο σημείο με βάση το σύστημα αναφοράς  $C_1$ , οι ομογενής συντεταγμένες του είναι:

$$\mathbf{P}_{C_1} = \begin{bmatrix} X_1 \\ Y_1 \\ Z_1 \\ 1 \end{bmatrix} \quad (2.2)$$

Για το ίδιο σημείο στον χώρο με βάση το σύστημα αναφοράς  $C_2$  ισχύει:

$$\mathbf{P}_{C_2} = \begin{bmatrix} X_2 \\ Y_2 \\ Z_2 \\ 1 \end{bmatrix} \quad (2.3)$$

Ο μετασχηματισμός οπού μετατρέπει το  $\mathbf{P}_{C_1}$  στο  $\mathbf{P}_{C_2}$  προς το  $C_2$  γράφεται ως:

$$\mathbf{P}_{C_2} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ 0 & 1 \end{bmatrix} \mathbf{P}_{C_1} \quad (2.4)$$

Οπού  $\mathbf{R}$  πίνακας περιστροφής και  $\mathbf{t}$  διάνυσμα μετατόπισης:

$$\mathbf{R} = \begin{bmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \\ R_{31} & R_{32} & R_{33} \end{bmatrix}, \quad \mathbf{t} = \begin{bmatrix} t_1 \\ t_2 \\ t_3 \\ 1 \end{bmatrix} \quad (2.5)$$

Με

$$\begin{aligned}
 R_{11} &= \cos(\theta_z)\cos(\theta_y) \\
 R_{12} &= -\cos(\theta_z)\sin(\theta_y)\cos(\theta_x) - \sin(\theta_z)\sin(\theta_x) \\
 R_{13} &= \cos(\theta_z)\sin(\theta_y)\sin(\theta_x) - \sin(\theta_z)\cos(\theta_x) \\
 R_{21} &= \sin(\theta_z)\cos(\theta_y) \\
 R_{22} &= -\sin(\theta_z)\sin(\theta_y)\cos(\theta_x) + \cos(\theta_z)\sin(\theta_x) \\
 R_{23} &= \sin(\theta_z)\sin(\theta_y)\sin(\theta_x) + \cos(\theta_z)\cos(\theta_x) \\
 R_{31} &= \sin(\theta_y) \\
 R_{32} &= \cos(\theta_y)\sin(\theta_x) \\
 R_{33} &= \cos(\theta_y)\cos(\theta_x)
 \end{aligned} \tag{2.6}$$

Για τον αντίστροφο μετασχηματισμό:

$$\mathbf{P}_{C_1} = \begin{bmatrix} \mathbf{R}^T & \mathbf{v} \\ 0 & 1 \end{bmatrix} \mathbf{P}_{C_2} \tag{2.7}$$

Όπου  $\mathbf{v} = -\mathbf{t}$  και  $\mathbf{R}^T = \mathbf{R}^{-1}$ .

#### 2.1.3.3.3 ΕΝΔΟΓΕΝΗΣ ΒΑΘΜΟΝΟΜΗΣΗ

Συγκεκριμένα για κάμερες, ο ενδογενής (εσωτερικός) μετασχηματισμός ή βαθμονόμηση προβάλει ένα σημείο από το σύστημα αναφοράς της κάμερας στο επίπεδο αποτύπωσης της (προβολή σημείου στο επίπεδο της κάμερας – οπτική αντιστοίχιση).

Για το τρισδιάστατο σημείο σε ομογενείς συντεταγμένες:

$$\mathbf{P}_{C_o} = \begin{bmatrix} X_1 \\ Y_1 \\ Z_1 \\ 1 \end{bmatrix} \tag{2.8}$$

Στο πλαίσιο αναφοράς  $C_o$  της κάμερας. Για την χαρτογράφηση του σημείου αυτού στον χώρο των pixel της κάμερας πρέπει:

$$\mathbf{P}_{C_1} = \mathbf{K} \times \mathbf{P}_{C_o} \tag{2.9}$$

Με  $\mathbf{K}$ η μήτρα ενδογενούς βαθμονόμησης που ορίζεται ως:

$$\mathbf{K} = \begin{bmatrix} f_x & s & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \tag{2.10}$$

Όπου:

$f_x, f_y$  Οι εστιακές αποστάσεις. Αντιπροσωπεύουν τη μεγεθυντική ισχύ της κάμερας κατά μήκος των μοναδιαίων διανυσμάτων βάσης του πλαισίου  $i$  και  $j$ . Εξαρτώνται από τα οπτικά χαρακτηριστικά του φακού και του αισθητήρα της κάμερας.

$s$ : Συντελεστής διαγώνιας ασυμμετρίας. Εκφράζει την παραμόρφωση που μπορεί να υπάρχει μεταξύ των δύο αξόνων του αισθητήρα (τάνυση).

$c_x, c_y$ : Συντεταγμένες του κύριου σημείου. Αναφέρονται στο κέντρο του αισθητήρα της κάμερας και υποδεικνύουν το κύριο σημείο (οπτικό κέντρο) της εικόνας. Καθορίζουν τη μετατόπιση του κύριου σημείου από το κέντρο του αισθητήρα στον χώρο του επιπέδου εικόνας ( pixel world ).

Το σημείο  $P_{C_1}$  είναι ένα σημείο στο επίπεδο της κάμερας, άρα πρέπει να κανονικοποιηθεί στην διάσταση του βάθους ώστε να προβληθεί ορθά.

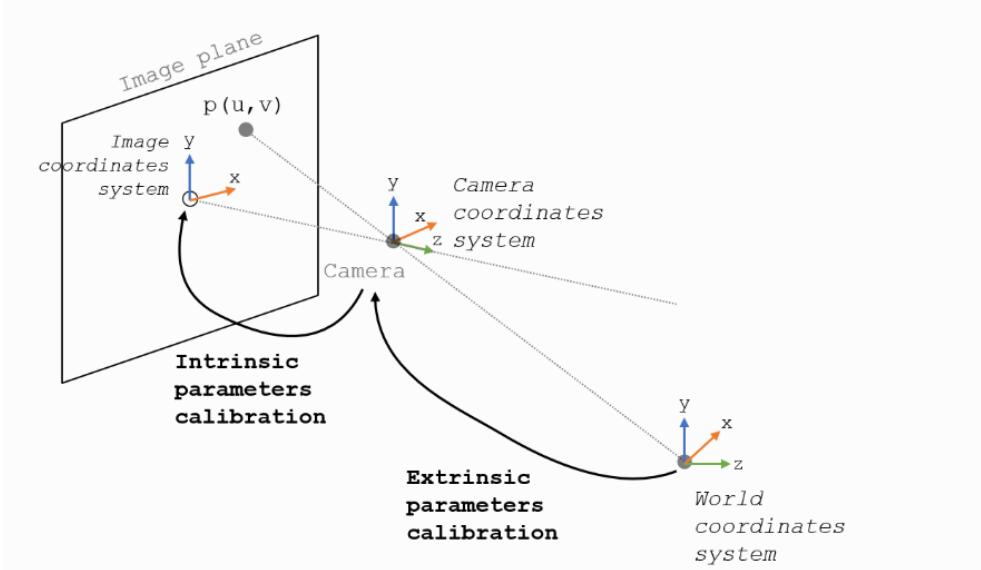
Συνεπώς:

$$\begin{bmatrix} x_p \\ y_p \\ 1 \end{bmatrix} = \begin{bmatrix} X \\ \bar{Z} \\ Y \\ \bar{Z} \\ 1 \end{bmatrix} \quad (2.11)$$

Το διάνυσμα  $u = \begin{bmatrix} x_p \\ y_p \end{bmatrix}$  είναι το σημείο στον χώρο των εικονοστοιχείων της κάμερας το οποίο αντιστοιχεί στο σημείο του χώρου που αντικρύζει.

Η αντίστροφη διαδικασία δεν μπορεί να πραγματοποιηθεί αφού η αίσθηση βάθους δεν υπάρχει κατ' εξοχήν στην εικόνα.

Οι πιο συνηθισμένες μέθοδοι για τον υπολογισμό του πίνακα εσωτερικής βαθμονόμησης περιλαμβάνουν τη χρήση παλινδρόμησης επάνω σε εικόνες με γνωστά αντικείμενα γνωστών διαστάσεων όπως την χρήση ειδικών προτύπων π.χ. μία σκακιέρα [6].



Εικόνα 2-1: Ενδογενής και εξωγενής μετασχηματισμοί<sup>2</sup>

#### 2.1.4 ΔΙΓΡΑΜΜΙΚΗ ΠΑΡΕΜΒΟΛΗ

Η μέθοδος διγραμμικής παρεμβολής υλοποιεί δύο γραμμικές παρεμβολές για να υπολογίσει μία δισδιάστατη εκτίμηση. Έτσι για τα σημεία με τιμές  $(x_1, y_1) = Q_{11}$ ,  $(x_2, y_1) = Q_{21}$ ,  $(x_1, y_2) = Q_{12}$ ,  $(x_2, y_2) = Q_{22}$  σε ένα διακριτό πλέγμα, η εκτίμηση μέσω διγραμμικής παρεμβολής για το σημείο  $(x, y)$  όπου  $x_1 \leq x \leq x_2$ ,  $y_1 \leq y \leq y_2$  είναι:

$$f(x, y) = \frac{(x_2 - x)(y_2 - y)}{(x_2 - x_1)(y_2 - y_1)} Q_{11} + \frac{(x - x_1)(y_2 - y)}{(x_2 - x_1)(y_2 - y_1)} Q_{21} \\ + \frac{(x_2 - x)(y - y_1)}{(x_2 - x_1)(y_2 - y_1)} Q_{12} + \frac{(x - x_1)(y - y_1)}{(x_2 - x_1)(y_2 - y_1)} Q_{22} \quad (2.12)$$

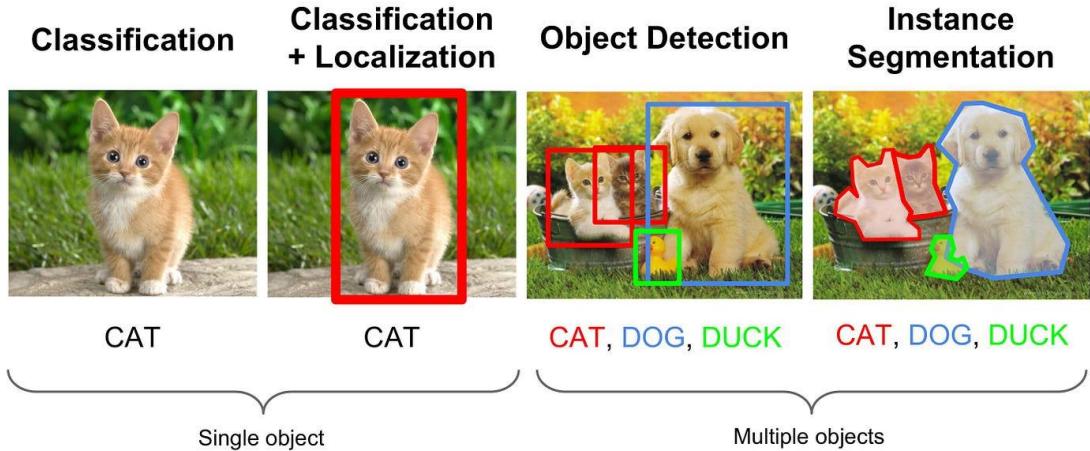
Το οποίο είναι πρακτικά είναι ένα βεβαρημένο μέσο μεταξύ των τεσσάρων σημείων και του σημείου ενδιαφέροντος.

#### 2.1.5 ΤΑΞΙΝΟΜΗΣΗ, ΕΝΤΟΠΙΣΜΟΣ ΚΑΙ ΚΑΤΑΤΜΗΣΗ

Το πρόβλημα της ταξινόμησης (classification) απαιτεί την ετικετοποίηση (labeling) διαφόρων οντοτήτων (objects) σε κλάσεις (classes) με βάση τα χαρακτηριστικά τους. Ο εντοπισμός ή ανίχνευση (detection) απαιτεί την ταυτόχρονη τοποθέτηση ενός πλαισίου το οποίο εμπεριέχει την κάθε οντότητα, οι οποίες μπορεί να είναι πολλές σε κάθε εικόνα σε αντίθεση με την τοποθέτηση (localization) η οποία λειτουργεί για ένα αντικείμενο ανά εικόνα. Η

<sup>2</sup> [https://mpphy0026.readthedocs.io/en/latest/calibration/camera\\_calibration.html](https://mpphy0026.readthedocs.io/en/latest/calibration/camera_calibration.html)

τμηματοποίηση ή κατάτμηση χωρίζει την εικόνα σε τμήματα (συστάδες, ομάδες), ξεχωρίζοντας τις οντότητες σε επίπεδο εικονοστοιχείου (pixel-wise) [7].



Εικόνα 2-2: Ταξινόμηση, Εντοπισμός και Κατάτμηση

Υπάρχουν δύο βασικά είδη κατάτμησης, η σημασιολογική τμηματοποίηση (semantic segmentation) και η κατάτμηση αντικειμένων (instance segmentation). Η πρώτη ομαδοποιεί τα εικονοστοιχεία με βάση τις κλάσεις ως σύνολα, ενώ η δεύτερη ξεχωρίζει κάθε οντότητα κάθε κλάσης μεμονωμένα [7].

### 2.1.6 ΣΥΝΔΙΑΚΥΜΑΝΣΗ

Η συνδιακύμανση ή συνδιασπορά (covariance) έχει να κάνει με την δεσμευμένη μεταβλητότητα (jointed variability) μεταξύ τυχαίων μεταβλητών. Η συνδιασπορά ορίζεται ως η αναμενόμενη τιμή του γινομένου των αποκλίσεων από τις εκάστοτε αναμενόμενες τιμές των τυχαίων μεταβλητών. Δηλαδή, είναι η συσχέτιση μεταξύ τυχαίων μεταβλητών. Η μονάδα μέτρησης είναι μονάδες μέτρησης του  $X$  επί μονάδες μέτρησης του  $Y$ . Τα διανύσματα βάσης της συνδιακύμανσης λέγονται κύριες συνιστώσες (principal components) [8].

Για δύο τυχαίες μεταβλητές  $X, Y$  με αναμενόμενες τιμές  $E[X], E[Y]$  και μετρήσιμες δεύτερες ροπές ο πίνακας συνδιακύμανσης είναι:

$$\text{cov}(X, Y) = E[(X - E[X])(Y - E[Y])] \quad (2.13)$$

Για  $N$  τυχαίες μεταβλητές γράφουμε:

$$\Sigma = \begin{bmatrix} \text{cov}(X_1, X_1) & \text{cov}(X_2, X_1) & \dots & \text{cov}(X_n, X_1) \\ \text{cov}(X_1, X_2) & \text{cov}(X_2, X_2) & \dots & \text{cov}(X_n, X_2) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}(X_1, X_n) & \text{cov}(X_2, X_n) & \dots & \text{cov}(X_n, X_n) \end{bmatrix} \quad (2.14)$$

Αντίστοιχα σε μορφή διανυσμάτων  $\mathbf{X}$  και  $\mathbf{Y}$ :

$$\text{cov}(\mathbf{X}, \mathbf{Y}) = K_{\mathbf{XY}} = E[(\mathbf{X} - E[\mathbf{X}])(\mathbf{Y} - E[\mathbf{Y}])^T] \quad (2.15)$$

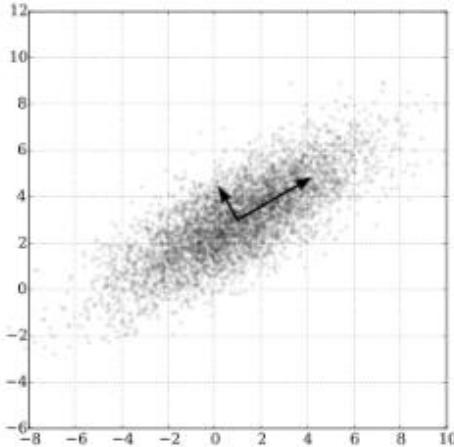
Η διακύμανση μίας τυχαίας μεταβλητής  $X$  είναι:

$$\text{Var}(X) = \text{Cov}(X, X) = E[(X - \mu)^2] \quad (2.16)$$

Ο πίνακας συνδιακύμανσης  $\Sigma$  είναι τετραγωνικός και συμμετρικός.

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \dots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} & \dots & \sigma_{2n} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} & \dots & \sigma_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \sigma_{n3} & \dots & \sigma_{nn} \end{bmatrix} \quad (2.17)$$

Η συνδιακύμανση είναι μία υπερέλλειψη με οποιονδήποτε προσανατολισμό στον χώρο  $n$ , όπως φαίνεται στην Εικόνα 2-3. Οι τυχαίες μεταβλητές έχουν διαφορετικές διακυμάνσεις και είναι μεταξύ τους εξαρτημένες, οπού θετική συνδιακύμανση υπονοεί ότι τείνουν να έχουν την ίδια συμπεριφορά, ενώ αρνητική το αντίθετο. Εάν υπάρχει, ο αντίστροφος του πίνακα συνδιακύμανσης ορίζεται ως πίνακας ακρίβειας ή συγκέντρωσης [9].



Εικόνα 2-3: Κύριοι άξονες δισδιάστατης συνδιακύμανσης

Για μία Γκαουσιανή τυχαία μεταβλητή  $x \sim \mathcal{N}(\mu, \mathbf{Q})$  με μέση τιμή  $\mu$  και συνδιακύμανση  $\mathbf{Q}$  η πυκνότητα πιθανότητας της είναι:

$$p(x) = \mathcal{N}(x; \mu, \mathbf{Q}) = \frac{1}{\sqrt{|2\pi\mathbf{Q}|}} \exp\left(-\frac{1}{2}(x - \mu)^T \mathbf{Q}^{-1}(x - \mu)\right) \quad (2.18)$$

Εξαρτάται μόνο από την μέση τιμή και την συνδιακύμανση. Όταν το διάνυσμα μέσης τιμής είναι μηδέν τότε η μέτρηση ακολουθείται από Γκαουσιανό θόρυβο. Ο Γκαουσιανός θόρυβος

με μηδενική μέση τιμή, σημαίνει ότι δεν υπάρχει συστηματική απόκλιση στις μετρήσεις, αλλά υπάρχει διασπορά γύρω από την πραγματική τιμή [9].

### 2.1.7 ΨΕΥΔΟ-ΑΠΟΣΤΑΣΗ MAHALANOBIS

Η ψευδό-απόσταση Mahalanobis υπολογίζει την απόσταση μεταξύ ενός σημείου και μιας κατανομής πιθανότητας. Αποτελεί μία γενικοποίηση της τετραγωνικής απόκλισης. Είναι 0 όταν σημείο βρίσκεται στο κέντρο της κατανομής και αυξάνεται καθώς το σημείο απομακρύνεται παράλληλα των κύριων συνιστωσών [8].

Για ένα σημείο  $\mathbf{x} = (x_1, x_2, x_3, \dots, x_N)^\top$  και μία κατανομή  $\mathbf{Q}$  στο  $\mathbb{R}^n$  με μέσο  $\mu = (\mu_1, \mu_2, \mu_3, \dots, \mu_N)^\top$  η απόσταση Mahalanobis είναι:

$$d_M(\mathbf{x}, Q) = \sqrt{(\mathbf{x} - \mu)^\top \mathbf{Q}^{-1} (\mathbf{x} - \mu)} \quad (2.19)$$

$\mathbf{Q}^{-1}$  είναι ο πίνακας ακρίβειας ή συγκέντρωσης.

Παράγωγα, η απόσταση μεταξύ δύο σημείων με βάση την συνδιακύμανση των δεδομένων που αντικατοπτρίζεται με τον πίνακα  $\mathbf{S}$  είναι:

$$d_M(\mathbf{x}, \mathbf{y}, Q) = \sqrt{(\mathbf{x} - \mathbf{y})^\top \mathbf{S}^{-1} (\mathbf{x} - \mathbf{y})} \quad (2.20)$$

Η απόσταση είναι αδιάστατη, μη μεταβαλλόμενη της κλίμακας, και λαμβάνει υπόψιν την συσχέτιση μεταξύ των οντοτήτων του χώρου. Εάν ο πίνακας συνδιακύμανσης (διανύσματα βάσης του χώρου) μετασχηματιστεί με τρόπο ώστε  $\mathbf{S} = \mathbf{I}$ , τότε η απόσταση Mahalanobis αντιστοιχεί στην ευκλείδεια [8].

### 2.1.8 ΠΑΡΑΤΗΡΗΤΕΣ ΚΑΤΑΣΤΑΣΗΣ

#### 2.1.8.1 ΘΟΡΥΒΟΣ ΠΑΡΑΤΗΡΗΣΗΣ

Η αβεβαιότητα παρατήρησης (measurement) μίας μέτρησης αποτελείται από την μέση τιμή της μέτρησης του διανύσματος κατάστασης ακολουθούμενη από την συνδιακύμανση της. Έχει να κάνει με το γεγονός ότι οι μετρήσεις των αισθητήρες και το περιβάλλον δεν είναι ιδανικά [10].

#### 2.1.8.2 ΘΟΡΥΒΟΣ ΔΙΑΔΙΚΑΣΙΑΣ

Η αβεβαιότητα διαδικασίας (process) ή πρόβλεψης αντιπροσωπεύει την αβεβαιότητα στη δυναμική του συστήματος, έτσι η μελλοντική πρόβλεψης κίνησης κάθε αντικειμένου δεν είναι

ντετερμινιστική [10].

#### 2.1.8.3 ΘΟΡΥΒΟΣ ΚΑΤΑΣΤΑΣΗΣ

Ο θόρυβος κατάστασης εξηγεί την αβεβαιότητα της κατάστασης του κάθε αντικειμένου στο σύστημα. Η αβεβαιότητα συνήθως μοντελοποίεται ως Γκαουσιανός θόρυβος εξαιτίας των ιδιοτήτων της κανονικής κατανομής, που είναι καλά μελετημένες και επιτρέπουν κάποια απλοποίηση στους υπολογισμούς [10].

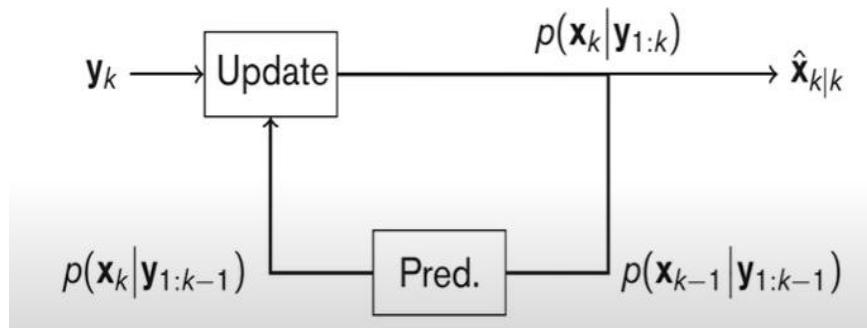
#### 2.1.8.4 ΠΑΡΑΚΟΛΟΥΘΗΣΗ

Για την παρακολούθηση ενός συστήματος κατά την πάροδο του χρόνου χρειάζεται το διάνυσμα κατάστασης του να είναι παρατηρήσιμο (αντίστοιχα ελέγχιμο, αν απαιτείται ο έλεγχος του). Αυτό το διάνυσμα εύλογο είναι να αποτελείται από τις ελάχιστες τιμές που γίνεται για την παραγωγή όλων των τιμών που χρειάζονται. Σε ένα σύστημα παρακολούθησης πολλαπλών στόχων υπάρχουν πολλά διανύσματα κατάστασης εκτός ελέγχου του ίδιου του συστήματος.

Μέσω των αισθητήρων δεν μπορεί να εξαχθεί κάποια απόλυτη ντετερμινιστική κατάσταση, αλλά ένα σύνολο πιθανών καταστάσεων με κάποιες πιο πιθανές από άλλες (Γκαουσιανή διακύμανση γύρω από το πιθανότερο διάνυσμα κατάστασης). Παράλληλα οι αισθητήρες έχουν κάποιο ρυθμό δειγματοληψίας, δηλαδή για κάποιο χρόνο το σύστημα έχει κλειστά τα μάτια και δεν μπορεί να γνωρίζει τις καταστάσεις των αντικειμένων, παρέχοντας μία ασυνέχεια. Αυτή η ασυνέχεια γίνεται πρόβλημα όταν αντικείμενα κινούνται γρήγορα ή σε περιβάλλοντα που υπάρχουν αντικείμενα πολύ κοντά το ένα στο άλλο, καθώς και όταν αντικείμενα φεύγουν και μπαίνουν στην εμβέλεια κάλυψης των αισθητήρων. Μία διαχείριση αυτής της ασυνέχειας λοιπόν είναι η μοντελοποίηση του κινηματικού συστήματος καθενός από τα αντικείμενα ενδιάμεσα του χρόνου δειγματοληψίας. Ακόμα και έτσι όμως το σύστημα παραμένει χαοτικό (ξαφνικές αλλαγές στον προσανατολισμό και την ταχύτητα) για αυτό και προστίθεται στην μοντελοποίηση ο θόρυβος κινηματικής του μοντέλου, ο οποίος αποτυπώνει την αβεβαιότητα της κίνησης. Απόρροια όλων αυτών είναι το διάνυσμα κατάστασης να ακολουθείται από τον θόρυβο κατάστασης ο οποίος αποτελεί μία μορφή συνδυασμού της συνδιακύμανσης της αβεβαιότητας παρατήρησης και της αβεβαιότητας κίνησης σε ένα παράθυρο χρόνου. Η αβεβαιότητα κατάστασης αυξάνεται όταν το σύστημα έχει κλειστά τα μάτια και μειώνεται όταν τελικά ανοίξει τα μάτια.

### 2.1.8.5 ΦΙΛΤΡΟ KALMAN

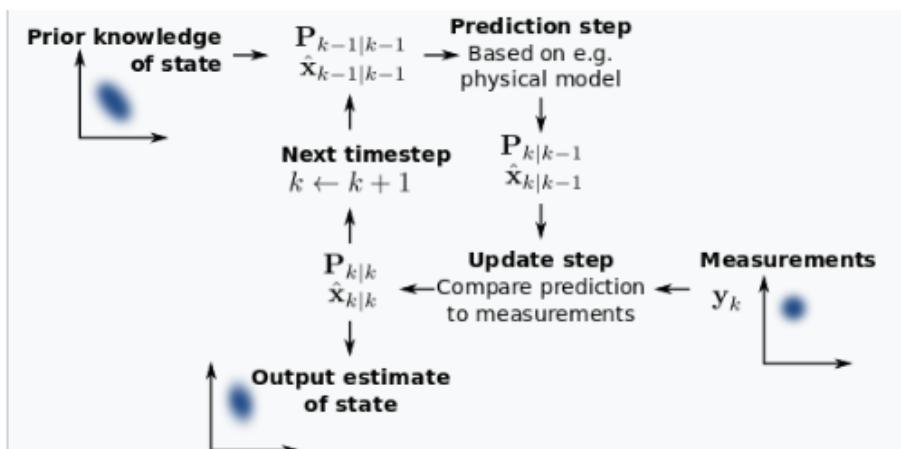
Το φίλτρο Kalman είναι ένας παρατηρητής κατάσταση (state observer) και αποτελεί αναλυτική επίλυση των εξισώσεων φιλτραρίσματος. Είναι επαναληπτικός αλγόριθμος που εκτιμά την πιο πιθανή κατάσταση ενός γραμμικού δυναμικού συστήματος βασισμένο σε μια σειρά παρατηρήσεων υπό γκαουσιανό θόρυβο, είναι ιδανικό σε εφαρμογές πραγματικού χρόνου.



Εικόνα 2-4: Φίλτρο Kalman

Τα στάδια του είναι:

- Βήμα Πρόβλεψης (Prediction Step): Μέσω του κινηματικού μοντέλου του συστήματος και των εισόδων ελέγχου (εάν υπάρχουν) γίνεται πρόβλεψη της επόμενης κατάστασης.
- Βήμα Διόρθωσης (Correction Step): Χρησιμοποιεί τις νέες μετρήσεις που λαμβάνονται από το σύστημα για να προσαρμόσει την προβλεπόμενη κατάσταση. Αυτό επιτυγχάνεται με τη χρήση του κέρδους (gain) Kalman, το οποίο είναι ένα σύνολο βαρών που προσδιορίζει πόσο οι μετρήσεις επηρεάζουν την τελική εκτίμηση. Υπολογίζεται έτσι ώστε να ελαχιστοποιείται η μέση τετραγωνική απόκλιση της εκτίμησης, ταξινομώντας έτσι το Kalman ως βέλτιστο φίλτρο.



Εικόνα 2-5: Μεθοδολογία Φίλτρου Kalman

#### 2.1.8.6 EKTETAMENO ΦΙΛΤΡΟ KALMAN

Το εκτεταμένο φίλτρο Kalman (Extended Kalman Filter, EKF) στην ουσία είναι μια γραμμικοποίηση γύρω από το διάνυσμα κατάστασης, για μη γραμμικά συστήματα, έτσι δεν αποτελεί αναλυτική λύση του προβλήματος αλλά μια προσέγγιση αυτής [11].

Για τη μη γραμμική στοχαστική διαφορική εξίσωση:

$$\begin{aligned}\mathbf{x}_k &= f(\mathbf{x}_{k-1}, \mathbf{u}_{k-1}, \mathbf{w}_{k-1}) \\ \mathbf{z}_k &= h(\mathbf{x}_k, \mathbf{v}_k)\end{aligned}\quad (2.21)$$

Όπου  $f$  ένα μη γραμμικό μοντέλο μετάβασης,  $h$  ένα μη γραμμικό μοντέλο παρατήρησης  $\mathbf{w}$ ,  $\mathbf{v}$  θόρυβος διαδικασίας και παρατήρησης, με συνδιακύμανση  $\mathbf{Q}$  και  $\mathbf{R}$ . Αρχικό θόρυβο κατάστασης με συνδιακύμανση  $P$ . Εάν το Kalman χρησιμοποιείται ως φίλτρο για αντικείμενα εκτός της δυνατότητας ελέγχου τότε δεν υπάρχει είσοδος ελέγχου  $\mathbf{u}$ .

Η συνάρτηση λοιπόν  $f$  χρησιμοποιείται για να προβλέψει την επόμενη κατάστασης του συστήματος με βάση την προηγούμενη του. Η συνάρτηση  $h$  για να προβλέψει την επόμενη μέτρηση του συστήματος με βάση την προβλεπόμενη κατάσταση.

Ωστόσο λόγω μη γραμμικότητας του συστήματος νοείται να υπολογιστούν για κάθε επανάληψη οι Ιακωβιανές των συναρτήσεων  $f$  και  $h$  ως προς  $x$  και θόρυβο (μια μορφή αναπτύγματος Taylor δηλαδή). Για αυτό και λέμε ότι το εκτεταμένο φίλτρο Kalman γραμμικοποιεί το σύστημα γύρω από την τρέχουσα κατάσταση.

Ο Ιακωβιανός πίνακας ορίζεται ως:

$$J = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \dots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \dots & \frac{\partial f_m}{\partial x_n} \end{bmatrix} \quad (2.22)$$

Και ο ολοκληρωμένος αλγόριθμος για το EKF είναι:

Αν

$$F^{(x)} = \frac{\partial f}{\partial x}, \quad F^{(w)} = \frac{\partial f}{\partial w} \quad (2.23)$$

Ιακωβιανές μήτρες της συνάρτησης  $f$  ως προς την κατάσταση και τον θόρυβο διαδικασίας.

Και

$$H^{(x)} = \frac{\partial h}{\partial x}, \quad H^{(v)} = \frac{\partial h}{\partial v} \quad (2.24)$$

Ιακωβιανές μήτρες της συνάρτησης  $h$  ως προς την κατάσταση και τον θόρυβο μέτρησης.

Βήμα πρόβλεψης:

$$\begin{aligned} x_{k+1|k} &= f(x_{k|k}, u^k) \\ P_{k+1|k} &= F_k^{(x)} P_{k|k} F_k^{(x)T} + F_k^{(w)} Q_k F_k^{(w)T} \\ z_{k+1|k} &= h(x_{k+1|k}) \end{aligned} \quad (2.25)$$

Η τρέχουσα εκτίμηση της κατάστασης και της συνδιακύμανσης της χρησιμοποιούνται για την πρόβλεψη της μελλοντικής κατάστασης και της αντίστοιχης συνδιακύμανσης. Η συνδιακύμανση αυξάνεται λόγω αβεβαιότητας.

Βήμα διόρθωσης:

$$\begin{aligned} y_{k+1} &= (z_{k+1} - z_{k+1|k}) \\ S_{k+1} &= H_{k+1}^{(x)} P_{k+1|k} H_{k+1}^{(x)T} + H_{k+1}^{(v)} R_{k+1} H_{k+1}^{(v)T} \\ K_{k+1} &= P_{k+1} H_{k+1}^{(x)T} S_{k+1}^{-1} \\ x_{k+1|k+1} &= x_{k+1|k} + K_{k+1} y_{k+1} \\ P_{k+1|k+1} &= P_{k+1|k} - K_{k+1} S_{k+1} K_{k+1}^T \end{aligned} \quad (2.26)$$

Το βήμα διόρθωσης προσαρμόζει την προβλεπόμενη κατάσταση με βάση την πραγματική μέτρηση. Όπου  $z_{k+1}$  πραγματική παρατήρηση,  $y_{k+1}$  σφάλμα παρατηρητή,  $S_{k+1}$  πίνακας καινοτομίας (innovation),  $K_{k+1}$  κέρδος Kalman. Η τελική αβεβαιότητα κατάστασης (ο πίνακας συνδιακύμανσης της κατάστασης  $P_{k+1|k+1}$ ) μειώνεται.

## 2.2 ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

### 2.2.1 ENTOPIΣMOΣ ANTIKEIMENΩN ΣE 2D ΔΕΔΟΜΕΝΑ KAMERAS

#### 2.2.1.1 H APXITEKTONIKH TOY MASK R-CNN

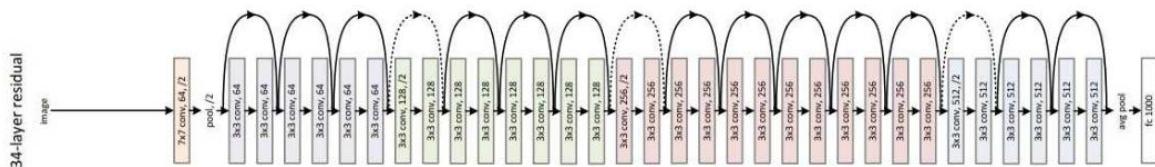
Το Mask R-CNN (Mask Region-based Convolutional Neural Network) [12] προέρχεται από μία πλούσια οικογένεια εντοπιστών (R-CNN, FAST R-CNN, FASTER R-CNN) [13-15]. Δεν εντοπίζει μόνο αντικείμενα αλλά κάνει και κατάτμηση. Τα δομικά στοιχεία του Mask R-CNN αντιμετωπίζονται σχεδόν σε όλες τις αρχιτεκτονικές δισδιάστατων εντοπιστών.

#### 2.2.1.2 PAXOKOKALIA

Η ραχοκοκαλιά (backbone) αναφέρεται στο δίκτυο εξαγωγής χαρακτηριστικών το οποίο μετασχηματίζει την αρχική εικόνα σε έναν χάρτη χαρακτηριστικών (feature map). Η κατασκευή αυτού του χάρτη είναι το πρώτο βήμα των περισσότερων αρχιτεκτονικών εντοπισμού καθώς αποτελεί τον τρόπο με τον οποίο το υπόλοιπο δίκτυο αντιλαμβάνεται την εικόνα κατ' αντιστοιχία με τον άνθρωπο προς την αρχική εικόνα.

Μία σημαντική αρχιτεκτονική ραχοκοκαλιάς για εργασίες εντοπισμού και τμηματοποίησης είναι το ResNet [16] (Residual neural network) το οποίο αποτελείται από πολύπλοκες τεχνικές αναπήδησης ή προσπέρασης συνδέσεων (skip connections) και επαναληπτικών (Residual) μονάδων.

Κάθε backbone έχει καλύτερα αποτελέσματα σε συγκεκριμένες εργασίες, συνήθως χρησιμοποιούνται προ-εκπαίδευμένα και ως βάση για κάποια καινούργια λογική εντοπιστή.



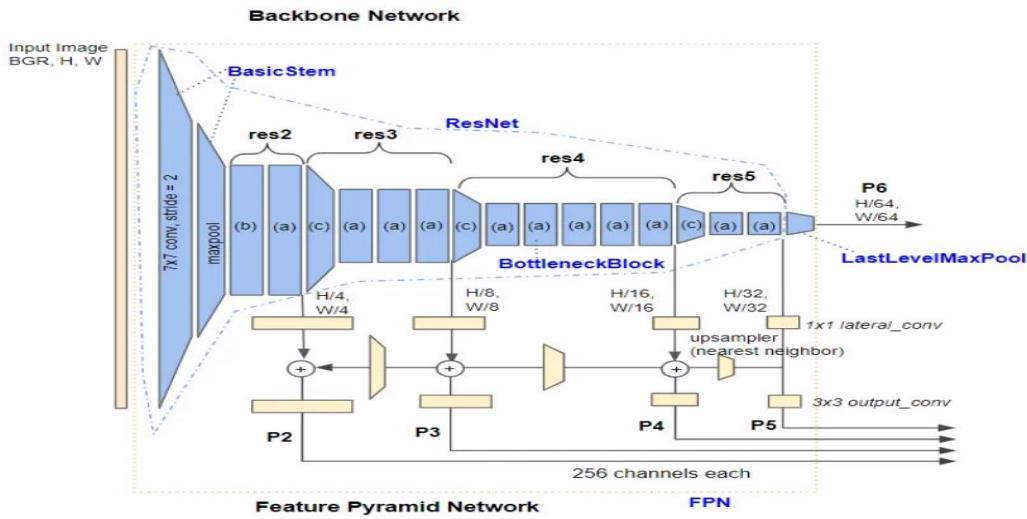
Eikόνα 2-6: Paxokokaliá ResNet

#### 2.2.1.3 ΔΙΚΤΥΑ ΠΥΡΑΜΙΔΑΣ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ

Το δίκτυο πυραμίδας χαρακτηριστικών (Feature Pyramid Network, FPN) κατασκευάζει αφαιρετικά επίπεδα με το καθένα να περιέχει είδη σημασιολογικής πληροφορίας (υψηλά-χαμηλά χαρακτηριστικά μοτίβα) και διαφόρων χωρικών (γεωμετρικών) χαρακτηριστικών

υψηλής και χαμηλής ευκρίνειας [17].

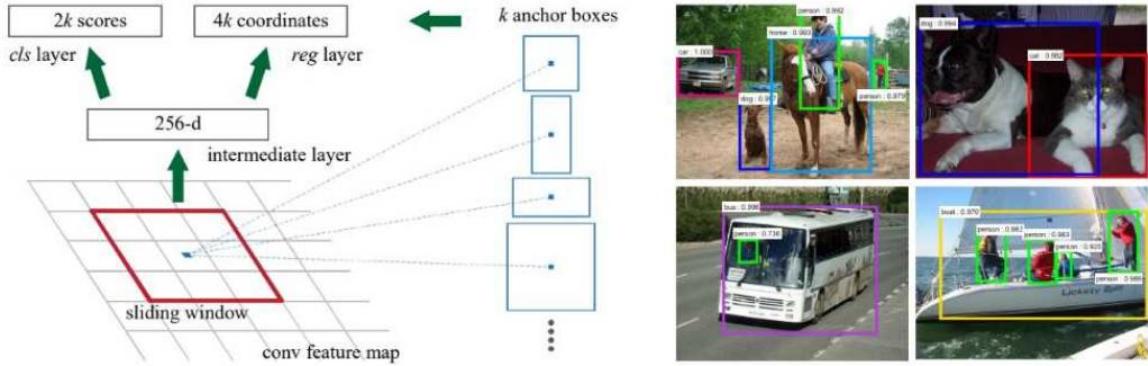
Λειτουργεί συγχέοντας διάφορα επίπεδα της ραχοκοκαλιάς, συμπλέκοντας μοτίβα και χαρακτηριστικά. Το αποτέλεσμα είναι μια πυραμίδα πολλαπλής κλίμακας με κάθε επίπεδο της να αντιστοιχεί σε χαρακτηριστικά διαφορετικής ευκρίνειας. Τα χαμηλότερα επίπεδα αποτυπώνουν χαμηλής ευκρίνειας χαρακτηριστικά, ενώ τα υψηλότερα όλο και πιο λεπτομερές μοτίβα. Έτσι το FPN προσδίδει στο μοντέλο την δυνατότητα να αντιλαμβάνεται ουσιώδης μοτίβα πληροφορίας σε μία συγχωνευμένη μορφή ευκρίνειας και κλίμακας, κάνοντας το αγνωστικιστικό (agnostic) σε σχετικές θέσεις, μεγέθη και προσανατολισμό των αντικειμένων.



Εικόνα 2-7: Δίκτυο Πυραμίδας Χαρακτηριστικών του Mask R-CNN

#### 2.2.1.4 ΔΙΚΤΥΑ ΠΡΟΤΑΣΗΣ ΠΕΡΙΟΧΩΝ

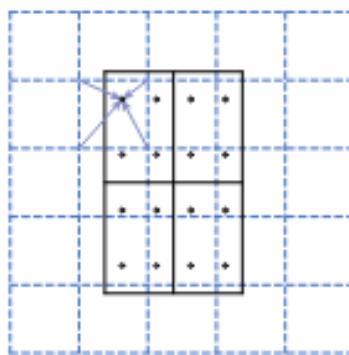
Τα δίκτυα πρότασης περιοχών (Region Proposal Networks, RPNs) εντοπίζουν σημεία ενδιαφέροντος σε όλη την έκταση της πυραμίδας χαρακτηριστικών μέσω των επιπέδων συγκέντρωσης περιοχών ενδιαφέροντος (Regions of interest pooling, ROI Pooling). Πρακτικά χρησιμοποιούνται με σκοπό το υπόλοιπο του δικτύου να μην χρειαστεί να εξερευνήσει όλον τον χάρτη αλλά κάποια σημεία που δείχνουν ενδιαφέρον. Αυτά τα σημεία ενδιαφέροντος είναι τα κουτιά αγκύρωσης (anchor boxes) μέσα στα οποία πολύ πιθανόν να υπάρχει κάποιο αντικείμενο ενδιαφέροντος (blobs). Για κάθε κουτί, το RPN υπολογίζει μια βαθμολογία αντικειμένου η οποία εκφράζει την πιθανότητα να περιέχει όντως αντικείμενο. Παράλληλα, υπολογίζεται μια βαθμολογία περιοχής η οποία υποδεικνύει πόσο καλά ταιριάζει το κουτί με το πιθανό σχήμα του αντικειμένου. Με βάση τις βαθμολογίες αυτές, επιλέγεται ένα σύνολο προτάσεων περιοχής οι οποίες αποτελούν τα υποψήφια αντικείμενα εντοπισμού. Για το Mask R-CNN οι προτάσεις είναι 1000 [12].



Εικόνα 2-8: Δίκτυο Πρότασης Περιοχών του Mask R-CNN

### 2.2.1.5 ΕΥΘΥΓΡΑΜΜΙΣΗ ΠΕΡΙΟΧΗΣ ΕΝΔΙΑΦΕΡΟΝΤΟΣ

Το δίκτυο ευθυγράμμισης περιοχών ενδιαφέροντος (Regions of interest Alignment, ROI Align) έχει σκοπό να ευθυγραμμίσει τις προτάσεις περιοχής με τα χαρακτηριστικά της πυραμίδας FPN και να επιστρέψει τις τελικές απαντήσεις των σημείων αυτών σε έναν ενιαίο χάρτη, εξασφαλίζοντας ότι τα χαρακτηριστικά που χρησιμοποιούνται για την ανάλυση κάθε αντικειμένου προέρχονται από την ακριβή του θέση στην αρχική εικόνα. Μία ειδοποιός διαφορά του Mask R-CNN με το FASTER R-CNN [12] είναι ότι το Mask R-CNN χρησιμοποιεί τη μέθοδο της διγραμμικής παρεμβολής (bilinear interpolation) για να μπορέσει να ευθυγραμμίσει με ακόμα μεγαλύτερη ακρίβεια τα χαρακτηριστικά του κάθε πλαισίου ενδιαφέροντος, το οποίο μπορεί να έχει μετατεθεί λόγω του εξαγωγέα περιοχών ενδιαφέροντος (ROI Pooling). Ο μέγιστος αριθμός αυτών των τελικών κουντιών είναι 100.

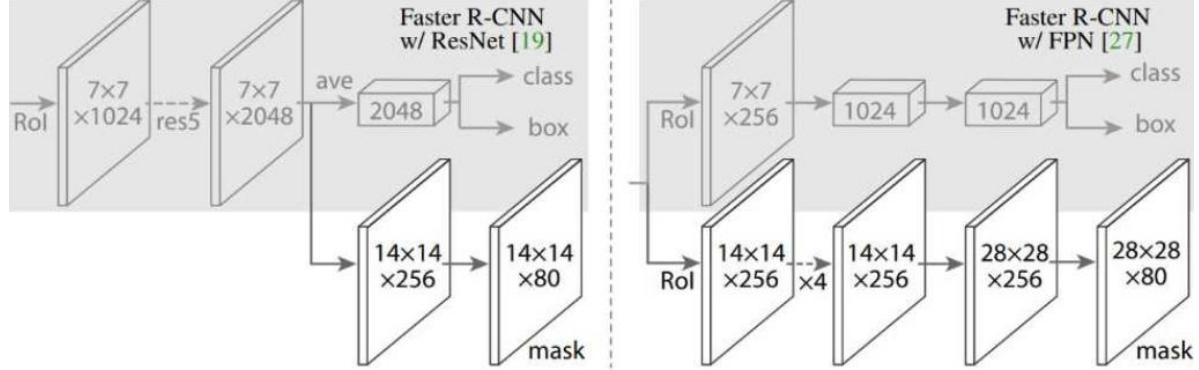


Εικόνα 2-9: Διγραμμική Παρεμβολή

### 2.2.1.6 ΚΕΦΑΛΗ MASK R-CNN

Στο τελευταίο στάδιο ταξινομούνται τα αντικείμενα ενδιαφέροντος, υπολογίζονται τα πλαίσια τους και δημιουργούνται μάσκες τμηματοποίησης για το καθένα. Το ένα μέλος ταξινομεί μέσω ενός πολυστρωματικού δικτύου (MLP) και εξάγει τα αντίστοιχα πλαίσια κάθε περιοχής

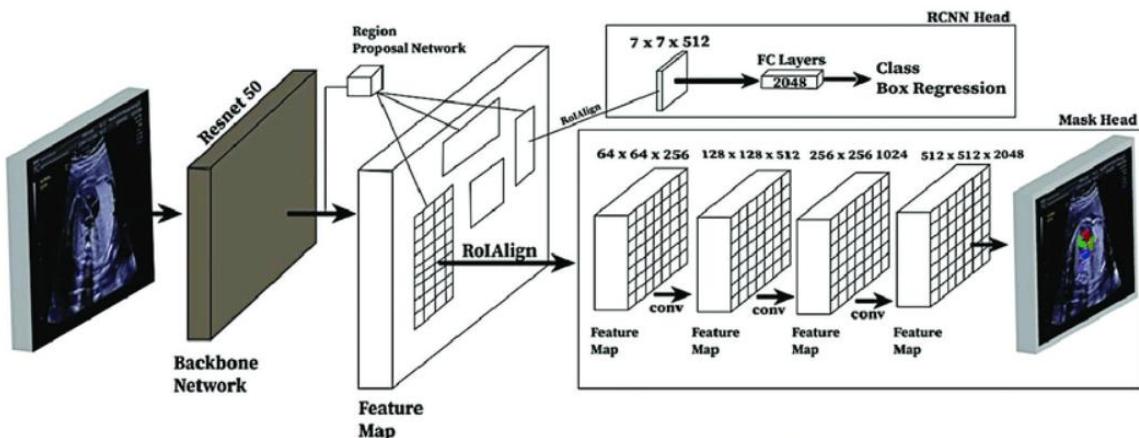
ενδιαφέροντος μέσω παλινδρόμησης. Το άλλο μέλος είναι αφιερωμένο στην κατασκευή μασκών σε επίπεδο εικονοστοιχείων για τμηματοποίηση. Χρησιμοποιεί ένα πλήρως συνδεδεμένο συνελικτικό δίκτυο (Full Connected convolutional Network, FCN) για κάθε σημείο ενδιαφέροντος κατασκευάζοντας μια δυαδική μάσκα μεταξύ του αντικειμένου και του παρασκηνίου (background) [12].



Εικόνα 2-10: Κεφαλή Mask R-CNN

#### 2.2.1.7 ΕΜΦΑΝΙΣΙΑΚΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ

Οι περιοχές (regions) που αντιστοιχούν σε σημεία ενδιαφέροντος στον χάρτη χαρακτηριστικών (feature map) είναι πλούσια σε μοτίβα της ιδιομορφίας της περιοχής εκείνης, δηλαδή σε εμφανισιακά χαρακτηριστικά του ίδιου του αντικειμένου.



Εικόνα 2-11: Αρχιτεκτονική Mask R-CNN

#### 2.2.2 ΕΝΤΟΠΙΣΜΟΣ ΑΝΤΙΚΕΙΜΕΝΩΝ ΣΕ 3D ΔΕΔΟΜΕΝΑ LIDAR

##### 2.2.2.1 Ο ΤΡΙΣΔΙΑΣΤΑΤΟΣ ΕΝΤΟΠΙΣΜΟΣ

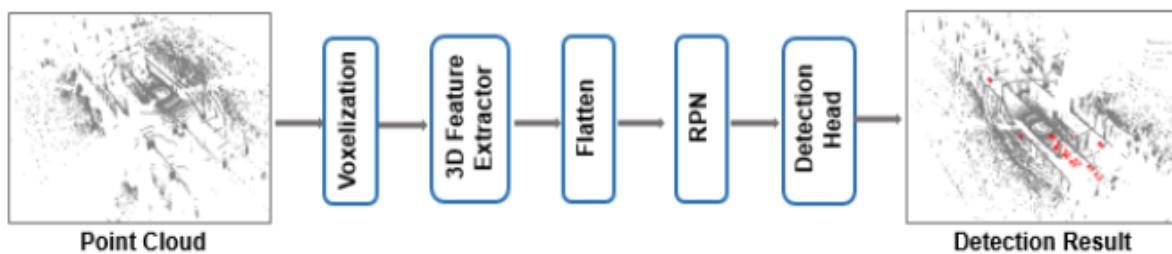
Η έξοδος ενός τρισδιάστατου ανιχνευτή είναι ένα διάνυσμα κατάστασης για κάθε αντικείμενο. Για κάθε αντικείμενο το διάνυσμα αυτό αποτελείται από το κέντρο του στον χώρο, τον

προσανατολισμός του (σχετικιστικά του πλαισίου αναφοράς) και το μέγεθος του καθώς και η πιθανότερη του κλάση. Οι τρισδιάστατοι εντοπιστές λειτουργούν με είσοδο το νέφος σημείων.

Μπορούμε να γράψουμε:

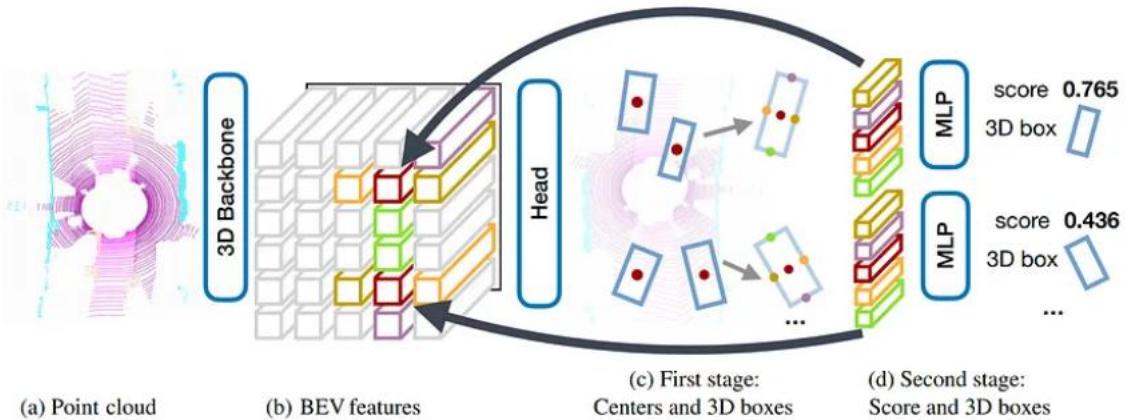
$$(\hat{\mathbf{C}}, \mathbf{V}) = G(\mathbf{PCD}) \quad (2.27)$$

Όπου **PCD** το νέφος σημείων,  $\hat{\mathbf{C}}$  δείκτης κλάσης,  $\mathbf{V}$  το διάνυσμα κατάστασης και  $G$  ο τρισδιάστατος εντοπιστής.



*Εικόνα 2-12: Τρισδιάστατος Εντοπισμός*

Το δίκτυο επεξεργάζεται το ακατέργαστο νέφος μέσω βιοξελοποίησης (voxelization) και προβάλει το νέφος σε χαρακτηριστικά μοτίβα (characteristic patterns). Η έξοδος εισάγεται στον εξαγωγέα 3d χαρακτηριστικών, στην μονάδα επιπεδοποίησης, στο δίκτυο πρότασης περιοχών και τέλος στην 3d κεφαλή εντοπισμού (detection head). Το δίκτυο CenterPoint διαφέρει από τους υπόλοιπους 3d ανιχνευτές που χρησιμοποιούν άγκυρες πλαισίων, αναπαριστώντας τα αντικείμενα ως σημεία (κέντρα). Η δομή του αποτελείται από δύο κύρια στάδια. Το πρώτο στάδιο περιλαμβάνει την εξαγωγή χαρακτηριστικών και την πρόβλεψη και το δεύτερο στην ταξινόμηση και βελτιστοποίηση των αποτελεσμάτων. Ένα ακόμα σημαντικό χαρακτηριστικό του CenterPoint είναι η ενσωμάτωση ενός δικτύου χαρακτηριστικών πυραμίδας (FPN) για την αποτελεσματική σύντηξη χαρακτηριστικών πολλαπλών κλιμάκων. Αυτή η δομή, όπως και στους 2d ανιχνευτές, συνδυάζει τα χαρακτηριστικά διαφορετικών χωρικών αναλύσεων.



Εικόνα 2-13: Αρχιτεκτονική Τρισδιάστατου Ανιχνευτή CenterPoint

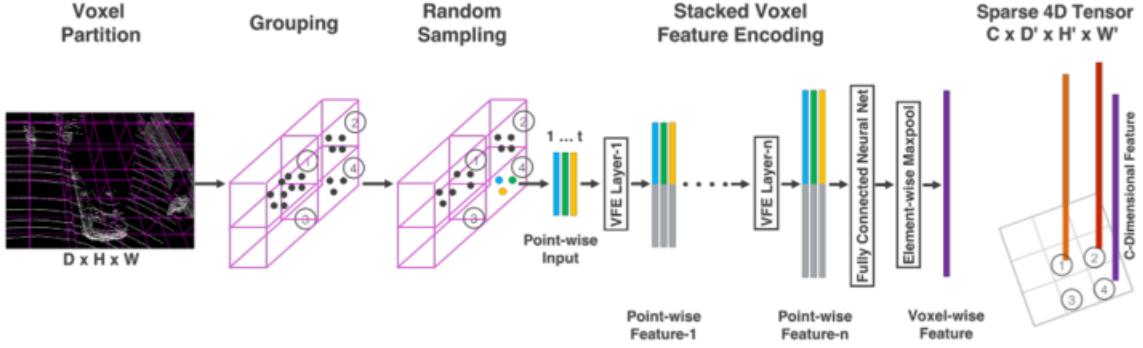
### 2.2.2.2 ΒΟΞΕΛΟΠΟΙΗΣΗ

Η διαδικασία της βοξελοποίησης αρχικά περιλαμβάνει την αναπαράσταση του νέφους σημείων σε τρισδιάστατα εικονοστοιχεία ή κυψέλες (voxels), χωρίζοντας έτσι τον χώρο σε ένα πλέγμα [18]. Κάθε κυψέλη αντιπροσωπεύει ένα διακριτό κομμάτι του χώρου, περιέχοντας πληροφορίες για τα σημεία εντός αυτού. Όσο πιο κοντά στον αισθητήρα βρίσκεται μία κυψέλη τόσο πιο πυκνή, αντίστοιχα, όσο πιο μακριά βρίσκεται μία κυψέλη τόσο λιγότερα σημεία εμπεριέχει (αραιά αποτύπωση). Στα αραιά τμήματα του νέφους, η μέθοδος υπερδειγματοληψίας (oversampling) δημιουργεί πολλαπλά αντίγραφα των υπαρχόντων σημείων αντισταθμίζοντας έτσι την ανισοκατανομή και εξασφαλίζοντας ότι καμία κυψέλη δεν παραλείπεται. Παράλληλα, αντιμετωπίζει την ανισορροπία των τάξεων, καθώς επαναλαμβάνει δείγματα των υπό-εκπροσωπούμενων τάξεων για να αυξήσει τον αριθμό τους, έτσι ώστε όλες οι τάξεις να έχουν περίπου ίσο αριθμό δειγμάτων κατά την εκπαίδευση.

Η μεθοδολογία VoxelNet κωδικοποιεί τα χαρακτηριστικά των σημείων μέσα σε κάθε κυψέλη, δημιουργώντας έναν τετραδιάστατο αραιό τανυστή (4D tensor). Αυτός ο τανυστής περιγράφει τις ιδιότητες του νέφους στον μετασχηματισμένο χώρο μειωμένων διαστάσεων ( $D'$ ,  $H'$ ,  $W'$ ), χρησιμοποιώντας ένα αντίστοιχο διάνυσμα χαρακτηριστικών  $C$  για κάθε κυψέλη.

Η διαδικασία ομαδοποίησης των σημείων γίνεται μέσω τυχαίας δειγματοληψίας από τον αρχικό χάρτη διάστασης ( $D$ ,  $H$ ,  $W$ ) η οποία αποτελείται από πολλαπλές επαναλήψεις, έτσι ώστε να διασφαλιστεί κάθε κυψέλη αναπαριστά άρτια τα εσωτερικά της σημεία. Τα χαρακτηριστικά κάθε κυψέλης κωδικοποιούνται στην συμπαγή αναπαράσταση μεγέθους  $C$ . Με αυτόν τον τρόπο το VoxelNet μπορεί να κατανοήσει τη χωροταξική κατανομή των σημείων και να εκμεταλλευτεί τη δομή του τρισδιάστατου χώρου για την αναγνώριση αντικειμένων

μέσω της αραιής αναπαράστασης του νέφους.



Eikόνα 2-14: Τεχνική Βοξελοποίησης

### 2.2.2.3 3D ΡΑΧΟΚΟΚΑΛΙΑ

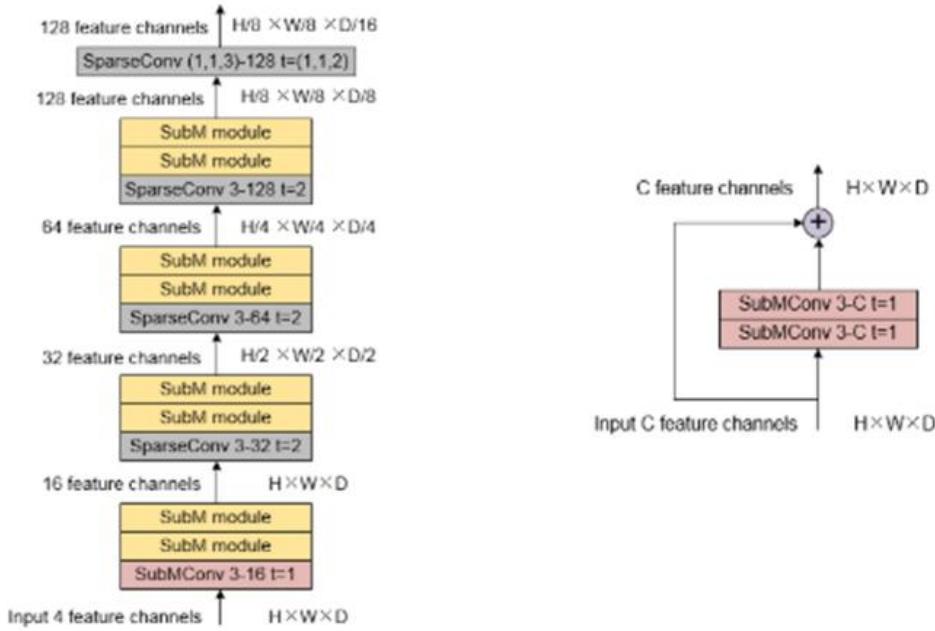
Η 3d ραχοκοκαλιά (3d backbone) είναι υπεύθυνη για την εξαγωγή πλούσιων χαρακτηριστικών από τον τετραδιάστατο αραιό χάρτη μέσω εκμάθησης χαρακτηριστικών όπως η καμπυλότητα των επιφανειών και η τοπολογία των αντικειμένων. Αποτελείται από πολλαπλά επίπεδα συνελίξεων, τα οποία εφαρμόζονται στις κυψέλες για την εξαγωγή χαρακτηριστικών, χρησιμοποιώντας τεχνικές αραιής συνέλιξης (sparse convolutions) και υπό-πολλαπλών αραιών μονάδων (Submanifold Sparse Units).

#### 2.2.2.3.1 ΑΡΑΙΕΣ ΣΥΝΕΛΙΞΕΙΣ

Οι αραιές συνελίξεις (sparse convolutions) είναι μια παραλλαγή των παραδοσιακών συνελίξεων που έχουν σχεδιαστεί για να είναι αποτελεσματικές σε δεδομένα με χαμηλή πυκνότητα, ιδανικές για 3d νέφη σημείων. Σε αντίθεση με τις κανονικές συνελίξεις, οι αραιές συνελίξεις εφαρμόζονται μόνο σε ενεργές κυψέλες (δηλαδή σε κυψέλες που περιέχουν τουλάχιστον ένα σημείο), επιτρέποντας την εξοικονόμηση υπολογιστικών πόρων και μνήμης [19].

#### 2.2.2.3.2 ΥΠΟΠΟΛΛΑΠΛΕΣ ΑΡΑΙΕΣ ΜΟΝΑΔΕΣ

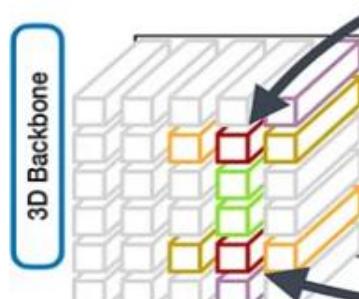
Οι υποπολλαπλές αραιές μονάδες (Submanifold Sparse Units) είναι μια ειδική κατηγορία αραιών συνελίξεων που διατηρούν την αρίθμηση των ενεργών κυψελών κατά τη διάρκεια της επεξεργασίας, κάτι ιδιαίτερα σημαντικό για την αποφυγή της εξάπλωσης της πυκνότητας, που θα μπορούσε να μειώσει την αποδοτικότητα της επεξεργασίας [20].



Εικόνα 2-15: Υποπολλαπλές Αραιές Μονάδες

#### 2.2.2.4 ΧΑΡΤΗΣ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ

Η έξοδος της 3d ραχοκοκαλιάς μία αναπαράσταση του αρχικού νέφους σε μορφή κάτοψης (Bird's Eye View, BEV), δηλαδή η προβολή των τρισδιάστατων δεδομένων στο δισδιάστατο επίπεδο. Πάνω σε αυτόν τον χάρτη χαρακτηριστικών γίνεται η αναγνώριση της χωρικής κατανομής των αντικειμένων και των αποστάσεων τους.



(b) BEV features

Εικόνα 2-16: Χάρτης Χαρακτηριστικών

#### 2.2.2.5 ΚΕΦΑΛΗ ΕΝΤΟΠΙΣΜΟΥ

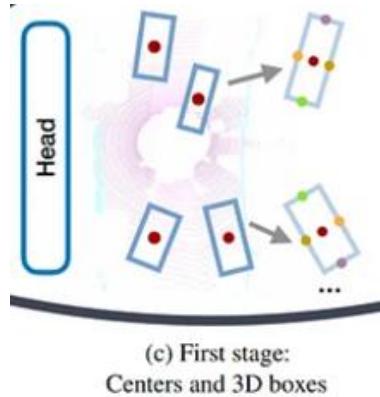
##### 2.2.2.5.1 ΠΡΩΤΟ ΣΤΑΔΙΟ: ΧΑΡΤΕΣ ΘΕΡΜΟΤΗΤΑΣ ΚΑΙ ΚΕΦΑΛΙΑ ΠΑΛΙΝΔΡΟΜΗΣΗΣ

Στο πρώτο στάδιο, η κεντρική κεφαλή χάρτη θερμότητας παράγει χάρτες θερμότητας για κάθε

κλάση αντικειμένου (Κ-κάναλο με Κ κλάσεις ενδιαφέροντος). Αυτοί οι χάρτες αναπαριστούν την πιθανότητα ύπαρξης (κέντρου) αντικειμένου σε κάθε θέση χάρτη χαρακτηριστικών. Η έξοδος του είναι Κ χάρτες θερμότητας Γκαουσιανής κατανομής [21].

Ανάλογα την εφαρμογή διατίθενται πέντε ή έξι κεφαλές παλινδρόμησης, καθεμία εξειδικευμένη στην πρόβλεψη συγκεκριμένων ιδιοτήτων των αντικειμένων. Η κεφαλή κέντρου (2 τιμές, x,y), ύψους (1 τιμή, z), η κεφαλή αντιστάθμισης για διορθώσεις στις προβλέψεις της κεφαλής κέντρου (2 τιμές, dx, dy), η κεφαλή πλαισίου (3 τιμές, l, w, h), και η κεφαλή προσανατολισμού (2 τιμές, cosθ, sinθ). Τέλος, η προαιρετική κεφαλή ταχύτητας, η οποία εκτιμά τις συνιστώσες της ταχύτητας του αντικειμένου.

Η αρχιτεκτονική κάθε μίας ακολουθεί παρόμοιο μοτίβο, χρησιμοποιώντας συνελικτικά επίπεδα, κανονικοποίηση και ReLU. Οι τελικές προβλέψεις γίνονται μέσω συνελίξεων 1x1 που προβάλλουν τα χαρακτηριστικά στον απαιτούμενο αριθμό καναλιών εξόδου.



Εικόνα 2-17: Πρώτο στάδιο Κεφαλής

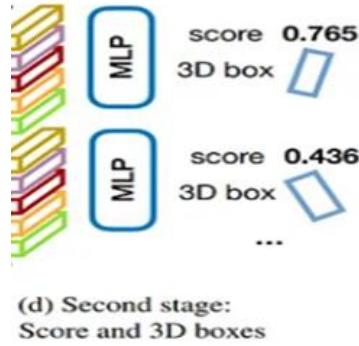
#### 2.2.2.5.2 ΔΕΥΤΕΡΟ ΣΤΑΔΙΟ: ΤΑΞΙΝΟΜΗΣΗ ΚΑΙ ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ

Το δεύτερο στάδιο του CenterPoint επεξεργάζεται περαιτέρω τις προβλέψεις του πρώτου σταδίου. Αρχικά, εντοπίζει τοπικά μέγιστα στους χάρτες θερμότητας και στη συνέχεια εξάγει χαρακτηριστικά για κάθε εντοπισμένο κέντρο. Τέλος, πραγματοποιεί την τελική ταξινόμηση και βελτιστοποίηση των προβλέψεων. Με αυτήν την μοντελοποίηση ο CenterPoint κατασκευάζει πλαίσια χωρίς να βασίζεται σε προκαθορισμένες άγκυρες [21].

Η εκπαίδευση γίνεται με ολιστικό τρόπο, με όλες τις κεφαλές να εκπαιδεύονται ταυτόχρονα. Για την αντιμετώπιση της ανισορροπίας κλάσεων στους χάρτες θερμότητας, χρησιμοποιείται εστιακή απώλεια (focal loss), για τις κεφαλές παλινδρόμησης εφαρμόζονται L1 ή ομαλή (smooth) L1 ως συναρτήσεις σφάλματος.

Κατά τη διάρκεια του συμπερασμού, ο CenterPoint εντοπίζει αρχικά τοπικά μέγιστα στους

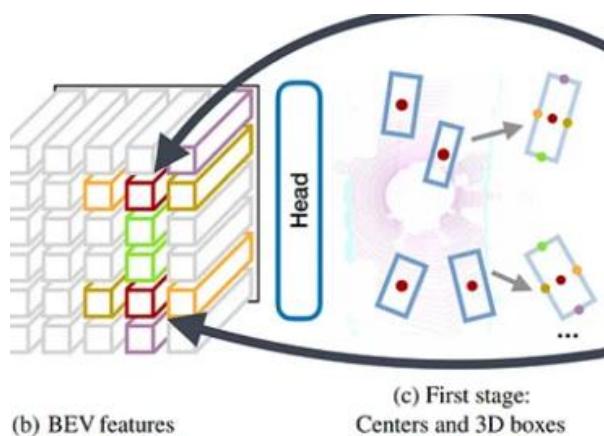
χάρτες θερμότητας και στη συνέχεια εξάγει προβλέψεις από τις κεφαλές παλινδρόμησης για κάθε εντοπισμένο κέντρο. Οι προβλέψεις αυτές συνδυάζονται για τη δημιουργία των τελικών τρισδιάστατων πλαισίων οριοθέτησης. Τέλος, εφαρμόζεται μη μέγιστη καταστολή (non-maximum suppression) για την απομάκρυνση επικαλυπτόμενων προβλέψεων.



Εικόνα 2-18: Δεύτερο στάδιο Κεφαλής

#### 2.2.2.6 ΤΑ MOTIBA ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ

Ο χάρτης χαρακτηριστικών περιέχει πλούσια χαρακτηριστικά αναπαράστασης της περιοχής σάρωσης και επακολούθως των γεωμετρικών χαρακτηριστικών των αντικειμένων. Ο μετασχηματισμός προς τον χάρτη BEV από τα εντοπιζόμενα κέντρα των αντικειμένων στο τελικό σύστημα αναφοράς του LiDAR, είναι εφικτός γνωρίζοντας το μέγεθος της κυψέλης (voxel size), το πεδίο τιμών του νέφους (point cloud range) και του βήματος (stride).



Εικόνα 2-19: Γεωμετρικά Χαρακτηριστικά

#### 2.2.3 ΜΕΘΟΔΟΙ ΜΕΙΩΣΗΣ ΔΙΑΣΤΑΣΙΜΟΤΗΤΑΣ

Η εκμάθηση πολλαπλών συλλογών (manifold learning) είναι ένα πεδίο της μη επιβλεπόμενης μηχανικής μάθησης που εστιάζει στην ανακάλυψη δομών χαμηλών διαστάσεων σε δεδομένα υψηλής διάστασης. Βασίζεται στην υπόθεση ότι πολλά σύνολα δεδομένων βρίσκονται κοντά

σε μια πολλαπλότητα χαμηλής διάστασης ενσωματωμένη στον χώρο υψηλότερης διάστασης. Έτσι, στόχος είναι η μείωση της διαστατικότητας διατηρώντας παράλληλα τη δομή των δεδομένων. Η χαμηλοδιάστατη (low-dimensional) αναπαράσταση πρέπει να διατηρεί ουσιαστικές ιδιότητες των αρχικών δεδομένων, δηλαδή τοπολογικές σχέσεις και δομές.

Οι τεχνικές χωρίζονται σε γραμμικές και μη γραμμικές. Οι γραμμικές, όπως η Ανάλυση Κύριων Συνιστωσών (PCA), υποθέτουν ότι τα δεδομένα βρίσκονται σε ή κοντά σε έναν γραμμικό υπό-χώρο (sub-space). Οι μη γραμμικές μέθοδοι, όπως η ισομετρική απεικόνιση (Isomap) ή η ομοιόμορφη προσέγγιση και προβολή πολλαπλότητων (Uniform Manifold Approximation and Projection, UMAP) [22], μπορούν να ανακαλύψουν πιο περίπλοκες καμπύλες δομές στα δεδομένα αλλά απαιτούν πιο σημαντική παραμετροποίηση.

### 2.2.3.1 ΓΡΑΜΜΙΚΕΣ ΜΕΘΟΔΟΙ

#### 2.2.3.1.1 ΑΝΑΛΥΣΗ ΚΥΡΙΩΝ ΣΥΝΙΣΤΩΣΩΝ

Η ανάλυση κύριων συνιστωσών (Principal Component Analysis - PCA) [23] χρησιμοποιείται για την προβολή δεδομένων σε μία νέα βάση, η οποία είναι γραμμικός συνδυασμός της αρχικής. Σκοπός της είναι να αναπαραστήσει τα δεδομένα με τρόπο που να μεγιστοποιεί την πληροφορία που περιέχουν, μειώνοντας παράλληλα τη μεταξύ τους συσχέτιση.

Για τον πίνακα δεδομένων  $\mathbf{X} \in \mathbb{R}^{m \times n}$  όπου  $m$  δείγματα και  $n$  χαρακτηριστικά εφαρμόζουμε τον μετασχηματισμό βάσης:

$$\mathbf{Y} = \mathbf{P}\mathbf{X} \quad (2.28)$$

Με  $\mathbf{P} \in \mathbb{R}^{m \times m}$  για τον οποίο πρέπει η διακύμανση των δεδομένων να μεγιστοποιηθεί και η συσχέτιση μεταξύ των μεταβλητών του να ελαχιστοποιηθεί.

Ο πίνακας συνδιακύμανσης υπολογίζεται ως:

$$\mathbf{C}_X = \frac{1}{n-1} \mathbf{X}\mathbf{X}^T \quad (2.29)$$

Ομοίως:

$$\mathbf{C}_Y = \frac{1}{n-1} \mathbf{Y}\mathbf{Y}^T = \frac{1}{n-1} (\mathbf{P}\mathbf{X})(\mathbf{P}\mathbf{X})^T = \frac{1}{n-1} (\mathbf{P}\mathbf{X})(\mathbf{X}^T\mathbf{P}^T) = \frac{1}{n-1} \mathbf{P}(\mathbf{X}\mathbf{X}^T)\mathbf{P}^T \quad (2.30)$$

Έτσι:

$$\mathbf{C}_Y = \frac{1}{n-1} \mathbf{P}\mathbf{S}\mathbf{P}^T, \quad \mathbf{S} = \mathbf{X}\mathbf{X}^T \quad (2.31)$$

Χρησιμοποιώντας το θεώρημα ότι κάθε τετραγωνικός συμμετρικός πίνακας είναι ορθογώνια διαγωνοποιήσιμος:

$$\mathbf{S} = \mathbf{E} \mathbf{D} \mathbf{E}^T \quad (2.32)$$

Καταλήγουμε:

$$\mathbf{C}_Y = \frac{1}{n-1} \mathbf{E}^T (\mathbf{E} \mathbf{D} \mathbf{E}^T) \mathbf{E} \quad (2.33)$$

Εφόσον υπολογιστούν οι ιδιοτιμές και τα ιδιοδιανύσματα του  $\mathbf{S}$  ταξινομούνται σε φθίνουσα σειρά και παίρνουν τη θέση του όρου  $\mathbf{E}^T (\mathbf{E} \mathbf{D} \mathbf{E}^T) \mathbf{E}$  εφόσον  $\mathbf{P} = \mathbf{E}^T$ , δηλαδή οι γραμμές του  $\mathbf{P}$  είναι τα ιδιοδιανύσματα του  $\mathbf{S}$ . Πλέον η μεγαλύτερη διακύμανσή αντιστοιχεί στον πιο σημαντικό άξονα, η δεύτερη στον δεύτερο και ούτω καθεξής.

#### 2.2.3.1.2 ΑΠΟΣΥΝΘΕΣΗ ΜΟΝΑΔΙΑΙΩΝ ΤΙΜΩΝ

Η αποσύνθεση μοναδιαίων τιμών (Singular Value Decomposition, SVD) καταλήγει στην ίδια μορφολογία με την ανάλυση κύριων συνιστωσών, μόνο που είναι πιο γρήγορη καθώς δεν απαιτείται ο υπολογισμός πίνακα συνδιακύμανσης, παράλληλα δίνει καλύτερα αποτελέσματα όταν τα δεδομένα είναι αραιά [24].

Για τον  $\mathbf{A} \in \mathbb{R}^{m \times n}$  η αποσύνθεση μοναδιαίων τιμών γράφεται ως:

$$\mathbf{A} = \mathbf{U} \mathbf{D} \mathbf{V}^T \quad (2.34)$$

Όπου:

$\mathbf{U} \in \mathbb{R}^{m \times m}$  : Ορθοκανονικός πίνακας, οι στήλες του είναι τα αριστερά μοναδιαία διανύσματα του  $\mathbf{A} \mathbf{A}^T$ . Εκφράζουν τη δομή των δεδομένων του πίνακα στον χώρο των γραμμών του  $\mathbf{A}$ .

$\mathbf{D} \in \mathbb{R}^{m \times n}$  : Διαγώνιος πίνακας, περιέχει τις μοναδιαίες τιμές του  $\mathbf{A}$ , οι οποίες αντιπροσωπεύουν τη "δύναμη" ή τη "σημασία" κάθε διαστασιακής συνιστώσας. Οι μεγαλύτερες τιμές αντιστοιχούν στις κύριες διαστάσεις.

$\mathbf{V} \in \mathbb{R}^{n \times n}$ : Ορθοκανονικός πίνακας, οι στήλες του είναι τα ιδιοδιανύσματα του  $\mathbf{A}^T \mathbf{A}$ , δηλαδή τα δεξιά μοναδιαία διανύσματα. Εκφράζουν τη δομή των δεδομένων στον χώρο των στηλών.

Ο SVD αναπαριστά την καλύτερη προσέγγιση χαμηλότερης τάξης του αρχικού πίνακα. Αν διατηρηθούν μόνο οι πρώτες  $k$  μοναδιαίες τιμές, η προσεγγιστική μήτρα:

$$A_k = \sum_{i=1}^k \sigma_i u_i v_i^T \quad (2.35)$$

Όπου  $v_1, v_2, \dots, v_r$  τα δεξιά μοναδιαία διανύσματα,  $u_1, u_2, \dots, u_r$  τα αριστερά μοναδιαία διανύσματα και  $\sigma_1, \sigma_2, \dots, \sigma_r$  οι μοναδιαίες τιμές, ελαχιστοποιεί το σφάλμα Frobenius:

$$\| A - A_k \|_F \quad (2.36)$$

### 2.2.3.2 ΜΗ ΓΡΑΜΜΙΚΕΣ ΤΕΧΝΙΚΕΣ

#### 2.2.3.2.1 t-ΚΑΤΑΝΕΜΗΜΕΝΗ ΣΤΟΧΑΣΤΙΚΗ ΕΝΣΩΜΑΤΩΣΗ ΓΕΙΤΟΝΩΝ

Η t-Κατανεμημένη Στοχαστική Ενσωμάτωση Γειτόνων (t-Distributed Stochastic Neighbor Embedding, t-SNE) [25] είναι μια από τις καλύτερες μη γραμμικές τεχνικές ελάττωσης διαστάσεων για την οπτικοποίηση και την εξερεύνηση ογκωδών δεδομένων. Ειδικεύεται στη δημιουργία ενός ενιαίου χάρτη 2 ή 3 διαστάσεων, ο οποίος αποκαλύπτει δομές σε πολλές διαφορετικής κλίμακας, αλλά σχετιζόμενες, χαμηλών διαστάσεων πολλαπλότητες, όπως πχ εικόνες αντικειμένων της ίδιας κλάσης από πολλαπλές γωνίες θέασης [26].

Δεδομένου του συνόλου  $N$  υψηλοδιάστατων αντικειμένων  $x_1, x_2, \dots, x_N$ , η t-SNE υπολογίζει τις πιθανότητες  $p_{j|i}$  ανάλογες της ομοιότητας των αντικειμένων  $x_j$  και  $x_i$  αντίστοιχα ως εξής για  $i \neq j$ :

$$p_{j|i} = \frac{\exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma_i^2}\right)}{\sum_{k \neq i} \exp\left(-\frac{\|x_i - x_k\|^2}{2\sigma_i^2}\right)} \quad (2.37)$$

Θέτοντας  $p_{i|i} = 0$ .

Ισχύει ότι το άθροισμα πιθανοτήτων κάθε  $i$  μεταξύ όλων των  $j$  του είναι 1, δηλαδή:

$$\sum_j p_{j|i} = 1, \forall i \quad (2.38)$$

Το εύρος ζώνης  $\sigma$  των Γκαουσιανών πυρήνων ρυθμίζεται έτσι ώστε η εντροπία της υπό συνθήκης κατανομής να ισούται με μια προκαθορισμένη εντροπία.

Δηλαδή η ομοιότητα του σημείου  $x_j$  προς το  $x_i$  είναι η συνθήκη πιθανότητας  $p_{j|i}$ , η οποία υποδεικνύει την πιθανότητα το  $x_j$  να επιλέξει το  $x_i$  ως γείτονά του, δεδομένου ότι οι γείτονες επιλέγονται ανάλογα με την πυκνότητα πιθανότητας τους υπό μία Γκαουσιανή κατανομή κεντραρισμένη στο  $x_i$ .

Η ομοιότητα μεταξύ των σημείων  $i$  και  $j$  ορίζεται ως:

$$p_{ij} = \frac{p_{j|i} + p_{i|j}}{2N} \quad (2.39)$$

Τελικός στόχος είναι να ο αλγόριθμος να μπορέσει όσο το δυνατόν καλύτερα να αντιπροσωπεύσει αυτές τις πιθανοκρατικές  $p_{ij}$  σχέσεις σε έναν  $d$ -διάστατο χάρτη όπου  $y_i \in \mathbb{R}^d$  με συνήθεις τιμές  $d$  να είναι 2 ή 3. Έτσι για τον υπολογισμό της ομοιότητας σε αυτόν τον χώρο μεταξύ  $y_i$  και  $y_j$  όπου  $i \neq j$ :

$$q_{ij} = \frac{(1 + \|y_i - y_j\|^2)^{-1}}{\sum_k \sum_{l \neq k} (1 + \|y_k - y_l\|^2)^{-1}} \quad (2.40)$$

Τελικά οι θέσεις των  $y_i$  στον καινούργιο χώρο καθορίζονται μέσω της ελαχιστοποίησης της μη συμμετρικής Kullback-Liebler απόκλισης μεταξύ των κατανομών  $P$  και  $Q$ :

$$\text{KL}(P \parallel Q) = \sum_{i \neq j} p_{ij} \log \frac{p_{ij}}{q_{ij}} \quad (2.41)$$

Όσο μικρότερη αυτή η τιμή τόσο καλύτερα τα τελικά σημεία του χάρτη αντικατοπτρίζουν τα αρχικά. Συνήθως η επίλυση γίνεται μέσω της κατάβαση της κλίσης.

#### 2.2.3.2.2 ΟΜΟΙΟΜΟΡΦΗ ΠΡΟΣΕΓΓΙΣΗ ΚΑΙ ΠΡΟΒΟΛΗ ΠΟΛΛΑΠΛΟΤΗΤΩΝ

Η ομοιόμορφη προσέγγιση και προβολή πολλαπλότητων (Uniform Manifold Approximation and Projection, UMAP) προσφέρει αρκετά πλεονεκτήματα σε σύγκριση με την t-SNE, με κυριότερα την αυξημένη ταχύτητα και βελτιωμένη διατήρηση της παγκόσμιας δομής των δεδομένων. Δημιουργεί ένα ασαφές σύμπλεγμα (fuzzy simplicial complex) που αντιπροσωπεύει τη δομή των δεδομένων στον αρχικό χώρο, και στη συνέχεια προσπαθεί και αυτή να διατηρήσει αυτή τη δομή στον χώρο εξόδου [22].

Αρχικά κατασκευάζεται ένα γράφημα  $G$   $k$ -κοντινότερων γειτόνων ( $k$ -nearest neighbor graph). Στην συνέχεια υπολογίζονται οι ασαφής σχέσεις μεταξύ τους, δηλαδή για κάθε σημείο  $x_i$  υπολογίζεται η πιθανότητα σύνδεσης του με τον γείτονα  $x_j$ :

$$p_{ij} = \exp \left( -\frac{\max(0, d(x_i, x_j) - \rho_i)}{\sigma_i} \right) \quad (2.42)$$

Οπου  $\rho_i = \min_{j \in N_k(i), j \neq i} d(x_i, x_j)$  ελάχιστη απόσταση αποδοχής γείτονα για αποφυγή τοπολογικών βραχυκυκλωμάτων και συσσωμάτωσης,  $d$  μετρική απόστασης ή ομοιότητας και  $\sigma_i$  η κλίμακα που ρυθμίζει το εύρος πιθανοτήτων η οποία βρίσκεται μέσω αριθμητικής αναζήτησης, για αραιές περιοχές, θα είναι μεγαλύτερο, ενώ για πυκνές περιοχές μικρότερο.

Οι συμμετρικές πιθανότητες σύνδεσης μεταξύ των σημείων συνδυάζονται ως εξής:

$$w_{ij} = p_{ij} + p_{ji} - p_{ij} \cdot p_{ji} \quad (2.43)$$

Οι οποίες εκφράζουν την ισχύ σύνδεσης και ορίζουν την ασαφή συμπλοκή στον αρχικό χώρο.

Τελικά στον ελαττωμένο χώρο τα δεδομένα πάλι προβάλλονται έτσι ώστε να ελαχιστοποιούν τις αντίστοιχες μεταξύ τους αποστάσεις. Οι πιθανότητες σύνδεσης υπολογίζονται ως:

$$q_{ij} = \frac{1}{1 + a \|y_i - y_j\|^{2b}} \quad (2.44)$$

Με  $a, b$  υπερπαράμετροι οι οποίες υπολογίζονται με βάση μιας προκαθορισμένης συνάρτηση κατανομής πιθανότητας η οποία βελτιστοποιεί την υποκείμενη πολλαπλότητα, λαμβάνοντας υπόψη τη μέση πυκνότητα των δεδομένων και την επιλεγμένη παράμετρο εξάπλωσης, η οποία καθορίζει τη συνολική έκταση των δεδομένων στον χώρο εξόδου. Το  $a$  καθορίζει την ταχύτητα με την οποία η πιθανότητα σύνδεσης πέφτει καθώς αυξάνεται η απόσταση μεταξύ των σημείων ενώ το  $b$  διαμορφώνει το σχήμα της καμπύλης πιθανότητας, επηρεάζοντας ιδιαίτερα τις αλλαγές μικρών αποστάσεων.

Τελικά, παρομοίως, η προσαρμογή μεταξύ των δομών στον αρχικό και τελικό χώρο επιτυγχάνεται μέσω της ελαχιστοποίησης της συνάρτησης διασταυρούμενης εντροπίας:

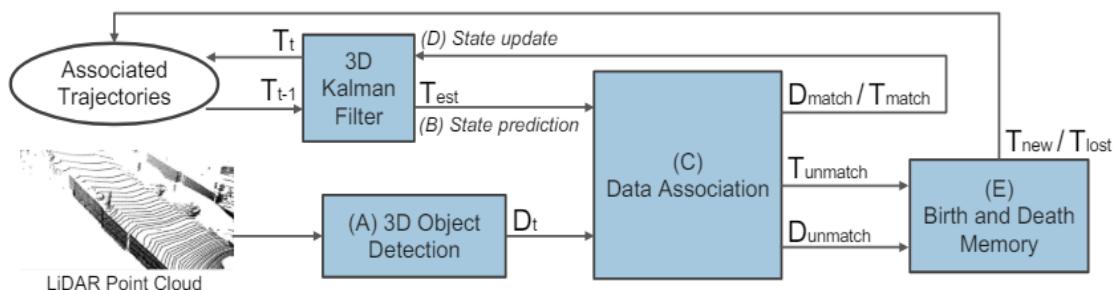
$$\mathcal{L} = \sum_{(i,j)} \left[ w_{ij} \log \frac{w_{ij}}{q_{ij}} + (1 - w_{ij}) \log \frac{1 - w_{ij}}{1 - q_{ij}} \right] \quad (2.45)$$

Η οποία προσεγγίζει το  $w_{ij}$  στο  $q_{ij}$ , διατηρώντας την τοπική και παγκόσμια δομή. Αλγόριθμοι όπως η στοχαστική κατάβαση μέσω της κλίσης μπορούν να χρησιμοποιηθούν για βελτίωση απόδοσης, καθώς και η κατασκευή του γραφήματος  $G$  μπορεί να επιταχυνθεί μέσω δομών όπως Ball Trees ή Annoy.

## 2.3 ΒΑΣΗ ΑΝΑΦΟΡΑΣ ΓΙΑ ΤΗΝ ΤΡΙΣΔΙΑΣΤΑΤΗ ΠΑΡΑΚΟΛΟΥΘΗΣΗ ΠΟΛΛΑΠΛΩΝ ΣΤΟΧΩΝ

Η βάση αναφοράς (baseline) της μονάδας παρακολούθησης προέρχεται από [27]. Για κάθε δειγματοληγία στον χρόνο  $t$ , γίνεται ο τρισδιάστατος εντοπισμός των στόχων ενδιαφέροντος γύρω της σκηνής του οχήματος (A). Το φίλτρο Kalman κάνει πρόβλεψη των καταστάσεων των υπάρχοντών ιχνών από τον χρόνο  $t-1$  στον  $t$  (B). Στην συνέχεια τα οι στόχοι και τα ίχνη εισάγονται στην μονάδα συσχέτιση δεδομένων (C) η οποία έχει ως έξοδο τρία (3) διαφορετικά, τα αντιστοιχημένα ίχνη στόχους ( $D_{match} / T_{match}$ ) τα μη αντιστοιχισμένα ίχνη ( $T_{unmatch}$ ) και τους μη αντιστοιχισμένα στόχους ( $D_{unmatch}$ ). Γίνεται η ενημέρωση των ιχνών (D) στο βήμα ενημέρωσης του φίλτρου Kalman. Στο τελευταίο βήμα της αρχιτεκτονικής οι μη αντιστοιχημένες οντότητες εισέρχονται στην μνήμη γέννησης και θανάτου (Birth and Death Memory). Εκεί αποφασίζεται εάν οι στόχοι θα ενταχθούν ως νέα ίχνη ( $T_{new}$ ) και εάν κάποια ίχνη θα σβηστούν ( $T_{lost}$ ).

Στο μοντέλο αναφοράς η συσχέτιση δεδομένων γίνεται μέσω χρήσης μήτρας συγγένειας μεταξύ στόχων και ιχνών. Η συγγένεια αυτή είναι η ευκλείδεια απόσταση μεταξύ κέντρων ή η τομή προς την ένωσή (Intersection over Union, IoU) των 3d πλαισίων του κάθε εντοπισμού προς ίχνους. Η ανάθεση των παρατηρήσεων στα αντίστοιχα ίχνη γίνεται επιλύοντας το πρόβλημα αντιστοιχίσης διμερούς γράφου (bipartite matching) πάνω στην μήτρα συγγένειας, κάτι που μπορεί να γίνει βέλτιστα, αλλά σε πολυωνυμικό χρόνο (αλγόριθμος Hungarian) [28]. Ο θάνατος των ιχνών γίνεται εάν το ίδιο ίχνος δεν έχει ανανεωθεί για κάποια συνεχόμενα βήματα του αλγόριθμου. Η γέννηση νέων ιχνών στην πιο απλή αντιμετώπιση γίνεται για κάθε μη αντιστοιχισμένο στόχο.



Εικόνα 2-20: Βασική Αρχιτεκτονική Παρακολούθησης

## 2.4 ΕΚΤΙΜΗΣΗ ΚΙΝΔΥΝΟΥ ΣΤΗΝ ΑΥΤΟΝΟΜΗ ΠΛΟΗΓΗΣΗ

Το πρόβλημα της εκτίμησης κινδύνου σύγκρουσης είναι θεμελιώδες για την αυτόνομη πλοήγηση και τα προηγμένα συστήματα υποβοήθησης οδήγησης (ADAS). Προκύπτει από την βασική ανάγκη των αυτόνομων οχημάτων να ελίσσονται με ασφάλεια σε δυναμικά περιβάλλοντα, με πολλαπλά κινούμενα - και μη - αντικείμενα γύρω τους. Η σημασία της μοντελοποίησης του κινδύνου έγκειται στην λήψη έγκαιρων και κατάλληλων αποφάσεων για την αποφυγή συγκρούσεων. Χωρίς αυτή την ικανότητα, η αυτόνομη οδήγηση θα ήταν εξαιρετικά επικίνδυνη και πρακτικά αδύνατη [29].

Πρώτον, απαιτείται ο ακριβής εντοπισμός και παρακολούθηση των αντικειμένων στο περιβάλλον. Δεύτερον, είναι απαραίτητη η πρόβλεψη των μελλοντικών τροχιών τόσο του αυτόνομου οχήματος όσο και των άλλων αντικειμένων, λαμβάνοντας υπόψη την αβεβαιότητα στις κινήσεις τους σε βάθος χρόνου. Τρίτον, απαιτείται ένας τρόπος να μοντελοποιηθεί και να υπολογιστεί η έννοια του κινδύνου σύγκρουσης. Η πιθανότητα κατάστασης σύγκρουσης (collision state probability, CSP), και πιθανότητα συμβάντος σύγκρουσης (collision event probability, CEP) σε βάθος χρόνου, αποτελούν ουσιώδη μαθηματικές ερμηνείες της πιθανότητας αυτής. Τέλος, όλοι αυτοί οι υπολογισμοί πρέπει να γίνονται σε πραγματικό χρόνο και με υψηλή ακρίβεια, ώστε το όχημα να μπορεί να αντιδρά άμεσα στις μεταβαλλόμενες συνθήκες [30].

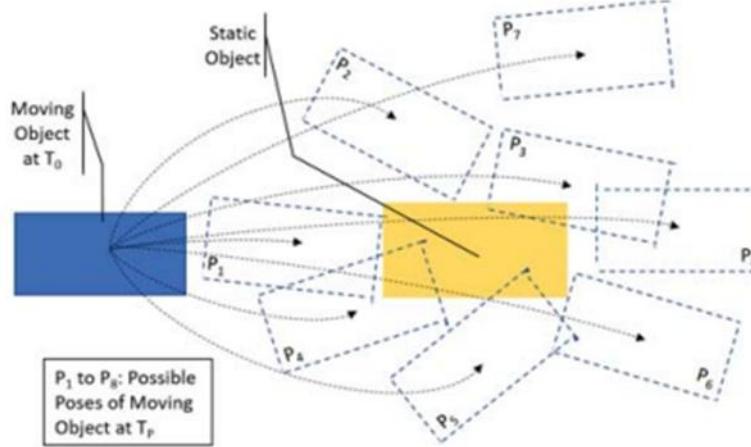
### 2.4.1 ΠΙΘΑΝΟΤΗΤΑ ΣΥΓΚΡΟΥΣΗΣ ΚΑΤΑΣΤΑΣΗΣ ΚΑΙ ΠΥΚΝΟΤΗΤΑ

#### ΠΙΘΑΝΟΤΗΤΑΣ ΣΥΜΒΑΝΤΟΣ ΣΥΓΚΡΟΥΣΗΣ

Η πιθανότητα σύγκρουσης κατάστασης ορίζεται ως η πιθανότητα χωρικής επικάλυψης μεταξύ δύο αντικειμένων σε μία συγκεκριμένη χρονική στιγμή για ένα παράθυρο πρόβλεψης κάποιων δευτερολέπτων. Η πιθανότητα κατάστασης σύγκρουσης αντιπροσωπεύει την πιθανότητα το αντικείμενο να βρίσκεται σε κατάσταση σύγκρουσης τη στιγμή της πρόβλεψης Tr, ενώ η πιθανότητα συμβάντος σύγκρουσης δείχνει την πιθανότητα να συμβεί ένα γεγονός σύγκρουσης μέχρι ή και ακριβώς τη στιγμή Tr.

Όπως φαίνεται στην Εικόνα 2-21 τα γεγονότα σύγκρουσης (P1, P2) αντιπροσωπεύουν την ακριβή στιγμή που το όριο του κινούμενου αντικειμένου μόλις αγγίζει ή τέμνει το στατικό αντικείμενο, σηματοδοτώντας την έναρξη μιας σύγκρουσης. Με αυτόν τον τρόπο εκφράζεται ο ακριβής χρόνος έναρξης μιας σύγκρουσης. Οι καταστάσεις σύγκρουσης (P1 έως P6)

περιλαμβάνουν οποιαδήποτε θέση όπου τα δύο αντικείμενα επικαλύπτονται, ακόμη και μερικώς, αντιπροσωπεύοντας συνεχείς καταστάσεις επαφής ή διείσδυσης. Το P6, για παράδειγμα, δείχνει μια κατάσταση σύγκρουσης αλλά όχι ένα γεγονός, καθώς η επικάλυψη έχει ήδη συμβεί. Οι θέσεις P7 και P8 δεν αποτελούν ούτε γεγονότα ούτε καταστάσεις σύγκρουσης, καθώς δεν υπάρχει επαφή ή επικάλυψη [31].



Εικόνα 2-21: Γεγονότα Σύγκρουσης

#### 2.4.2 ΓΕΝΙΚΗ ΛΥΣΗ

Για τον υπολογισμό του κινδύνου σύγκρουσης της κοινής κατανομής μεταξύ δύο ανεξάρτητων δισδιάστατων οντοτήτων απαιτείται η επίλυση του παρακάτω ολοκληρώματος:

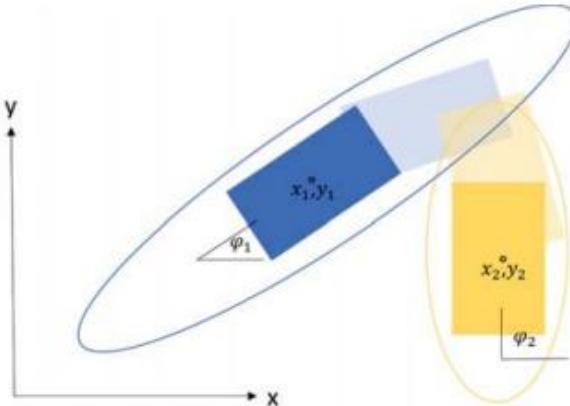
$$\int_{x_1} \int_{y_1} \int_{\varphi_1} \int_{x_2} \int_{y_2} \int_{\varphi_2} I_C(x_1, y_1, \varphi_1, x_2, y_2, \varphi_2) p(x_1, y_1, \varphi_1) p(x_2, y_2, \varphi_2) dx_1 dy_1 d\varphi_1 dx_2 dy_2 d\varphi_2 \quad (2.46)$$

Όπου  $(x_1, y_1, \varphi_1)$  και  $(x_2, y_2, \varphi_2)$  οι θέσεις και προσανατολισμοί των δύο αντικειμένων.  $p(x_1, y_1, \varphi_1)$  και  $p(x_2, y_2, \varphi_2)$  οι συναρτήσεις πυκνότητας πιθανότητας για τις καταστάσεις των αντικειμένων και  $I_C$  η συνάρτηση δείκτης σύγκρουσης η οποία επιστρέφει 1 εάν τα δύο ορθογώνια επικαλύπτονται, έστω και μερικώς.

$$I_C(x_1, y_1, \varphi_1, x_2, y_2, \varphi_2) = 1 \\ \text{if} \\ S(x_1, y_1, \varphi_1) \cap S(x_2, y_2, \varphi_2) \neq \emptyset \quad (2.47)$$

$S(x, y, \varphi)$  το σύνολο των σημείων που καταλαμβάνει ένα αντικείμενο στη δεδομένη κατάσταση, κάτι ασυνεχές εξού και η αδυναμία αναλυτικής επίλυσης του ολοκληρώματος

στην γενική μορφή. Επιπρόσθετα, οι προβλεπόμενες καταστάσεις ενός αντικειμένου σε διαφορετικές χρονικές στιγμές δεν είναι ανεξάρτητες εξ αυτού δεν έχει νόημα να αθροίσουμε τους κινδύνους σύγκρουσης σε διαφορετικές χρονικές στιγμές για να υπολογίσουμε τον συνολικό κίνδυνο σε ένα χρονικό διάστημα.



Εικόνα 2-22: Γενική Λύση στο πρόβλημα Υπολογισμού Κίνδυνου Σύγκρουσης

Για να υπολογισθεί με ακρίβεια ο συνολικός κίνδυνος σύγκρουσης σε ένα χρονικό διάστημα, θα έπρεπε να αξιολογηθεί η κοινή κατανομή των δύο αντικειμένων σε όλα τα χρονικά βήματα του ορίζοντα πρόβλεψης, κάτι που θα απαιτούσε την επίλυση ενός ολοκληρώματος ακόμη υψηλότερης διάστασης, το οποίο είναι πρακτικά αδύνατο λόγω υπολογιστικής πολυπλοκότητας [31].

#### 2.4.3 ΕΠΙΛΥΣΗ ΜΕΣΩ ΠΡΟΣΟΜΟΙΩΣΗΣ MONTE CARLO

Και οι δύο πιθανότητες μπορούν να υπολογιστούν με προσομοιώσεις Monte Carlo, χωρίς ωστόσο να ανταποκρίνονται στις απαιτήσεις πραγματικού χρόνου (real-time). Σε κάθε χρονικό βήμα, λαμβάνεται ένας αριθμός δειγμάτων από τις κατανομές καταστάσεων των δύο αντικειμένων και ελέγχεται αν τα ορθογώνια τους επικαλύπτονται. Το ποσοστό των περιπτώσεων επικάλυψης σε σχέση με το σύνολο των δειγμάτων αντιπροσωπεύει το CSP σε εκείνο το χρονικό σημείο. Ο υπολογισμός του CEP απαιτεί τη λήψη ενός αριθμού πλήρων δειγμάτων τροχιάς. Κατά την έναρξη της προσομοίωσης, λαμβάνονται δείγματα από την αρχική κατανομή για την αρχικοποίηση κάθε τροχιάς. Σε κάθε χρονικό βήμα, αυτές οι τροχιές προβλέπονται χρησιμοποιώντας ένα κατάλληλο μοντέλο διαδικασίας, και στη συνέχεια προστίθενται δείγματα από τις κατανομές θορύβου της διαδικασίας στην κατάσταση. Το ποσοστό των νέων τροχιών που συγκρούονται σε κάθε χρονικό βήμα αντιπροσωπεύει την πυκνότητα του CEP. Αυτές οι τροχιές αφαιρούνται πριν την εφαρμογή του επόμενου βήματος

πρόβλεψης. Το ποσοστό των τροχιών που συγκρούστηκαν σε ένα χρονικό διάστημα αποτελεί το CEP για αυτό το διάστημα.

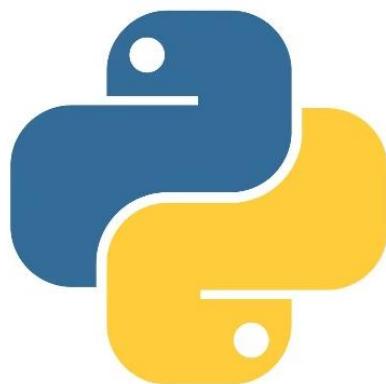
Καθώς αυξάνεται ο αριθμός των δειγμάτων, τα αποτελέσματα συγκλίνουν στην πραγματικότητα, αυξάνοντας ωστόσο τον χρόνο υπολογισμού. Οι μικρές τιμές κινδύνου απαιτούν περισσότερα δείγματα σε σύγκριση με τις υψηλότερες. Ο αριθμός των δειγμάτων που απαιτείται για ρεαλιστικά αποτελέσματα εξαρτάται κυρίως από τον αριθμό των πιθανοτικών μεταβλητών κατάστασης.

## **3 ΕΡΓΑΛΕΙΑ ΠΟΥ ΧΡΗΣΙΜΟΠΟΙΗΘΗΚΑΝ**

---

### **3.1 Python**

Η Python είναι μια γλώσσα προγραμματισμού υψηλού επιπέδου, δημοφιλής για την ευελιξία και την απλότητά της αλλά και πασίγνωστη για την τεράστια κοινότητα της τα τελευταία χρόνια. Χρησιμοποιείται για ανάπτυξη κώδικα και υλοποίηση μοντέλων και αλγορίθμων παρέχοντας μία πληθώρα δωρεάν βιβλιοθηκών για ανάλυση δεδομένων, μηχανική μάθηση και μαθηματικούς υπολογισμούς. Χρησιμοποιούνται οι εκδόσεις 3.6, 3.7, 3.9, 3.11 για τα μοντέλα της υλοποίησης.



*Eικόνα 3-1: Python*

### **3.2 GitHub**

Το GitHub είναι μια από τις πιο δημοφιλείς πλατφόρμες συνεργασίας και διαχείρισης λογισμικού, σχεδιασμένη για να διευκολύνει την ανάπτυξη, την αποθήκευση και τη συντήρηση κώδικα. Βασίζεται στο Git, ένα σύστημα ελέγχου εκδόσεων, που επιτρέπει στην κοινότητα να παρακολουθεί αλλαγές στον κώδικα, να εργάζεται σε διαφορετικά παρακλάδια (branches) και να ενσωματώνει νέες δυνατότητες χωρίς να επηρεαστεί ο βασικός κορμός. Υπάρχει μία κεντρική αποθήκευση έργων, ώστε να είναι προσβάσιμα από οπουδήποτε και να μπορούν να ανακτηθούν σε περίπτωση απώλειας. Περιέχει εργαλεία που διευκολύνουν την ανασκόπηση προηγούμενων εκδόσεων, την επαναφορά σε σταθερές εκδόσεις και την παρακολούθηση αλλαγών που έγιναν από αλλά μέλη. Μέσω αιτημάτων pull (pull requests), μπορούν να προταθούν αλλαγές από άλλους χρήστες, δίνοντας επίσης δυνατότητες σχολιασμού και συζήτησης πάνω σε συγκεκριμένα μέρη των υλοποιήσεων. Πολλές κοινότητες και οργανισμοί δημοσιεύουν έργα ανοιχτού κώδικα στο GitHub, επιτρέποντας συμβολή από όλο τον κόσμο.

Ενσωματώνει εργαλεία όπως πίνακες Kanban, διαχείριση ζητημάτων (issues), και αυτόματες ροές εργασίας (workflows) μέσω του GitHub Actions, επιτρέποντας την αυτοματοποίηση καθημερινών εργασιών για δοκιμές και ανάπτυξη. Τέλος, προσφέρει εργαλεία όπως έλεγχο ευπαθειών, εξασφαλίζοντας ότι τα έργα είναι ασφαλή από απειλές.



Εικόνα 3-2: GitHub

### 3.3 WSL2 και Ubuntu

Το WSL2 (Windows Subsystem for Linux 2) αποτελεί τεχνολογία της Microsoft. Επιτρέπει την εκτέλεση ενός πλήρους Linux περιβάλλοντος μέσα σε Windows. Η χρήση του σε συνδυασμό με Ubuntu, έναν από τους πιο δημοφιλείς διανομείς Linux, προσφέρει ευκολία ανάπτυξης και εκτέλεσης εφαρμογών βασισμένων σε Linux χωρίς την ανάγκη ξεχωριστού υπολογιστή ή dual boot. Το WSL2 χρησιμοποιεί μια πραγματική μηχανή εικονικοποίησης (Virtual Machine, VM) για να επιτύχει ταχύτητα και πλήρη συμβατότητα με τις βιβλιοθήκες και τα εργαλεία της Linux.

### 3.4 Anaconda

Η Anaconda είναι μια δημοφιλής πλατφόρμα ανοικτού κώδικα για ανάπτυξη εφαρμογών πάνω στην Python, ειδικότερα στην επιστήμη δεδομένων και τη μηχανική μάθηση. Παρέχει ένα ολοκληρωμένο περιβάλλον με εύκολη εγκατάσταση και διαχείριση βιβλιοθηκών και εξαρτήσεων μέσω του package manager conda. Περιλαμβάνει διάφορα εργαλεία όπως το Jupyter Notebook και Spyder, καθιστώντας το ιδανικό για ανάλυση δεδομένων, δημιουργία μοντέλων μηχανικής μάθησης και ανάπτυξη κώδικα. Με τη δυνατότητα δημιουργίας απομονωμένων εικονικών περιβαλλόντων, βοηθά στην αποφυγή συγκρούσεων μεταξύ βιβλιοθηκών και εκδόσεων, διευκολύνοντας τη διαχείριση διαφορετικών έργων.

## 3.5 NumPy

Η NumPy (Numerical Python) είναι μια από τις πιο βασικές και ευρέως χρησιμοποιούμενες βιβλιοθήκες για μαθηματικούς και αριθμητικούς υπολογισμούς στην Python. Παρέχει εξαιρετική υποστήριξη για πολυδιάστατους πίνακες και μια σειρά εργαλείων και λειτουργιών επιτάχυνσης αυτών υλοποιημένα σε C. Αποτελεί πυλώνα για την εκτέλεση επιστημονικών υπολογισμών στην Python και χρησιμεύει ως θεμέλιο για πολλές άλλες βιβλιοθήκες, όπως την SciPy και την Matplotlib. Χρησιμοποιούνται οι εκδόσεις 1.21.6, 1.22.3, 1.26.4.

## 3.6 FilterPy

Η βιβλιοθήκη FilterPy παρέχει εργαλεία για την εφαρμογή αλγορίθμων φιλτραρίσματος, όπως αυτή του Kalman, καθώς και άλλες αντίστοιχες ή και πιο πολύπλοκες μεθόδους όπως το φίλτρο σωματιδίων (particle filter). Είναι χρήσιμη σε εφαρμογές που απαιτούν την εκτίμηση ή την πρόβλεψη καταστάσεων πολυδιάστατων δυναμικών συστημάτων υπό θόρυβο. Χρησιμοποιείται ευρέως σε πεδία ρομποτικής, συχνότερα στην παρακολούθηση αντικειμένων. Χρησιμοποιείται η έκδοση 1.4.5.

## 3.7 SciPy

Η βιβλιοθήκη SciPy επεκτείνει τις δυνατότητες της NumPy, προσφέροντας ένα ευρύτερο και πιο εξεζητημένο φάσμα εργαλείων για πιο πολύπλοκους επιστημονικούς υπολογισμούς. Χρησιμοποιείται κυρίως σε τομείς γραμμικής άλγεβρας όπως στον υπολογισμό ιδιοτιμών και αποσυνθέσεων (π.χ. SVD), στην βελτιστοποίηση, στην επεξεργασία σήματος, στην στατιστική, στην επίλυση διαφορικών εξισώσεων μέσω αριθμητικών μεθόδων καθώς και στην διαχείριση και επεξεργασία χωρικών δεδομένων. Χρησιμοποιείται η έκδοση 1.10.1, 1.13.1.

### 3.7.1 ΣΥΝΑΡΤΗΣΗ ΟΛΟΚΛΗΡΩΣΗΣ ΤΗΣ SciPy

Η συνάρτηση `quad` από το πακέτο `scipy.integrate` υπολογίζει ορισμένα ολοκληρώματα μίας μεταβλητής. Χρησιμοποιεί την αριθμητική μέθοδο QUADPACK. Οι είσοδοι της είναι:

- `func`: Η συνάρτηση που θα ολοκληρωθεί. Μπορεί να είναι οποιαδήποτε συνάρτηση ή έκφραση.
- `a`: Το κατώτερο όριο ολοκλήρωσης.
- `b`: Το ανώτερο όριο ολοκλήρωσης.

Και οι έξοδοι της είναι:

- result: Το αποτέλεσμα του ολοκληρώματος.
- error: Η εκτιμώμενη αριθμητική απόκλιση.

## 3.8 TensorFlow

Η TensorFlow είναι ένα ανοιχτού κώδικα πλαίσιο μηχανικής μάθησης, το οποίο αναπτύχθηκε από την Google Brain. Χρησιμοποιείται για την κατασκευή, την εκπαίδευση και την ανάπτυξη μοντέλων βαθιάς μάθησης και μηχανικής μάθησης. Είναι γνωστή για την ευελιξία, την επεκτασιμότητα και τη δυνατότητα λειτουργίας της σε διάφορες πλατφόρμες, όπως υπολογιστές, διακομιστές, κινητές συσκευές και ενσωματωμένα συστήματα. Χρησιμοποιείται η έκδοση 2.1.0.



*Eικόνα 3-3: TensorFlow*

## 3.9 PyTorch

Η PyTorch αναπτύχθηκε από το Facebook AI Research (FAIR). Χρησιμοποιείται και αυτή ευρέως για την ανάπτυξη και εκπαίδευση μοντέλων μηχανικής μάθησης. Το δυναμικό υπολογιστικό της γράφημα (dynamic computation graph) προσφέρει ευελιξία στις προσεγγίσεις μοντελοποίησης υπολογισμών, κάνοντάς την πιο εύκολη και εύχρηστή σε σύγκριση με άλλα πλαίσια, όπως την TensorFlow. Χρησιμοποιείται οι εκδόσεις 1.13.1, 2.3.0.



Eikόνα 3-4: PyTorch

### 3.9.1 ΠΑΚΕΤΟ ΝΕΥΡΩΝΙΚΩΝ ΔΙΚΤΥΩΝ ΤΗΣ PyTorch

Το `torch.nn` είναι το βασικότερο πακέτο της PyTorch για τη δημιουργία, των σχεδιασμό, και την εκπαίδευση νευρωνικών δικτύων. Παρέχει εργαλεία και δομές για τη διαχείριση πολυεπίπεδων μοντέλων, την υλοποίηση διαφόρων τύπων στρωμάτων, συναρτήσεων ενεργοποίησης και συναρτήσεων κόστους. Η κλάση `nn.Module` είναι η βάση όλων των νευρωνικών δικτύων. Όλα τα μοντέλα καθορίζονται ως υποκλάση του κάνοντας τα πάντα εύκολα διαχειρίσιμα.

### 3.9.2 ΣΥΝΑΡΤΗΣΗ ΠΛΕΓΜΑΤΟΣ ΤΗΣ PyTorch

Η συνάρτηση `grid_sample` χρησιμοποιείται σε υπολογισμούς χαρακτηριστικών και γενικευμένων μορφολογικών μετασχηματισμών σε έναν πίνακα ή τένσορα χρησιμοποιώντας ένα καθορισμένο πλέγμα, το οποίο συνήθως είναι κανονικοποιημένο σε τιμές μεταξύ [-1, 1]. Οι παράμετροι της είναι:

- Input: Πίνακας (ή εικόνα) από τον οποίο θα γίνει η δειγματοληψία. Συνήθως είναι ένα 4D τανυστής με σχήμα (παρτίδα, κανάλια, πλάτος, μήκος).
- Grid: Τανυστής με τις συντεταγμένες του πλέγματος για τη δειγματοληψία. Συνήθως σχήματος (παρτίδα κανάλια εξόδου, πλάτος εξόδου, 2) όπου οι τελευταίοι δύο άξονες είναι οι κανονικοποιημένες (x, y) συντεταγμένες του πλέγματος.
- mode: Ο τρόπος δειγματοληψίας. Υποστηρίζονται:

'bilinear': Δειγματοληψία μέσω γραμμικής παρεμβολής.

'nearest': Δειγματοληψία μέσω της κοντινότερης τιμής (nearest neighbor).

- padding\_mode: Ο τρόπος χειρισμού των σημείων του πλέγματος που βρίσκονται εκτός των ορίων του πίνακα εισόδου. Υποστηρίζονται:

'zeros': Στοχεύει στις μηδενικές τιμές για τις εκτός ορίων περιοχές.

'border': Χρησιμοποιεί τις τιμές του ορίου για τις εκτός ορίων περιοχές.

'reflection': Αντιστρέφει το περιεχόμενο της εικόνας για τις εκτός ορίων περιοχές.

- align\_corners: Εάν είναι `True`, οι γωνίες του κανονικοποιημένου πλέγματος θα ευθυγραμμιστούν με τις γωνίες της εικόνας εισόδου. Εάν είναι `False`, το πλέγμα δεν θα ευθυγραμμιστεί πλήρως με τις γωνίες.

Έξοδος:

Ένας νέος πίνακας με τα δειγματοληπτημένα δεδομένα, το σχήμα του οποίου εξαρτάται από το μέγεθος της εξόδου του πλέγματος.

### 3.9.3 ΒΕΛΤΙΣΤΟΠΟΙΗΤΗΣ Adam THΣ PyTorch

Ο αλγόριθμος Adam (Adaptive Moment Estimation) είναι ένας αλγόριθμος βελτιστοποίησης για την εκπαίδευση νευρωνικών δικτύων. Στην PyTorch, παρέχεται μέσω της κλάσης `torch.optim.Adam`. Προτείνεται τόσο για μικρά όσο και για μεγάλα δίκτυα λόγω της αποτελεσματικότητάς του και της δυνατότητάς του να προσαρμόζει αυτόματα τον ρυθμό εκμάθησης, υπολογίζοντας ξεχωριστούς ρυθμούς μάθησης για κάθε παράμετρο χρησιμοποιώντας την κλίση πρώτης και δεύτερης τάξης (gradient και squared gradient). Υπολογίζει την εκθετική κινούμενη μέση τιμή των κλίσεων (1ης τάξης) για να εξομαλύνει τις ταλαντώσεις και ενσωματώνει την εκθετική μέση τιμή του τετραγώνου των κλίσεων για να εξισορροπήσει τη συμβολή των παραμέτρων μεγάλης ή μικρής κλίσης.

## 3.10 scikit-learn

Η scikit-learn (ή `sklearn`) παρέχει εργαλεία για την εφαρμογή αλγορίθμων ταξινόμησης (πχ SVM, Random Forests) παλινδρόμησης (πχ Lasso), ομαδοποίησης (πχ DBSCAN), μείωσης διάστασης (πχ PCA), εξόρυξης συσχετίσεων και αναγνώριση ανωμαλιών (πχ Isolation Forest), καθώς και για άλλες τεχνικές όπως βελτιστοποίηση μοντέλων και αξιολόγηση της απόδοσης. Κύρια χρήση στην εργασία βρίσκει στις τεχνικές μείωσης διαστασιμότητας.

## 3.11 Cuda και cuDNN

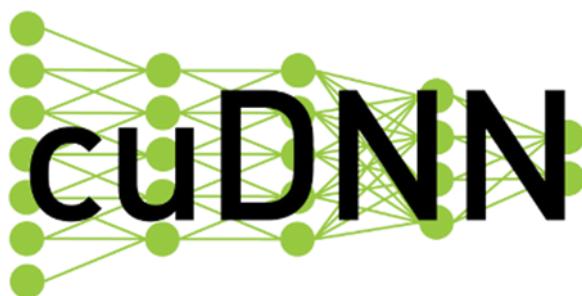
Η CUDA (Compute Unified Device Architecture) είναι μια πλατφόρμα και αρχιτεκτονική προγραμματισμού που έχει αναπτυχθεί από την NVIDIA, σχεδιασμένη να αξιοποιεί την

επεξεργαστική ισχύ των GPUs (Graphics Processing Units) για γενικούς υπολογισμούς πέρα από την απεικόνιση γραφικών, το οποίο μπορεί να χρησιμοποιηθεί για ταχύτερη επεξεργασία σε πολλές γλώσσες προγραμματισμού όπως python. Είναι ιδανική για υπολογισμούς σε μεγάλα σύνολα δεδομένων. Αρχικά κατασκευάστηκε για χρήση υπολογιστικών προσομοιώσεων και συγκεκριμένα ρευστών. Χρησιμοποιείται η έκδοση 11.7 και 11.8.



*Eικόνα 3-5: Cuda*

Η cuDNN (CUDA Deep Neural Network Library) είναι μια βελτιστοποιημένη βιβλιοθήκη της NVIDIA, ειδικά σχεδιασμένη για την επιτάχυνση αλγορίθμων βαθιάς μάθησης (Deep Learning) που εκτελούνται σε GPUs. Αποτελεί βασικό κομμάτι των βιβλιοθηκών TensorFlow, PyTorch, MXNet, και Caffe, προσφέροντας μέγιστη απόδοση στους υπολογισμούς νευρωνικών δικτύων.



*Eικόνα 3-6: CuDNN*

### 3.12 ΠΡΩΤΟΚΟΛΛΟ SSH

Το πρωτόκολλο SSH (Secure Shell) είναι ένα κρυπτογραφημένο πρωτόκολλο δικτύου που επιτρέπει την απομακρυσμένη διαχείριση συστημάτων και τη μεταφορά δεδομένων με ασφάλεια. Χρησιμοποιείται ευρέως για την απομακρυσμένη σύνδεση και τον έλεγχο

υπολογιστών ή διακομιστών προσφέροντας ισχυρή κρυπτογράφηση προστασίας των δεδομένων που μεταφέρονται καθ' όλην τη διάρκεια σύνδεσης. Παρέχει δυνατότητες για την ανίχνευση και την αντιμετώπιση τυχόν προβλημάτων ή δυσλειτουργιών που μπορεί να προκύψουν κατά τη διάρκεια της εκπαίδευσης των μοντέλων.

### 3.13 ΥΠΟΛΟΓΙΣΤΙΚΗ ΥΠΟΔΟΜΗ

Για την εκτέλεση όλων των μονάδων του έργου χρησιμοποιήθηκε διακομιστής (server) 125 GB RAM, 20 CPU cores και δύο GPUs (RTX 3090 24 GB), με τον οποίο η σύνδεση έγινε μέσω πρωτοκόλλου Ασφαλούς Κελύφους (Secure Shell, SSH), επιτρέποντας τον απομακρυσμένο έλεγχο και διαχείριση των συστημάτων και GitHub μοντέλων.

### 3.14 ΑΠΟΘΕΤΗΡΙΑ

Για την υλοποίηση του αλγορίθμου εξαγωγής γεωμετρικών χαρακτηριστικών επιλέχθηκε το αποθετήριο OpenPCDet<sup>3</sup> το οποίο περιέχει μία πληθώρα τρισδιάστατων ανιχνευτών μαζί και τον CenterPoint με πολλαπλά προ-εκπαιδευμένα μοντέλα υποστηρίζοντας τα πιο γνωστά σύνολα δεδομένων. Χρησιμοποιούνται πολλαπλά προ-εκπαιδευμένα βάρη δικτύων διαφορετικών κυψελών.



Εικόνα 3-7: OpenMMLab<sup>4</sup>

Για την ανάπτυξη του αλγορίθμου εξαγωγής εμφανισιακών χαρακτηριστικών χρησιμοποιήθηκε ως βάση το κοινοποιημένο αποθετήριο Mask R-CNN από το GitHub. Χρησιμοποιείται το προ-εκπαιδευμένο δίκτυο στο σύνολο δεδομένων COCO [32] με όνομα mask\_rcnn\_coco.h5.

---

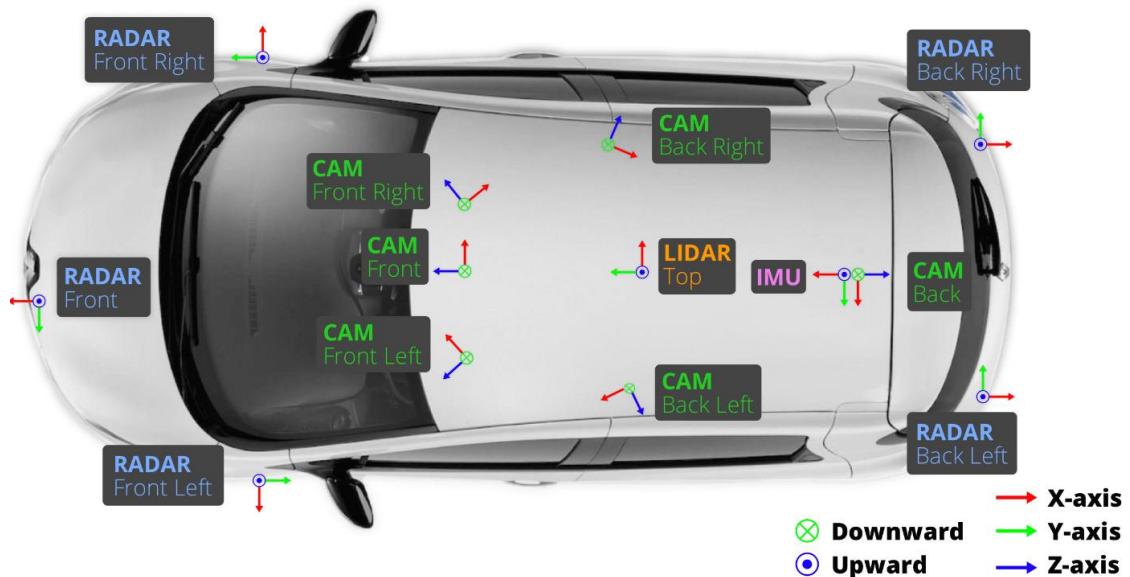
<sup>3</sup> <https://github.com/open-mmlab/OpenPCDet>

<sup>4</sup> <https://github.com/open-mmlab>

Το σύστημα παρακολούθησης κτίζεται πάνω στο αποθετήριο<sup>5</sup> με το οποίο γίνεται και σύγκριση.

### 3.15 ΤΟ ΣΥΝΟΛΟ ΔΕΔΟΜΕΝΩΝ NUSCENES

Το nuScenes<sup>6</sup>, είναι ένα ανταγωνιστικό πυκνό σύνολο δεδομένων (dataset) αποτελούμενο από 6 κάμερες, 1 LiDAR, 5 radar, 1 IMU και GNSS για εργασίες στην αυτόνομη οδήγηση. Αποτελείται από 1000 σκηνές, κάθε μία από 40 δείγματα (samples) με συγχρονισμένη δειγματοληγία. Κάθε σκηνή διαρκεί 20 δευτερόλεπτα. 700 σκηνές διατίθενται για μάθηση, 150 για συμπερασμό (train-val split) και οι υπόλοιπες για τεστ (test split). Συνολικά το nuScenes απαιτεί 360 GB μνήμης [33].



Εικόνα 3-8: Οψη και αισθητήρες του αυτόνομου οχήματος από το σύνολο δεδομένων

Αποτελείται από πολλές καταστάσεις καιρού σε μέρα και νύχτα, κάνοντας το ένα από τα πιο απαιτητικά σύνολα, ταυτόχρονα έχει σοβαρή ανισορροπία κλάσεων [34]. Παρακάτω εμφανίζεται ένα στιγμιότυπο, με πολλαπλά αντικείμενα προς εντοπισμό και παρακολούθηση με τις προβολές του νέφους στις 6 κάμερες, μέρος της σκηνής 1048.

<sup>5</sup> [https://github.com/eddyhkchiu/mahalanobis\\_3d\\_multi\\_object\\_tracking.git](https://github.com/eddyhkchiu/mahalanobis_3d_multi_object_tracking.git)

<sup>6</sup> <https://www.nuscenes.org/>



Εικόνα 3-9: Στιγμιότυπο σκηνής συνόλου δεδομένων

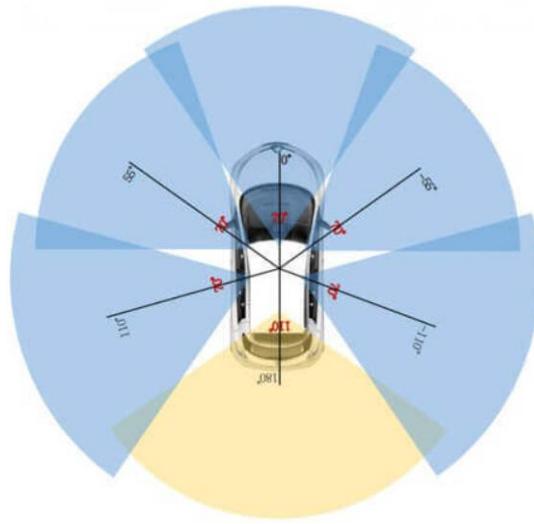
Οι ενδογενείς μετασχηματισμοί κάθε κάμερας και οι εξωγενής μετασχηματισμοί για κάθε αισθητήρα με βάση το ρομποκεντρικό σύστημα αναφοράς (μεσοδιάστημα του πίσω άξονα) είναι δεδομένοι. Οι παγκόσμιες συντεταγμένες βρίσκονται κάτω αριστερά του γνωστού χάρτη, η ακριβής τοποθέτηση του οχήματος θεωρείται ντετερμινιστική και λαμβάνεται από τα δεδομένα του συνόλου [33].



Εικόνα 3-10: Αισθητήρες που χρησιμοποιούνται

Ο ρυθμός δειγματοληγίας είναι συγχρονισμένος στα 2 Hz (μπορεί να υπάρχουν αμελητέες αποκλίσεις) για τις κάμερες και το LiDAR, με τα δεδομένα που αποδίδουν να ονομάζονται καρέ-κλειδιά (keyframes). Η χρονική βαθμονόμηση επιτυγχάνεται ενεργοποιώντας τις

κάμερες όταν το LiDAR διέρχεται από το κέντρο του πεδίου όρασης της κάθε μίας<sup>7</sup>.

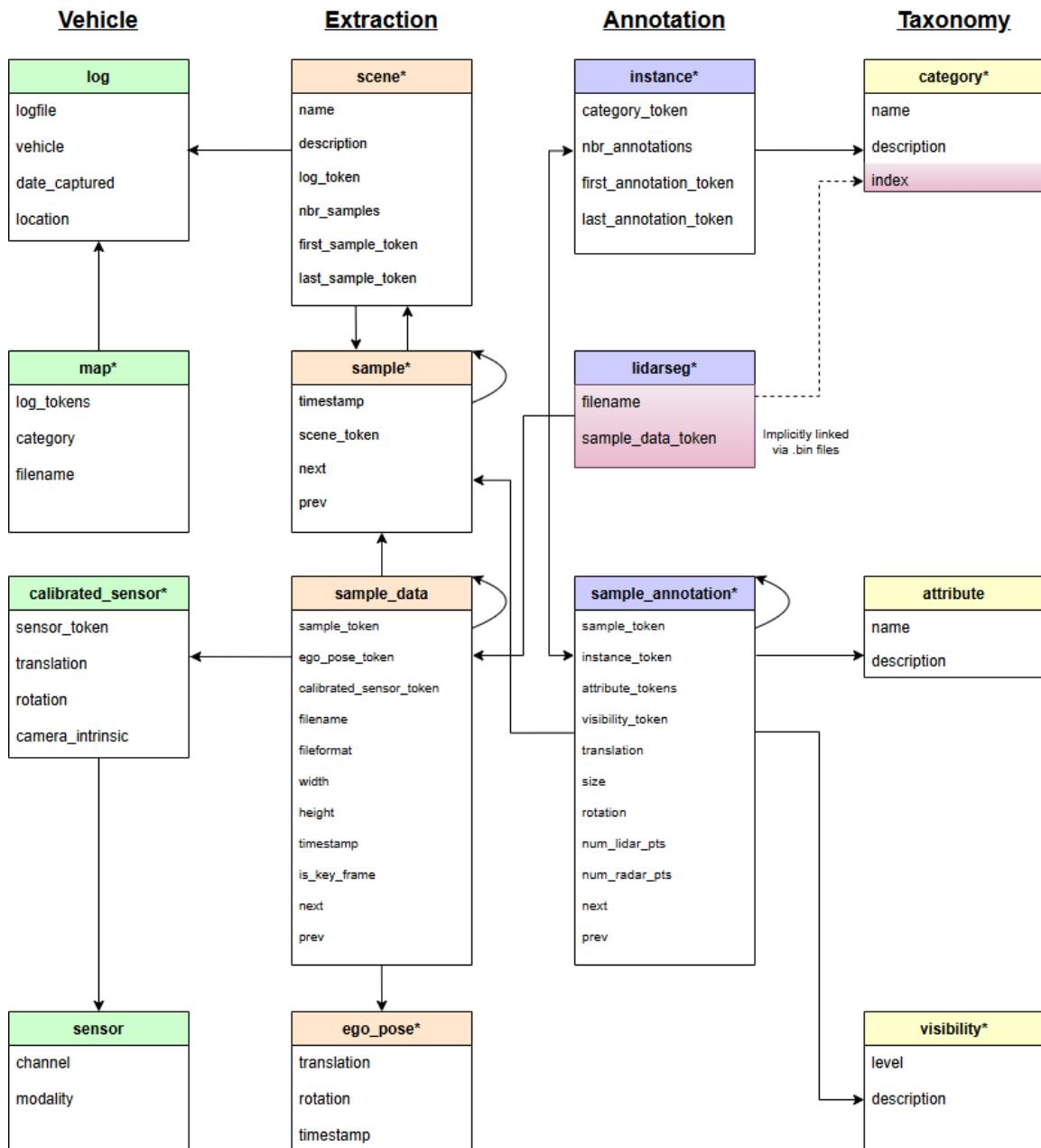


Εικόνα 3-11: Πεδίο όρασης αισθητήρων

Στην Εικόνα 3-11 φαίνεται το πεδίο όρασης των καμερών του συνόλου δεδομένων. Στην Εικόνα 3-12 φαίνεται η αρχιτεκτονική δεδομένων του συνόλου. Μέσω διαφόρων κλάσεων και συναρτήσεων γίνεται η ανάκληση πληροφοριών από διάφορα λεξικά και λίστες για την κατάσταση διαφόρων αντικειμένων. Διαφορετικά κλειδιά λεξικών, όπως παραδείγματος χάριν τα κλειδιά σκηνών μπορούν να χρησιμοποιηθούν για ανάκληση συγκεκριμένων σχετικών πληροφοριών, όπως το πρώτο και τελευταίο στιγμιότυπο της. Μέσα από αυτό το κλειδί στιγμιότυπου είναι δυνατή η ανάκλαση μίας πληθώρας σχετικών πληροφοριών κατώτερου επιπέδου όπως το διάνυσμα κατάστασης του ρομπότ, αλλά ακόμα και η αναζήτηση προς τα πίσω, δηλαδή η ανάκλαση της σκηνής που περιέχει αυτό το στιγμιότυπο.

---

<sup>7</sup> <https://www.nuscenes.org/>



Εικόνα 3-12: Αρχιτεκτονική Δεδομένων

Υπάρχουν 23 κλάσεις αντικείμενων στο nuScenes. Για εντοπισμό χρησιμοποιούνται 10 από αυτές.

- Car
- Truck
- construction\_vehicle
- bus
- trailer
- barrier
- motorcycle
- bicycle
- pedestrian
- traffic\_cone

Σύμφωνα με τα πρότυπα του συνόλου [33], επιλέγεται η παρακολούθηση των:

- bicycle – ποδήλατο
- bus – λεωφορείο
- car – αυτοκίνητο
- motorcycle – μοτοσικλέτα
- pedestrian – πεζός
- trailer – τροχοφόρο
- truck – φορτηγό

## **4 ΕΞΑΓΩΓΗ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ ΑΠΟ 3D ΚΑΙ 2D ANIXNEYTEΣ**

---

### **4.1 ΜΟΝΑΔΑ ΕΞΑΓΩΓΗΣ ΓΕΩΜΕΤΡΙΚΩΝ**

#### **ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ ΝΕΦΟΥΣ ΑΠΟ ΤΟΝ ΤΡΙΣΔΙΑΣΤΑΤΟ ANIXNEYTH CENTERPOINT**

Η πλήρης υλοποίηση βρίσκεται σε αυτόνομο προσωπικό αποθετήριο<sup>8</sup> στο GitHub.

Για να εξαχθούν τα χαρακτηριστικά νέφους κάθε στόχου πρέπει πρώτα να γίνει ο εντοπισμός τους ώστε να γνωρίζουμε τα κέντρα τους. Στο κεφάλι του CenterPoint μετασχηματίζεται κάθε κέντρο αντικειμένου από το σύστημα αναφοράς του LiDAR, πίσω στον χάρτη χαρακτηριστικών και εξάγεται μία περιοχή 3\*3 γύρω του μαζί με τα αντίστοιχα χαρακτηριστικά της, διάστασης 512.

$$x_s = \frac{x_c - x_{min}}{s_x \cdot v_x}, \quad y_s = \frac{y_c - y_{min}}{s_y \cdot v_y} \quad (4.1)$$

Όπου  $x_c, y_c$  κέντρο αντικειμένου στο σύστημα αναφοράς του LiDAR,  $x_{min}, y_{min}$  εύρος νέφους σημείων,  $v_x, v_y$  μέγεθος κυψέλης και  $s_x, s_y$  βήμα χάρτη χαρακτηριστικών.

Με έμπνευση τον τρόπο που το Mask R-CNN ευθυγραμμίζει τα χαρακτηριστικά πολλαπλών επιπέδων, χρησιμοποιείται η μέθοδος διγραμμικής παρεμβολής για βελτιωμένη ακρίβεια στον μετασχηματισμό των διακριτών σημείων του χάρτη χαρακτηριστικών από τα κέντρα των εντοπισμών.

Η υλοποίηση ανταποκρίνεται σε όλα τα μεγέθη κυψέλης, περιορίζεται στην κεφαλή Center\_Head. Όσο μικρότερη η κυψέλη, τόσο πιο λεπτομερής ο χάρτης χαρακτηριστικών (128\*128 για κυψέλη όγκου 0.1 m<sup>3</sup> και 180\*180 για κυψέλη όγκου 0.075 m<sup>3</sup>) τόσο πιο ακριβής η εξαγόμενη περιοχή. Η εξαγωγή χαρακτηριστικών μπορεί να γίνει και στο ακριβώς επόμενο επίπεδο του δικτύου εξήντα τεσσάρων καναλιών. Τα τρισδιάστατα διανύσματα κατάστασης αποθηκεύονται με το νέφος χαρακτηριστικών τους, για κάθε στιγμιότυπο κάθε σκηνής σε ένα λεξικό.

---

<sup>8</sup> [https://github.com/DimSpathoulas/Point\\_Cloud\\_Feature\\_Extractor.git](https://github.com/DimSpathoulas/Point_Cloud_Feature_Extractor.git)

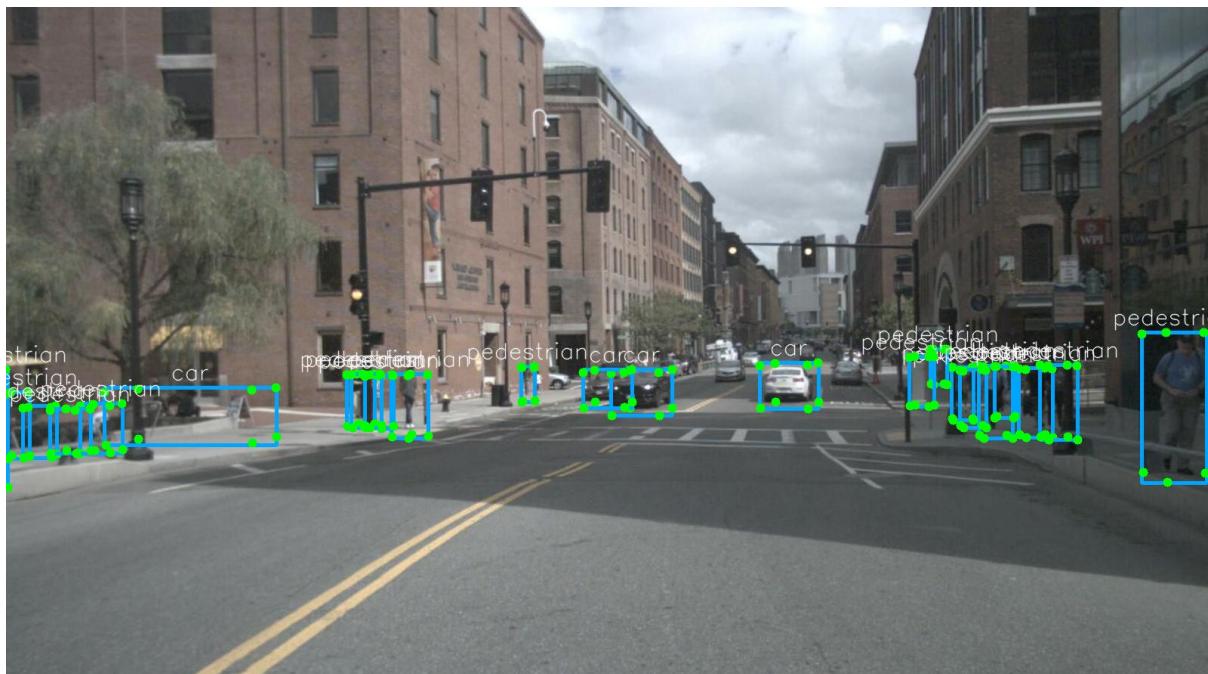
## 4.2 ΜΟΝΑΔΑ ΕΞΑΓΩΓΗΣ ΕΜΦΑΝΙΣΙΑΚΩΝ

### ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ ΕΙΚΟΝΟΣΤΟΙΧΕΙΩΝ ΑΠΟ ΤΟΝ ΔΙΣΔΙΑΣΤΑΤΟ ANIXNEYTH MASK R-CNN

#### 4.2.1 ΠΡΟΒΟΛΗ 3D ANTIKEIMENΩΝ ΣΤΗΝ ΑΝΤΙΣΤΟΙΧΗ ΚΑΜΕΡΑ

Για την εξαγωγή διανυσμάτων χαρακτηριστικών μέσω του Mask R-CNN για κάθε αντικείμενο πρέπει πρώτα να γίνει η προβολή τους πάνω στο επίπεδο της αντίστοιχης κάμερας που αποτυπώνονται, από τις έξι (6) που υπάρχουν και χρησιμοποιούνται. Για κάθε κάμερα, με ωρολογιακή φορά ξεκινώντας από την μπροστά, ελέγχεται κάθε αντικείμενο ξεχωριστά εάν εμφανίζεται στο επίπεδο της.

Για κάθε κάμερα, αρχικά κατασκευάζονται τα τρισδιάστατα πλαίσια κάθε αντικείμενου τοποθετημένα στον χώρο με βάση το πλαίσιο του LiDAR. Στην συνέχεια μεταφέρονται στο ρομποκεντρικό σύστημα αναφοράς και μετά στο παγκόσμιο. Από εκεί τα αντικείμενα μεταφέρονται στο σύστημα της κάμερας και προβάλλονται με τον πίνακα προβολής. Κάθε αντικείμενο ελέγχεται εάν βρίσκεται μπροστά από την κάμερα (αποβολή αντικατοπτρισμού του αντικειμένου στην αντισυμμετρική κάμερα από αυτήν που προβάλλεται πραγματικά) και εάν μπορεί να κατασκευαστεί ένα δισδιάστατο πλαίσιο το οποίο να βρίσκεται στα όρια του μεγέθους της εικόνας (εικονοστοιχεία).



Εικόνα 4-1: Προβαλλόμενα πλαίσια εντοπισμών CenterPoint στην αντίστοιχη κάμερα (κατώφλι αποδοχής 0.03)



Εικόνα 4-2: Προβαλλόμενα πλαίσια εντοπισμών στην αντίστοιχη κάμερα (κατώφλι εντοπισμού 0.81)

Είναι πιθανό ένα αντικείμενο να εμφανίζεται στο πεδίο όρασης παραπάνω από μίας κάμερας (πχ λεωφορείο κοντά στο Α.Ο.). Για την αντιμετώπιση αυτού του φαινομένου, γίνεται έλεγχος των κέντρων των στόχων της προηγούμενης κάμερας με της τωρινής, εάν δύο κέντρα είναι ίδια (δηλαδή είναι το ίδιο αντικείμενο) κρατείται το διάνυσμα χαρακτηριστικών οπού αντιστοιχεί στην μεγαλύτερη επιφάνεια εκ των δύο προβολών. Κάθε αντικείμενο κωδικοποιείται επίσης με ένα διάνυσμα 6 αντιστοίχισης θέσεων (μία για κάθε κάμερα) δυαδικής λογικής, το οποίο υποδεικνύει σε ποια κάμερα το αντικείμενο έχει προβληθεί (one-hot vector).

#### 4.2.2 ΕΞΑΓΩΓΗ ΔΙΑΝΥΣΜΑΤΟΣ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ ΑΠΟ MASK R-CNN

Κάθε εικόνα μαζί με τα δισδιάστατα πλαίσια στόχων της εισάγεται στο προ-εκπαιδευμένο Mask R-CNN. Σκοπός είναι για κάθε μία από τις προβολές να εξαχθεί το αντίστοιχο διάνυσμα χαρακτηριστικών της πάνω στην εικόνα από το Mask R-CNN.

Οι προβολές εισάγονται ως περιοχές ενδιαφέροντος στον χάρτη χαρακτηριστικών της πυραμίδας χαρακτηριστικών (FPN). Το FPN όπως έχει αναφερθεί, είναι εκπαιδευμένο να περιμένει 1000 προτεινόμενες περιοχές, έτσι κατασκευάζεται δυναμικά ένας πίνακας  $1000 \times 4$  θέσεων με τις  $N$  πρώτες να είναι οι πραγματικές προβολές οι οποίες στην συνέχεια φορμάρονται (mold) για να ταιριάζουν στον μετασχηματισμό που αναμένει το μοντέλο σε εκείνο το επίπεδο, και οι άλλες γεμίζονται με μηδενικά. Τα τελικό εξαγόμενο διανύσματα

$\mathbb{R}^{N \times (1024)}$  βρίσκονται στο επίπεδο-στρώμα ακριβώς πριν την ταξινόμηση και πλαισίωση.

### 4.3 ΤΕΛΙΚΟ ΔΙΑΝΥΣΜΑ ΠΛΗΡΟΦΟΡΙΑΣ

Η υλοποίηση επιστρέφει τα διανύσματα ως την τελική έξοδο του δικτύου, μειώνοντας έτσι τον χρόνος επεξεργασίας κάθε εικόνας, καθώς η τμηματοποίηση, η πλαισίωση και η ταξινόμηση δεν ολοκληρώνονται, ωστόσο τα αναγκαία προηγούμενα βήματα γίνονται κανονικά. Ένα σημάδι της ορθής υλοποίησης του αλγορίθμου είναι ότι μόνον αυτά τα πρώτα  $N$  χαρακτηριστικά έχουν αριθμούς ενώ τα άλλα είναι γεμάτα μηδενικά.

Η έξοδος της μονάδας περιέχει το διάνυσμα κατάστασης (σχετικό του παγκόσμιου χάρτη), το διάνυσμα γεωμετρικών χαρακτηριστικών, το διάνυσμα εμφανισιακών χαρακτηριστικών και τον κωδικοποιητή κάμερας, χωρισμένα για κάθε κλάση αντικειμένου για κάθε στιγμιότυπο κάθε σκηνής του επιλεγόμενου υποσυνόλου nuScenes:

$$\boldsymbol{\Omega}_{s,t,c} = \{\mathbf{x}_{s,t,c}, \mathbf{g}_{s,t,c}, \mathbf{a}_{s,t,c}, \mathbf{d}_{s,t,c}, \tau_{s,t,c}\} \quad (4.2)$$

Όπου  $\boldsymbol{\Omega}_{s,t,c}$  το αντικείμενο στην συγκεκριμένη σκηνή  $s$  του στιγμιότυπου  $t$  κλάσης  $c$  με αντίστοιχο διάνυσμα κατάστασης σε παγκόσμιες συντεταγμένες  $\mathbf{x}_{s,t,c} \in \mathbb{R}^{(7)}$ , γεωμετρικά χαρακτηριστικά  $\mathbf{g}_{s,t,c} \in \mathbb{R}^{(512 \times 3 \times 3)}$ , εμφανισιακά χαρακτηριστικά  $\mathbf{a}_{s,t,c} \in \mathbb{R}^{(1024)}$ ,  $\mathbf{d}_{s,t,c} \in \mathbb{R}^{(6)}$  διάνυσμα αποκωδικοποιητή κάμερας και κατώφλι εντοπισμού  $\tau_{s,t,c} \in \mathbb{R}$ . Η υλοποίηση βρίσκεται σε αυτόνομο προσωπικό αποθετήριο<sup>9</sup>.

---

<sup>9</sup> [https://github.com/DimSpathoulas/2D\\_FEATURE\\_EXTRACTOR.git](https://github.com/DimSpathoulas/2D_FEATURE_EXTRACTOR.git)

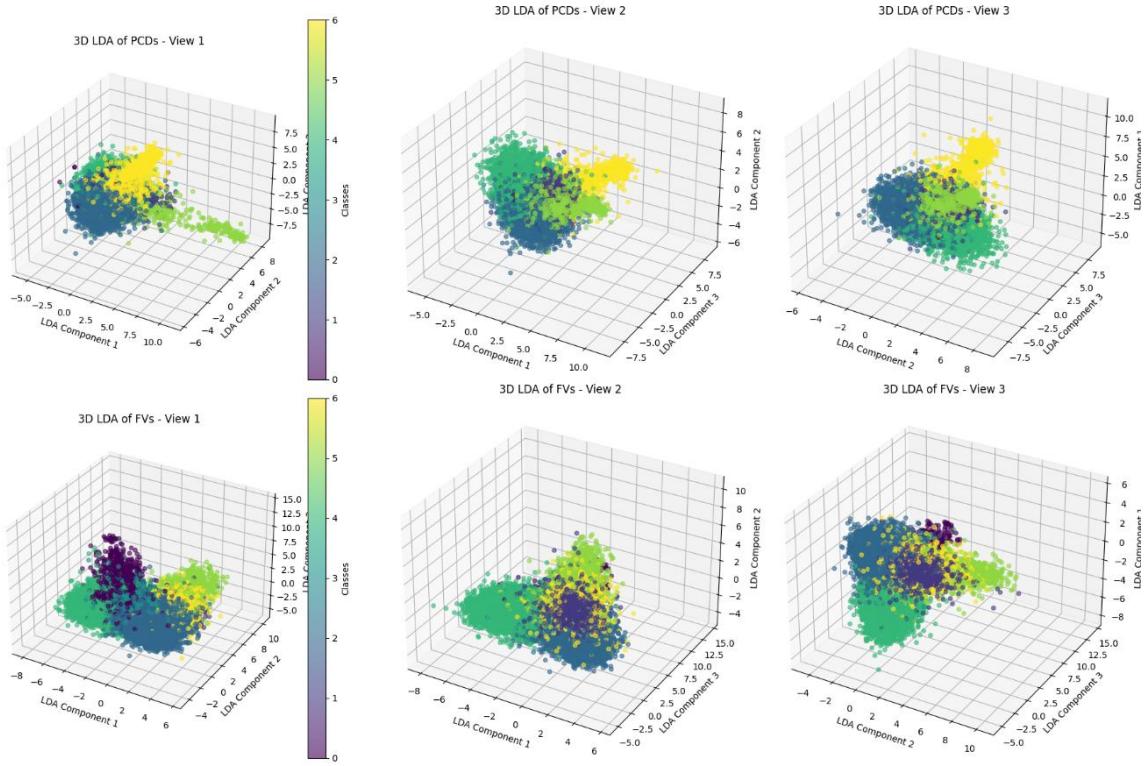
## **5 ΜΑΘΗΣΗ ΠΟΛΛΑΠΛΩΝ ΣΥΛΛΟΓΩΝ ΣΤΑ ΓΕΩΜΕΤΡΙΚΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΤΟΥ ΤΡΙΣΔΙΑΣΤΑΤΟΥ ΑΝΙΧΝΕΥΤΗ CENTERPOINT ΚΑΙ ΣΤΑ ΕΜΦΑΝΙΣΙΑΚΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΤΟΥ ΔΙΣΔΙΑΣΤΑΤΟΥ ΑΝΙΧΝΕΥΤΗ MASK R-CNN**

---

### **5.1 ΓΡΑΜΜΙΚΗ ΑΝΑΛΥΣΗ ΔΙΑΚΡΙΣΗΣ**

Μέσω της Γραμμική Ανάλυσης Διάκρισης (Linear Discriminant Analysis, LDA) επιβεβαιώνεται ότι τα εξαγόμενα χαρακτηριστικά και από τους χάρτες χαρακτηριστικών είναι ιδανικά για ταξινόμηση. Τα δεδομένα είναι θορυβώδη, καθώς δεν πραγματοποιείται κάποια επιπλέον επεξεργασία με τα διαθέσιμα πραγματικά δεδομένα (groundtruths), έτσι ώστε ο LDA να αντικατοπτρίζει όσο το δυνατό καλύτερα την διάκριση των δεδομένων που αντιλαμβάνεται ο ανιχνευτής. Η κωδικοποίηση είναι:

0. bicycle
1. bus
2. car
3. motorcycle
4. pedestrian
5. trailer
6. truck



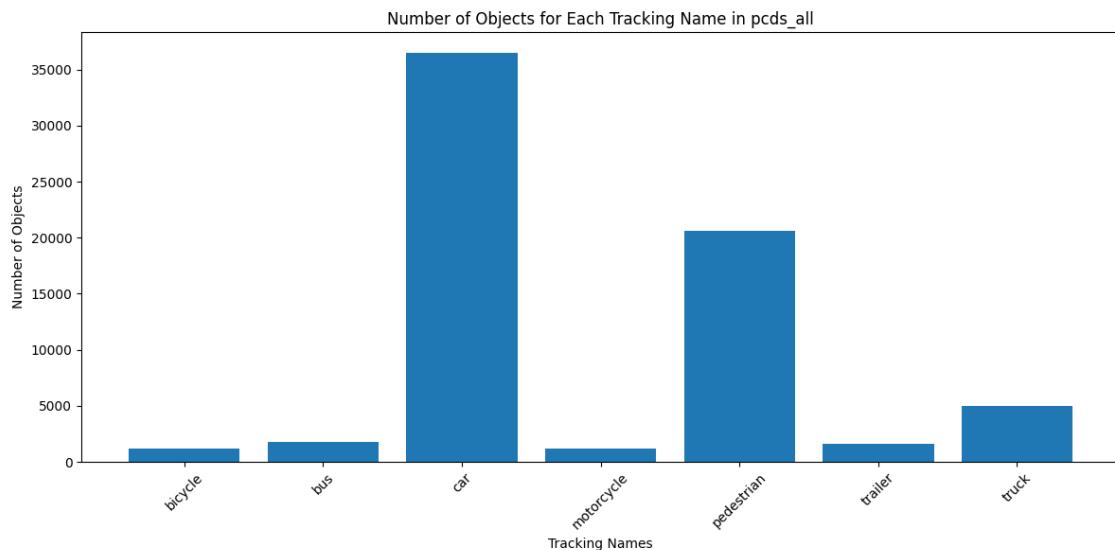
Εικόνα 5-1: Γραμμική ανάλυση διάκρισης στα εμφανισιακά και γεωμετρικά χαρακτηριστικά εντοπισμών (κατώφλι αποδοχής 0.57)

## 5.2 ΑΠΟΜΟΝΩΜΕΝΕΣ ΠΟΛΛΑΠΛΟΤΗΤΕΣ

Έχοντας ήδη εξακριβώσει ότι τα χαρακτηριστικά μπορούν να χρησιμοποιηθούν για ταξινόμηση παρακάτω, γίνεται μία μελέτη πάνω στην μάθηση των εκάστοτε γεωμετριών και ιδιόμορφών ως απομονωμένες οντότητες. Κατά βάση χρησιμοποιούνται οι τεχνικές PCA, t-SNE, UMAP και SVD. Εξετάζονται μέσα από το σύνολο των 140 (4 τεχνικές \* 7 κλάσεις \* 5 καταστάσεις) φωτογραφιών οι σημαντικότερες και πιο διακριτές για σύγκριση και επεξήγηση. Τα χαρακτηριστικά εξάγονται από χάρτη μεγέθους 180\*180 δηλαδή κυψελίδας όγκου 0.075 m<sup>3</sup> στο σύνολο δεδομένων συμπερασμού. Οι υπερπαράμετροι της UMAP είναι 20 γείτονες για μεγαλύτερη διατήρηση τοπικών δομών, ελάχιστη απόσταση 0.1 για αποφυγή βραχυκυκλώματος – συσσωμάτωσης και cosine μετρική απόστασης μεταξύ οντοτήτων. Όλες οι αναλύσεις έγιναν στο υποσύνολο συμπερασμού λόγω όγκου.

### 5.2.1 ΚΑΤΩΦΛΙ ΑΠΟΔΟΧΗΣ

Αρχικά οι πολλαπλότητες κατασκευάζονται μόνο μέσω ενός κατωφλίου αποδοχής της εμπιστοσύνης του ανιχνευτή. Στην Εικόνα 5-2 φαίνεται η σημαντικότατη ανισορροπία στις κλάσεις του συνόλου δεδομένων.

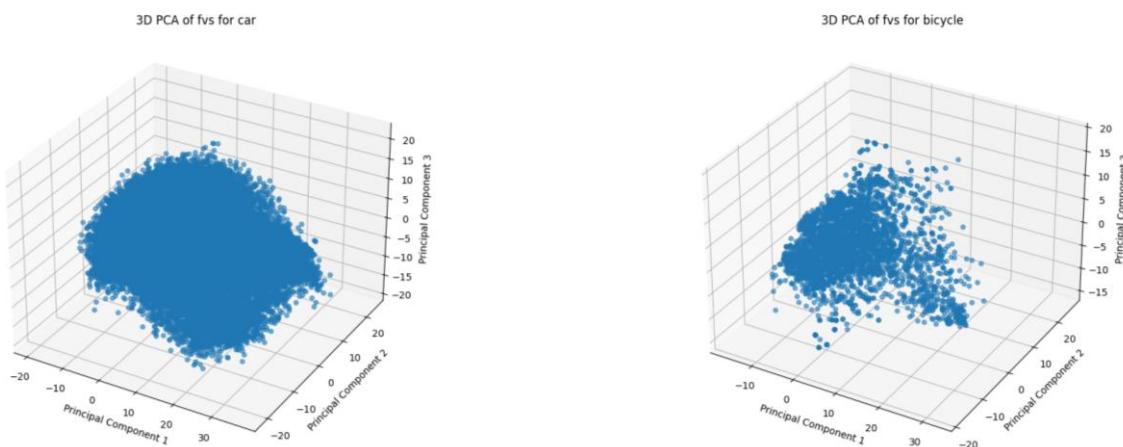


Εικόνα 5-2: Πλήθος εντοπισμών (κατώφλι αποδοχής 0.57)

Η πληθώρα των τιμών που υπάρχουν σε κάθε χαρακτηριστικό στο σύνολο τους είναι μηδενικά και πολύ μικρές τιμές κοντά του, τα χαρακτηριστικά θα μπορούσαν να θεωρηθούν αραιά.

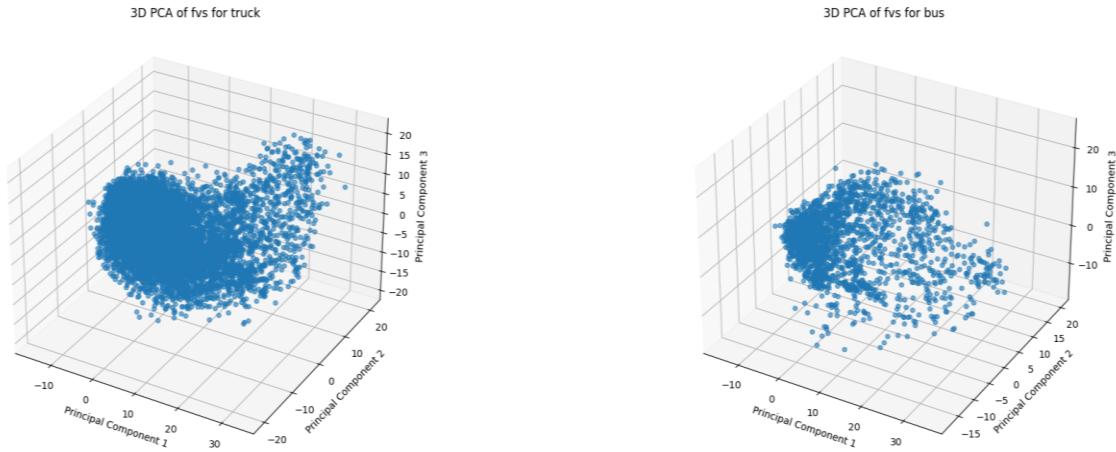
### 5.2.1.1 ΑΝΑΛΥΣΗ ΚΥΡΙΩΝ ΣΥΝΙΣΤΩΣΩΝ

#### 5.2.1.1.1 ΕΜΦΑΝΙΣΙΑΚΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ



Εικόνα 5-3: Τρισδιάστατη αναπαράσταση εμφανισιακών χαρακτηριστικών μέσω ανάλυσης κύριων συνιστωσών (αριστερά αυτοκίνητα, δεξιά ποδήλατα)

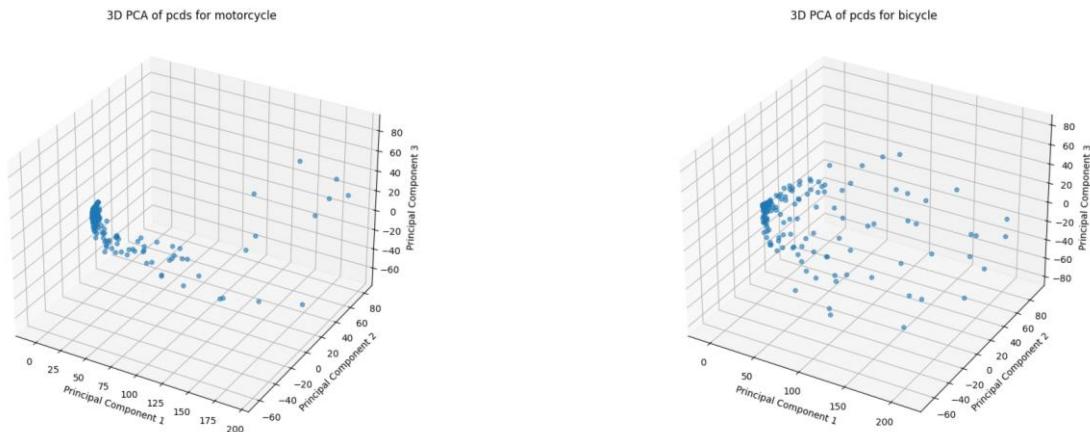
Και τα δύο μοτίβα φαίνονται αρκετά πυκνά, η κλάση αυτοκινήτου είναι περισσότερο πυκνή, αφού έχει περισσότερα αντικείμενα. Και οι δύο πολλαπλότητες αν και αρκετά συμπαγής φαίνεται να εμφανίζουν μία ομοιομορφία (Εικόνα 5-3). Οι μη γραμμικές μέθοδοι μπορεί να είναι καλύτερες στο να την αποτυπώσουν.



Εικόνα 5-4: Τρισδιάστατη αναπαράσταση εμφανισιακών χαρακτηριστικών μέσω ανάλυσης κύριων συνιστωσών (αριστερά φορτηγά, δεξιά λεωφορεία)

Στην κλάση φορτηγό εμφανίζεται ήδη μία ιδιαίτερη μορφολογία, ωστόσο δεν ισχύει το ίδιο για το λεωφορείο, γιατί το σύνολο δεδομένων αποτελείται από πιο διακριτά φορτηγά παρόμοια μεταξύ τους (Εικόνα 5-4).

#### 5.2.1.1.2 ΓΕΩΜΕΤΡΙΚΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ

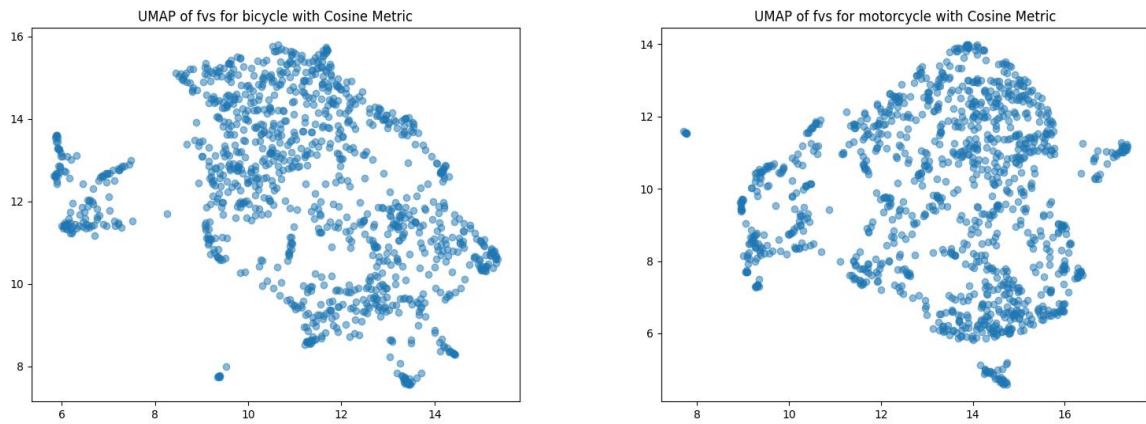


Εικόνα 5-5: Τρισδιάστατη αναπαράσταση γεωμετρικών χαρακτηριστικών μέσω ανάλυσης κύριων συνιστωσών (αριστερά μοτοσικλέτες, δεξιά ποδήλατα)

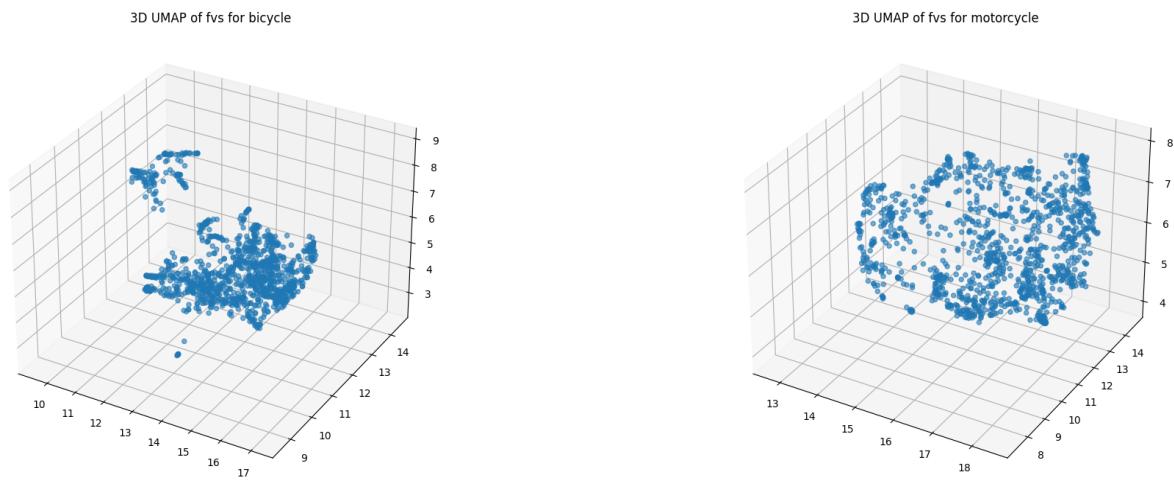
Στην Εικόνα 5-5 φαίνεται η έντονη ανισορροπία και η αραιότητα των κλάσεων μηχανής και ποδηλάτου, ακόμα και έτσι φαίνεται και οι δύο κλάσεις να εμφανίζουν ομοιότητες.

#### 5.2.1.2 UMAP

Παρακάτω (Εικόνα 5-6 και Εικόνα 5-7) εμφανίζεται η αποτύπωση των εμφανισιακών χαρακτηριστικών των ποδηλάτων και των μοτοσυκλετών σε 2d και 3d. Κατά βάση η κατανομή των σημείων για τις μοτοσυκλέτες φαίνεται να είναι πιο ομαδοποιημένη υποδηλώνοντας ότι υπάρχουν πιο πολλές εμφανισιακές ομοιότητες μεταξύ τους από ό,τι στα ποδήλατα.

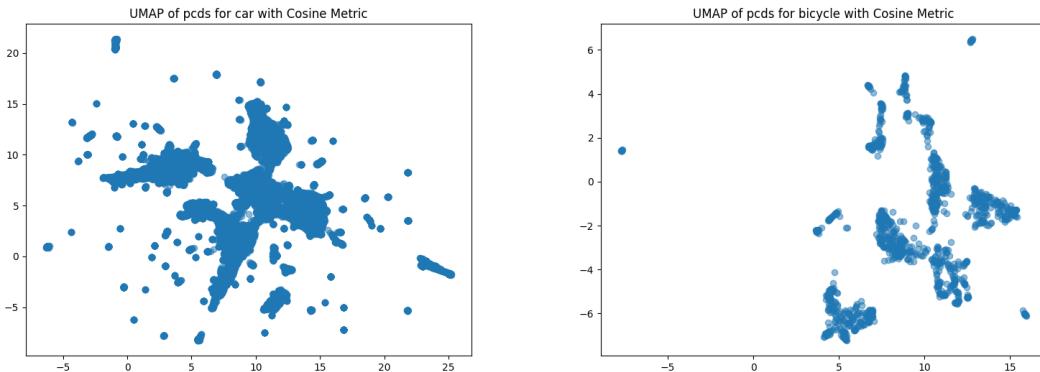


Εικόνα 5-6: Δισδιάστατή αποτύπωση εμφανισιακών χαρακτηριστικών μέσω UMAP (αριστερά ποδήλατα, δεξιά μοτοσικλέτες)



Εικόνα 5-7: Τρισδιάστατη αποτύπωση εμφανισιακών χαρακτηριστικών μέσω UMAP (αριστερά ποδήλατα, δεξιά μοτοσικλέτες)

Στην Εικόνα 5-8 φαίνεται ότι τα γεωμετρικά χαρακτηριστικά των ποδηλάτων είναι διάσπαρτα παρόλα αυτά αρκετά διακριτές περιοχές εμφανίζονται οι οποίες υποδεικνύουν τις σημαντικότερες αυτές γεωμετρίες. Αντίθετα τα αυτοκίνητα τα οποία έχουν πάρα πολλά περισσότερα σημεία είναι πιο εμφανή η κύρια διάσταση του. Υπάρχουν ξεκάθαρες γεμάτες μπλε περιοχές ακόμα και εάν είναι σημαντικά πλατιές με παρακλάδια και απομονωμένες περιοχές οι οποίες το πιθανότερο να είναι έντονα διαφορετικά μοντέλα οχημάτων.

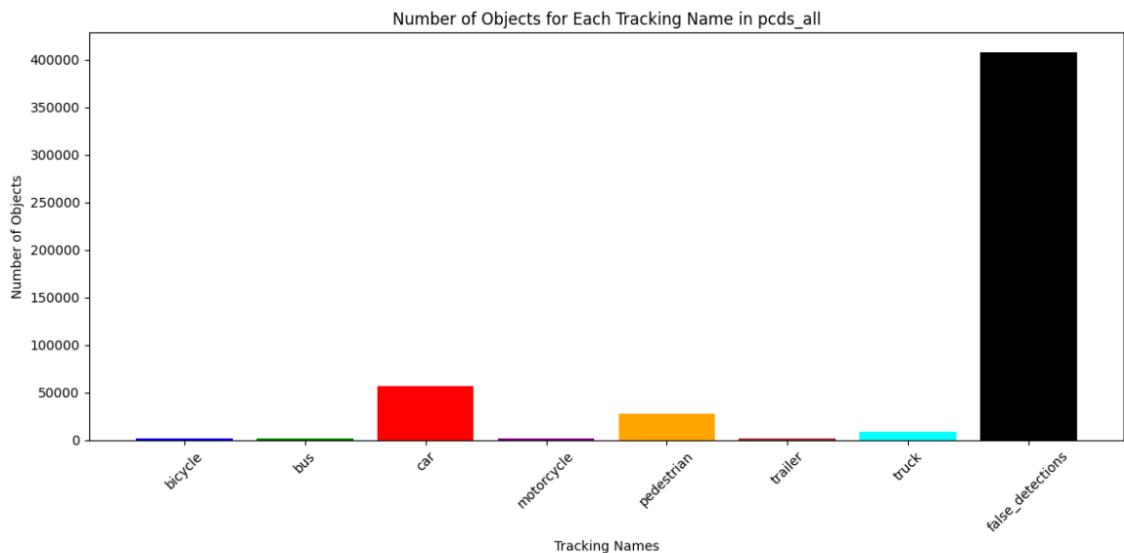


Εικόνα 5-8: Δισδιάστατή αποτύπωση γεωμετρικών χαρακτηριστικών μέσω UMAP (αριστερά αυτοκίνητα, δεξιά ποδήλατα)

### 5.2.2 ΑΠΟΘΟΡΥΒΟΠΟΙΗΣΗ

Εν συνεχεία τα δεδομένα αποθορυβοποιούνται συγκρίνοντας τα με τα groundtruths, το κατώφλι αποδοχής θέτεται ως 0.03 ώστε να διατηρούνται όλα τα ίχνη. Για κάθε στιγμιότυπο συγκρίνεται κάθε εντοπισμός με την αντίστοιχη groundtruth κλάση του με άπληστο τρόπο (ελάχιστη απόσταση), έτσι ώστε κάθε εντοπισμός να αντιστοιχείτε σε ένα groundtruth μόνο μία φορά. Αυτό κυρίως βοηθάει στο φιλτράρισμα της κλάσης πεζών, καθώς εκτός των πιθανών και αναπόφευκτων συστάδων, με την επιλογή κατωφλίου 2 μέτρων μπορεί να υπάρχουν 2 και 3 εντοπισμοί μέσα στο όριο ενός groundtruth κάτι που σε κλάσεις με πιο μεγάλη γεωμετρία είναι πιο δύσκολο.

Η ανισορροπία των κλάσεων είναι ακόμα πιο έντονη (Εικόνα 5-9), πλησιάζοντας την πραγματική κατανομή του συνόλου δεδομένων, παράλληλα εμφανίζεται και ο συνολικός αριθμός λανθασμένων εντοπισμών.



Εικόνα 5-9: Πλήθος αντικειμένων (κατώφλι αποδοχής 0.03)

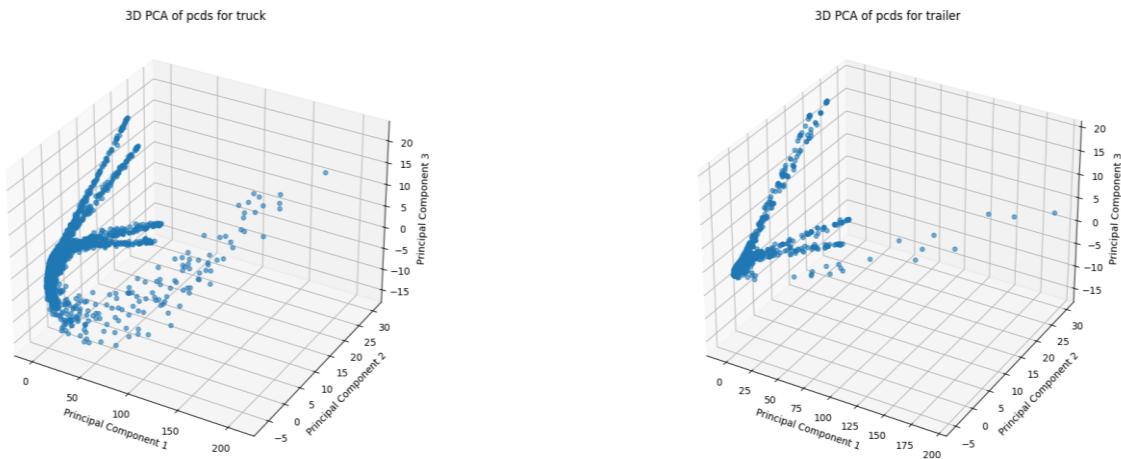
Αριθμός Σκηνών: 150 (6019 στιγμιότυπα)

- Αριθμός Ποδηλάτων: 1869
- Αριθμός Λεωφορείων: 2074
- Αριθμός Αυτοκινήτων: 57269
- Αριθμός Μοτοσικλετών: 1849
- Αριθμός Πεζών: 28373
- Αριθμός Τροχοφόρων: 2125
- Αριθμός Φορτηγών: 9346
- Αριθμός Ψευδών Ανιχνεύσεων: 408163

### 5.2.2.1 ΑΝΑΛΥΣΗ ΚΥΡΙΩΝ ΣΥΝΙΣΤΩΣΩΝ

#### 5.2.2.1.1 ΓΕΩΜΕΤΡΙΚΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ

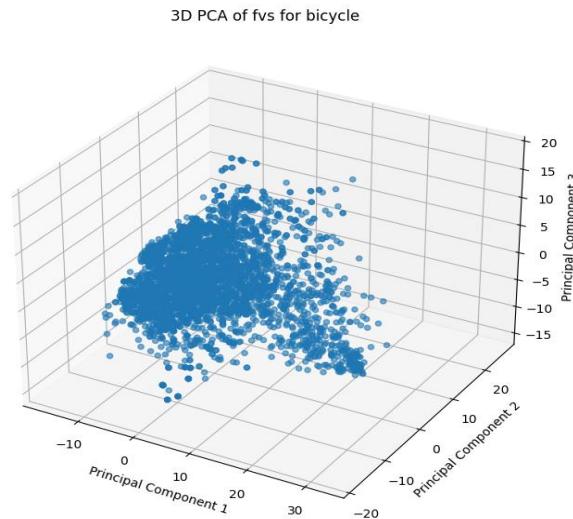
Στην Εικόνα 5-10 φαίνεται μια ξεκάθαρη τάση στα γεωμετρικά χαρακτηριστικά των φορτηγών και τροχοφόρων αφού και τα δύο είναι αρκετά παρόμοια. Αρκετές φορές ο ανιχνευτής μπερδεύει τις δύο κλάσεις μεταξύ τους κάτι που έχει γίνει αντιληπτό κατά την προβολή των εντοπισμών για εξαγωγή εμφανισιακών χαρακτηριστικών. Παρατηρείται ότι υπάρχουν τουλάχιστον 5 διακριτές κύριες συνιστώσες, η μία εκ των οποίων φαίνεται αρκετά χαοτική, κάτι που υποδεικνύει ότι υπάρχουν τουλάχιστον 5 διακριτές γεωμετρίες φορτηγών και τροχοφόρων, κάτι που ενισχύεται και παρακάτω.



Εικόνα 5-10: Τρισδιάστατη αναπαράσταση γεωμετρικών χαρακτηριστικών μέσω ανάλυσης κύριων συνιστωσών (αριστερά φορτηγά, δεξιά τροχοφόρα)

#### 5.2.2.1.2 ΕΜΦΑΝΙΣΙΑΚΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ

Τα εμφανισιακά χαρακτηριστικά των ποδηλάτων μετά από την ανάλυση κύριων συνιστωσών δείχνουν μία μορφολογία αλλά ακόμα δεν υπάρχει κάποια ξεκάθαρη γεωμετρία (Εικόνα 5-11).



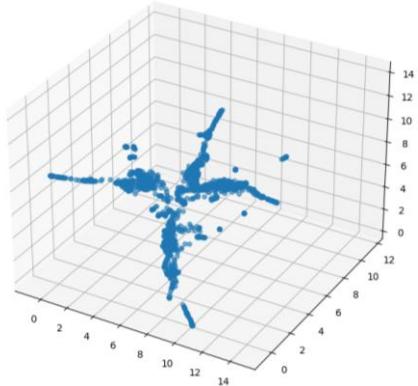
Εικόνα 5-11: Τρισδιάστατη αναπαράσταση εμφανισιακών χαρακτηριστικών ποδηλάτων μέσω ανάλυσης κύριων συνιστωσών

#### 5.2.2.2 UMAP

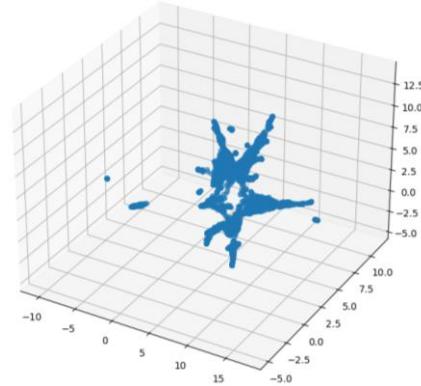
Στις εικόνες παρακάτω εμφανίζονται οι πολλαπλότητες των γεωμετρικών χαρακτηριστικών των κλάσεων. Όλα όσα έχουν συζητηθεί προηγουμένως ενισχύονται καθώς υπάρχουν κύριοι άξονες σε όλες τις κλάσεις υποδηλώνοντας διαφορετικά αλλά διακριτά είδη ίδιων αντικειμένων, υπάρχει ορατή διαφοροποίηση μεταξύ των κλάσεων αλλά παράλληλα ακολουθείτε ένα μοτίβο μεταξύ τους καθώς ο ανιχνευτής θέλει και προσπαθεί να συμπιέσει τις πληροφορίες όσο καλύτερα γίνεται. Τελικά αν και μπορεί να φανεί ότι το ποδήλατο και η

μοτοσυκλέτα θα έπρεπε να μοιάζουν ακόμα περισσότερο, η μοτοσυκλέτα σε φαίνεται να πλησιάζει πιο πολύ την μορφολογία των αυτοκινήτων και το ποδήλατο να είναι κάτι μοναδικό στην μορφολογία του. Ισως εξαιτίας του τρόπου υπερδειγματοληψίας του εντοπιστή αλλά και της ίδιας της κοίλης (τρύπιας) πολλαπλότητας των ποδηλάτων καθώς αποτελούνται από σκελετό και δεν είναι όσο πυκνό το νέφος τους ακόμα και στο κέντρο του εντοπισμού τους, αντίθετα με τις μοτοσυκλέτες οι οποίες αποτελούνται από πολλά ακόμα κομμάτια στο σασί τους.

3D UMAP of pcds for bus

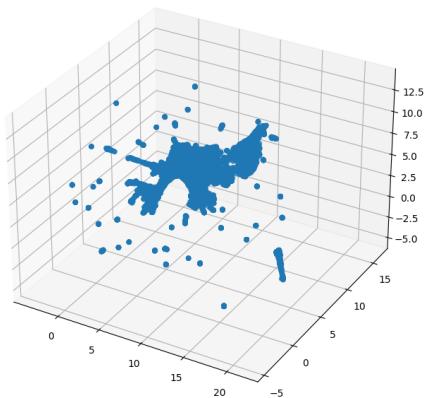


3D UMAP of pcds for truck

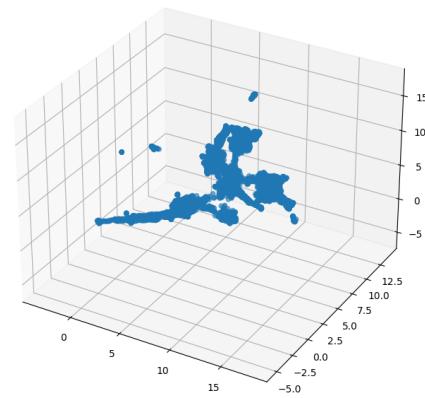


Εικόνα 5-12: Τρισδιάστατη αποτύπωση γεωμετρικών χαρακτηριστικών μέσω UMAP (αριστερά λεωφορεία, δεξιά φορτηγά)

3D UMAP of pcds for car

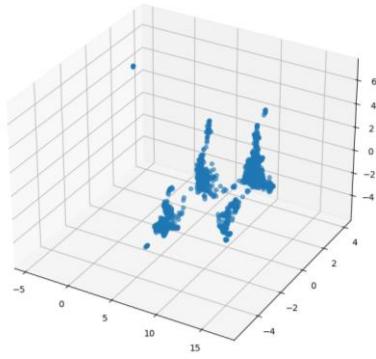


3D UMAP of pcds for pedestrian

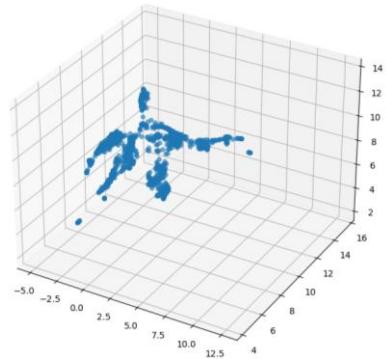


Εικόνα 5-13: Τρισδιάστατη αποτύπωση γεωμετρικών χαρακτηριστικών μέσω UMAP (αριστερά αυτοκίνητα, δεξιά πεζοί)

3D UMAP of pcds for bicycle



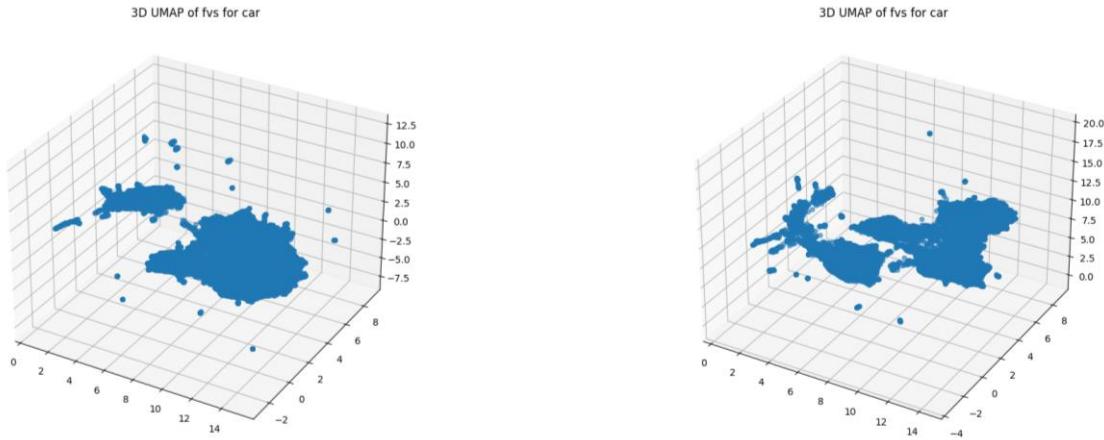
3D UMAP of pcds for motorcycle



Εικόνα 5-14: Τρισδιάστατη αποτύπωση γεωμετρικών χαρακτηριστικών μέσω UMAP (αριστερά ποδήλατα, δεξιά μοτοσικλέτες)

### 5.2.3 ΣΥΓΚΡΙΣΗ ΜΕΤΑΞΥ ΚΑΤΩΦΛΙΟΥ ΚΑΙ ΑΠΟΘΟΡΥΒΟΠΟΙΗΣΗΣ

Δεξιά στην Εικόνα 5-15, βρίσκεται η πολλαπλότητα μέσω φιλτραρίσματος και αριστερά μέσω κατωφλίου αποδοχής. Και οι δύο είναι παρόμοιες, εάν το κατώφλι αποδοχής μειωνόταν ακόμα περισσότερο πιθανών οι πολλαπλότητες να έμοιαζαν ακόμα περισσότερο μεταξύ τους. Φαίνεται ότι διάφορα μεταξύ τους μέρη συνδέθηκαν με το φιλτράρισμα που μόνο του το κατώφλι εντοπισμών άνω του 57% δεν μπόρεσε να αντιληφθεί πλήρως, διαχωρίζοντας την πολλαπλότητα σε δύο σώματα. Το γεγονός αυτό είναι ίσως λίγο πιο περίπλοκο στην εξήγηση του από την πιθανότερή απάντηση η οποία είναι απλώς ο ανιχνευτής να δυσκολεύεται να εντοπίσει συγκεκριμένες καταστάσεις αυτοκινήτων κάνοντας τον λιγότερο βέβαιο για τον εντοπισμό τους. Όπως αναφέρθηκε τα εμφανισιακά χαρακτηριστικά έρχονται μόνο από την προβολή των εντοπισμών του CenterPoint, έτσι καταλήγουμε στο συμπέρασμα ότι ο CenterPoint μπορεί να υστερεί να εντοπίζει αντικείμενα με υψηλή βεβαιότητα, κάτι που ισχύει, καθώς ο αριθμός εντοπισμών αλλάζει δραματικά μειώνοντας το κατώφλι αποδοχής εξαγωγής και διατήρησης των ίδιων.

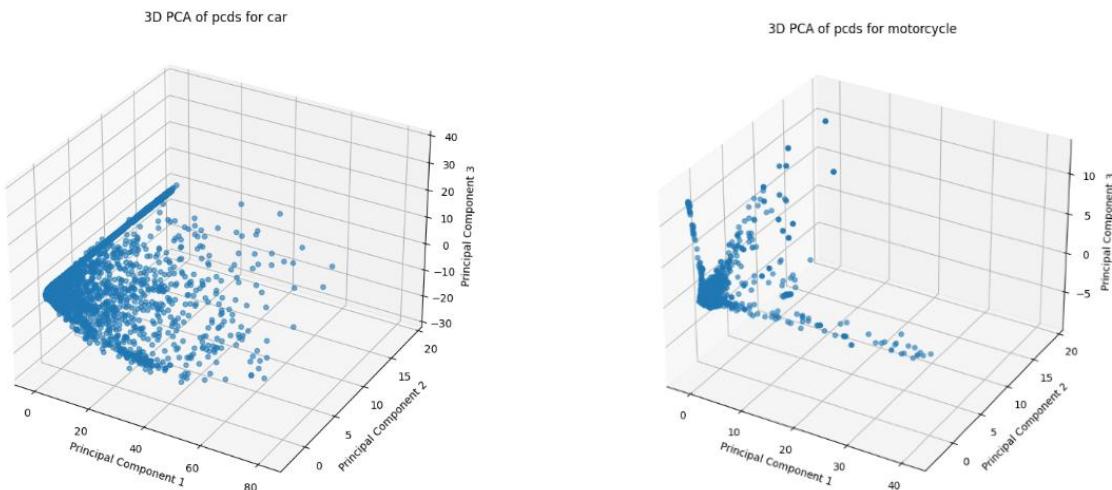


Εικόνα 5-15: Σύγκριση τρισδιάστατης αποτύπωσης εμφανισιακών χαρακτηριστικών μέσω UMAP για τα αυτοκίνητα

#### 5.2.4 ΣΥΓΚΡΙΣΗ ΜΕΤΑΞΥ ΚΥΨΕΛΩΝ

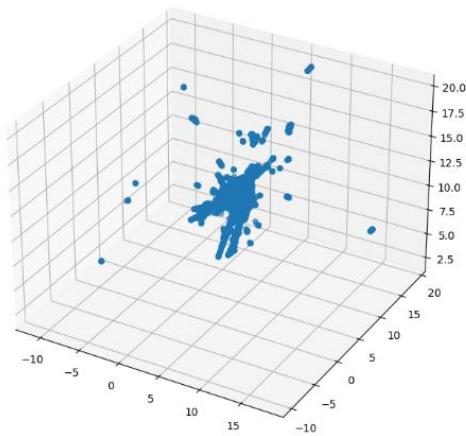
Παρακάτω (Εικόνα 5-16, Εικόνα 5-17) εμφανίζονται και γίνεται μία σύγκριση μεταξύ των χαρακτηριστικών που εξάγονται από τον χάρτη χαρακτηριστικών κυψέλης  $0.075 \text{ m}^3$  και κυψέλης  $0.1 \text{ m}^3$ . Ο πρώτος χάρτης είναι μεγέθους  $180*180$  και ο δεύτερος  $128*128$ . Όλες οι υπερπαράμετροι της UMAP παραμένουν ίδιες ώστε να μην υπάρχει αλλοίωση στην σύγκριση. Τα δεδομένα αυτή τη φορά είναι φιλτραρισμένα και έχουν κατώφλι αποδοχής το 0.57.

Συμπεραίνεται ότι ο ανιχνευτής έχει μάθει διαφορετικές αναπαραστάσεις για τα γεωμετρικά χαρακτηριστικά, έτσι καταλήγουμε ότι το μέγεθος της κυψέλης αλλάζει αρκετά σημαντικά την μορφολογία των πολλαπλοτήτων.

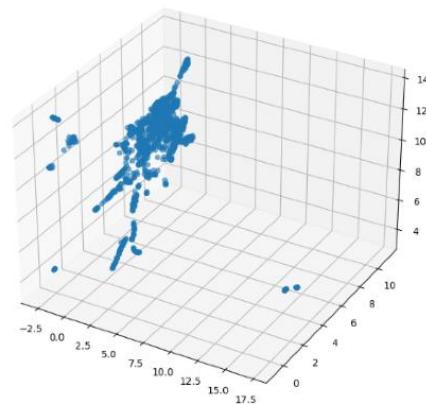


Εικόνα 5-16: Σύγκριση γεωμετρικών χαρακτηριστικών διαφορετικής κυψέλης μέσω PCA για τα αυτοκίνητα (αριστερά  $0.1$ , δεξιά  $0.075$ )

3D UMAP of pcds for truck



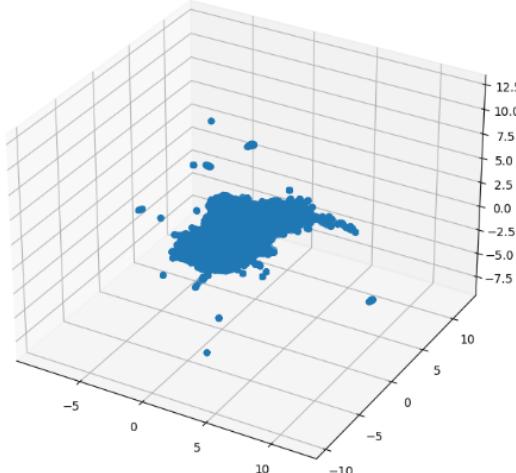
3D UMAP of pcds for trailer



Εικόνα 5-17: Σύγκριση γεωμετρικών χαρακτηριστικών διαφορετικής κυψέλης μέσω UMAP (αριστερά 0.1 και φορτηγά, δεξιά 0.075 και τροχοφόρα)

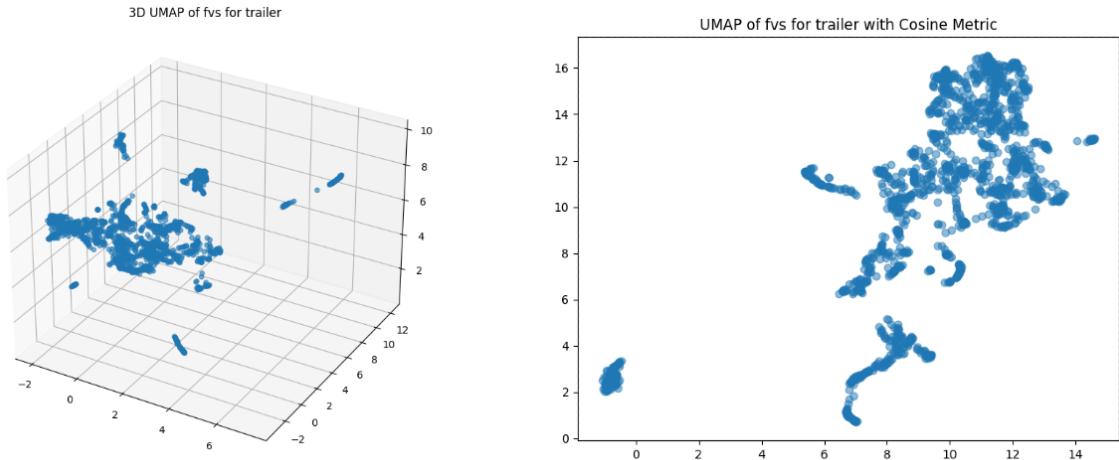
Για περεταίρω έλεγχο των αποτελεσμάτων ελέγχεται και η μορφολογία των πολλαπλοτήτων από τα εξαγόμενα εμφανισιακά χαρακτηριστικά. Παραμένει η ίδια (Εικόνα 5-18), σημειώνοντας ξανά ότι τα δεδομένα εισαγωγής έχουν κατώφλι αποδοχής 57% και έχουν φιλτραριστεί.

3D UMAP of fvs for car



Εικόνα 5-18: Έλεγχος και επαλήθευση μορφολογίας εμφανισιακών χαρακτηριστικών μέσω UMAP από διαφορετικά μεγέθη κυψέλης (φιλτράρισμα και κατώφλι ανίχνευσης 0.57)

Τέλος εμφανίζονται οι πολλαπλότητες των τροχοφόρων σε 2d και 3d αναπαράσταση (Εικόνα 5-19). Είναι εμφανές ότι είναι αρκετά παρόμοιες ωστόσο προέρχονται από δυο διαφορετικά σενάρια προβολών και εξαγωγής.



Εικόνα 5-19: Τρισδιάστατη και δισδιάστατη αναπαράσταση εμφανισιακών χαρακτηριστικών μέσω UMAP τροχοφόρων διαφορετικών σεναρίων εντοπισμού

### 5.2.5 ΧΡΗΣΗ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ

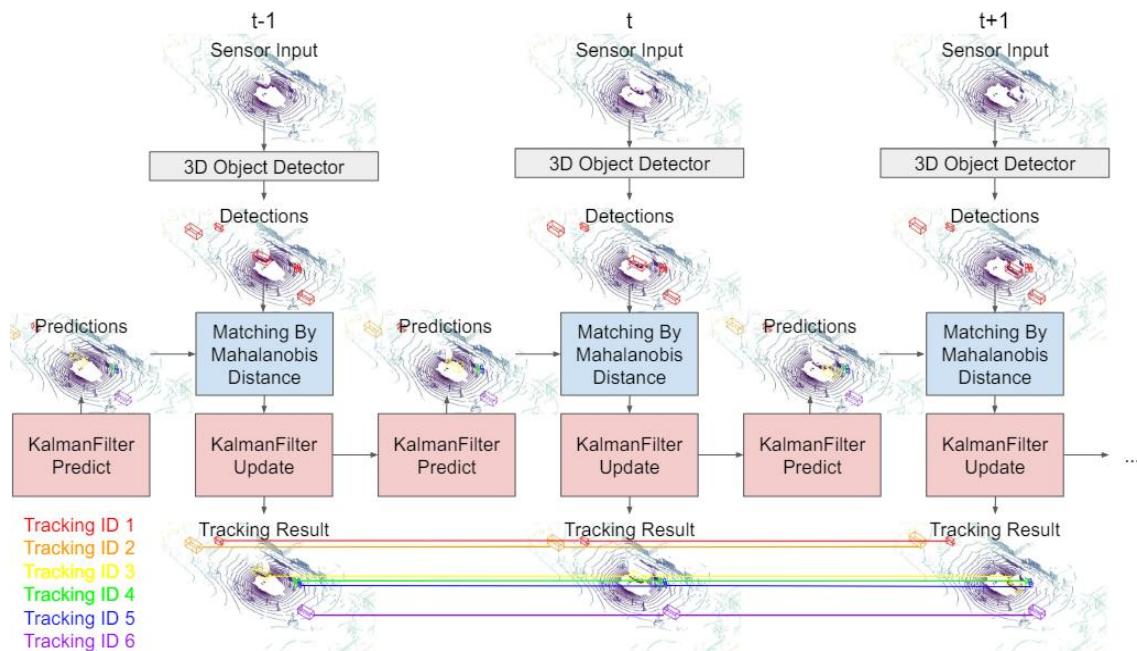
Συμπεραίνεται ότι υπάρχει συνοχή στα δεδομένα και είναι πιθανόν μέσω περαιτέρω επεξεργασιών να υπολογιστούν χωροχρονικές σχέσεις μεταξύ στις πολλαπλότητες τους ως εσωτερικές αποστάσεις. Με αυτόν τον τρόπο καταλήγουμε ότι τα χαρακτηριστικά μπορούν να χρησιμοποιηθούν για να ενισχύσουν επόμενες μονάδες πάνω στην αυτόνομη πλοιόγηση όπως την παρακολούθηση.

# 6 ΒΑΘΙΑ ΣΥΓΧΩΝΕΥΣΗ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ ΓΙΑ ΤΗΝ ΤΡΙΣΔΙΑΣΤΑΣΤΗ ΠΟΛΥΤΡΟΠΙΚΗ ΠΑΡΑΚΟΛΟΥΘΗΣΗ ΠΟΛΛΑΠΛΩΝ ΣΤΟΧΩΝ

## 6.1 ΠΙΘΑΝΟΚΡΑΤΙΚΗ ΤΡΙΣΔΙΑΣΤΑΣΤΗ ΠΑΡΑΚΟΛΟΥΘΗΣΗ ΠΟΛΛΑΠΛΩΝ ΣΤΟΧΩΝ ΓΙΑ ΑΥΤΟΝΟΜΗ ΠΛΟΗΓΗΣΗ

Η στρατηγική [35] βασίζεται σε δεδομένα μόνο από LiDAR για εντοπισμό και παρακολούθηση, χωρίς να υπάρχει συγχώνευση αισθητήρων. Στην αναφορά, ο εντοπισμός γίνεται με το μοντέλο MEGVII [34]. Για την μάθηση του ανιχνευτή έχει ληφθεί υπόψιν η ανισορροπία των κλάσεων.

Η συνολική αρχιτεκτονική της στρατηγικής φαίνεται παρακάτω:



Εικόνα 6-1: Πιθανοκρατική Τρισδιάσταση Παρακολούθηση Πολλαπλών Στόχων

Επαναληπτικά, έχοντας προηγούμενα ίχνη εκτιμάται η κατάσταση τους στον παρόν χρόνο μέσω του βήματος πρόβλεψης του Kalman. Υπολογίζεται η απόσταση Mahanalobis μεταξύ των  $m$  τωρινών ανιχνεύσεων και  $n$  προβλεπόμενων ιχνών, κατασκευάζοντας έναν πίνακα συγγένειας μεγέθους  $[m,n]$ . Μέσω αυτού του πίνακα αποφασίζεται άπληστα (greedy) η αντιστοίχιση στόχου με ίχνους. Για τα ίχνη που έγινε ανάθεση στόχου ενημερώνονται μέσω

του Kalman. Κάθε ίχνος έχει έναν μοναδικό δείκτη από την γένεση του. Πρακτικό είναι ο αλγόριθμος να επαναλαμβάνεται για κάθε κλάση αφού δεν υπάρχει περίπτωση ένα ίχνος να αλλάξει κλάση κατά την πορεία του χρόνου.

### 6.1.1 MONTELOPOIHSIS FILTROU KALMAN

Η κατάσταση για κάθε στόχο στον χρόνο είναι:

$$\mathbf{s}_t = (x, y, z, a, l, w, h, d_x, d_y, d_z, d_a)^T \quad (6.1)$$

Όπου  $x, y, z$  θέση του στόχου στο πλαίσιο αναφοράς,  $a$  ο προσανατολισμός,  $l, w, h$  οι διαστάσεις του πλαισίου και  $d_x, d_y, d_z, d_a$  η μεταβολή των  $x, y, z, a$  μεταξύ δύο διαδοχικών εντοπισμών. Θεωρείται δυναμικό μοντέλο σταθερής γραμμικής και γωνιακής ταχύτητας, με αποτέλεσμα το μοντέλο διαδικασίας (process model) να έχει την μορφή:

$$\begin{aligned} \hat{z}_{t+1} &= z_t + d_{z_t} + q_{z_t}, & \hat{d}_{z_{t+1}} &= d_{z_t} + q_{d_{z_t}} \\ \hat{y}_{t+1} &= y_t + d_{y_t} + q_{y_t}, & \hat{d}_{y_{t+1}} &= d_{y_t} + q_{d_{y_t}} \\ \hat{x}_{t+1} &= x_t + d_{x_t} + q_{x_t}, & \hat{d}_{x_{t+1}} &= d_{x_t} + q_{d_{x_t}} \\ \hat{a}_{t+1} &= a_t + d_{a_t} + q_{a_t}, & \hat{d}_{a_{t+1}} &= d_{a_t} + q_{d_{a_t}} \\ \hat{l}_{t+1} &= l_t, & \hat{w}_{t+1} &= w_t, & \hat{h}_{t+1} &= h_t \end{aligned} \quad (6.2)$$

Όπου  $(q_{x_t}, q_{y_t}, q_{z_t}, q_{a_t})$  και  $(q_{d_{x_t}}, q_{d_{y_t}}, q_{d_{z_t}}, q_{d_{a_t}})$  επιταχύνσεις ως γκαουσιανές τυχαίες μεταβλητές με μέση τιμή 0 και συνδιακύμανση  $\mathbf{Q}$ .

Το βήμα πρόβλεψης γράφεται:

$$\begin{aligned} \hat{\mu}_{t+1} &= \mathbf{A}\mu_t \\ \hat{\Sigma}_{t+1} &= \mathbf{A}\Sigma_t\mathbf{A}^T + \mathbf{Q} \end{aligned} \quad (6.3)$$

Το μοντέλο παρατήρησης έχει την μορφή:

$$\begin{aligned} \hat{\mathbf{o}}_{t+1} &= \mathbf{H}\hat{\mu}_{t+1} \\ \mathbf{S}_{t+1} &= \mathbf{H}\hat{\Sigma}_{t+1}\mathbf{H}^T + \mathbf{R} \end{aligned} \quad (6.4)$$

Με τον πίνακα παρατήρησης  $\mathbf{H}_{7 \times 11} = [\mathbf{I} \ \mathbf{0}]$ .

Το βήμα διόρθωσης μετά την συσχέτιση στόχου με ίχνους διαμορφώνεται ως:

$$\begin{aligned} \mathbf{K}_{t+1} &= \hat{\Sigma}_{t+1}\mathbf{H}^T\mathbf{S}_{t+1}^{-1} \\ \mu_{t+1} &= \hat{\mu}_{t+1} + \mathbf{K}_{t+1}(\mathbf{o}_{t+1} - \hat{\mathbf{o}}_{t+1}) \\ \Sigma_{t+1} &= (\mathbf{I} - \mathbf{K}_{t+1}\mathbf{H})\hat{\Sigma}_{t+1} \end{aligned} \quad (6.5)$$

$\mathbf{o}_{t+1}$  οι παρατηρήσεις και  $\hat{\mathbf{o}}_{t+1}$  οι προβλέψεις.

### 6.1.2 ΚΑΤΑΣΚΕΥΗ ΠΙΝΑΚΩΝ ΣΥΝΔΙΑΚΥΜΑΝΣΗΣ ΜΕΣΩ ΤΟΥ ΣΥΝΟΛΟ

#### ΔΕΔΟΜΕΝΩΝ

Αντί αρχικοποίησης των πινάκων συνδιακύμανσης  $\mathbf{Q}, \mathbf{R}, \Sigma_0$  του μοντέλου εμπειρικά, προτείνεται ένας τρόπος εκτίμησης αυτών με βάση τις διακυμάνσεις των πραγματικών μετρήσεων του συνόλου δεδομένων μάθησης - για να μην υπάρχει προκατάληψη (bias) του θορύβου κατά τον συμπερασμό - και τις μετρήσεις του ανιχνευτή [35]. Για τον πίνακα  $\mathbf{Q}$  χρησιμοποιήθηκαν οι θέσεις των πραγματικών μετρήσεων ανά τρία συνεχόμενα στιγμιότυπα - δεύτερης τάξης παράγωγος, δηλαδή επιτάχυνση - με σκοπό να εξαχθεί η διακύμανση της γραμμικής και γωνιακής ταχύτητας.

$$\begin{aligned} Q_{xx} &= \text{Var}\left(\left(x_{t+1}^{[m]} - x_t^{[m]}\right) - \left(x_t^{[m]} - x_{t-1}^{[m]}\right)\right) \\ Q_{yy} &= \text{Var}\left(\left(y_{t+1}^{[m]} - y_t^{[m]}\right) - \left(y_t^{[m]} - y_{t-1}^{[m]}\right)\right) \\ Q_{zz} &= \text{Var}\left(\left(z_{t+1}^{[m]} - z_t^{[m]}\right) - \left(z_t^{[m]} - z_{t-1}^{[m]}\right)\right) \\ Q_{aa} &= \text{Var}\left(\left(a_{t+1}^{[m]} - a_t^{[m]}\right) - \left(a_t^{[m]} - a_{t-1}^{[m]}\right)\right) \end{aligned} \quad (6.6)$$

Ο πίνακας  $\mathbf{R}$  κατασκευάστηκε μέσω των κέντρων της εξόδου του εντοπιστή και τις αντίστοιχες πραγματικές τιμές, δηλαδή η απόκλιση μεταξύ τους. Ο πίνακας  $\Sigma_0$  είναι ίσος με τον  $\mathbf{R}$ .

$$\begin{aligned} R_{xx} &= \text{Var}\left(D_{x_t}^{[k]} - G_{x_t}^{[k]}\right) \\ R_{yy} &= \text{Var}\left(D_{y_t}^{[k]} - G_{y_t}^{[k]}\right) \\ R_{zz} &= \text{Var}\left(D_{z_t}^{[k]} - G_{z_t}^{[k]}\right) \\ R_{aa} &= \text{Var}\left(D_{a_t}^{[k]} - G_{a_t}^{[k]}\right) \\ R_{ll} &= \text{Var}\left(D_{l_t}^{[k]} - G_{l_t}^{[k]}\right) \\ R_{ww} &= \text{Var}\left(D_{w_t}^{[k]} - G_{w_t}^{[k]}\right) \\ R_{hh} &= \text{Var}\left(D_{h_t}^{[k]} - G_{h_t}^{[k]}\right) \end{aligned} \quad (6.7)$$

Με:

$$D_t^{[k]} = \left(D_{x_t}^{[k]}, D_{y_t}^{[k]}, D_{z_t}^{[k]}, D_{a_t}^{[k]}, D_{l_t}^{[k]}, D_{w_t}^{[k]}, D_{h_t}^{[k]}\right) \quad (6.8)$$

$$G_t^{[k]} = \left(G_{x_t}^{[k]}, G_{y_t}^{[k]}, G_{z_t}^{[k]}, G_{a_t}^{[k]}, G_{l_t}^{[k]}, G_{w_t}^{[k]}, G_{h_t}^{[k]}\right) \quad (6.9)$$

$(D_t^{[k]}, G_t^{[k]})$  το ζεύγος  $k$  εντοπισμού και αληθινών μετρήσεων στον χρόνο  $t$ .

Χρησιμοποιώντας τον CenterPoint λήφθηκαν υπόψιν εκ νέου αυτοί οι υπολογισμοί. Η ολική διαφορά στα αποτελέσματα είναι σημαντικότατη (σελ. 80).

### 6.1.3 ΣΥΣΧΕΤΙΣΗ ΚΑΙ ΑΝΤΙΣΤΟΙΧΗΣΗ

#### 6.1.3.1 ΣΥΣΧΕΤΙΣΗ ΔΕΔΟΜΕΝΩΝ

Ο πίνακας συγγένειας κατασκευάζεται υπολογίζοντας την ψευδοαπόσταση Mahalanobis μεταξύ των παρατηρήσεων και των προβλεπόμενων στόχων. Ο πίνακας καινοτομίας (innovation matrix)  $\mathbf{S}_{t+1}$  αντιπροσωπεύει την μετάβαση και επαγωγικά την αύξηση της αβεβαιότητας της συνδιακύμανσης κατάστασης στο βήμα παρατήρησης. Ως  $\mathbf{o}_{t+1}$  οι παρατηρήσεις και  $\widehat{\mathbf{o}}_{t+1}$  οι προβλέψεις.

$$D = \sqrt{(\mathbf{o}_{t+1} - \widehat{\mathbf{o}}_{t+1})^T \mathbf{S}_{t+1}^{-1} (\mathbf{o}_{t+1} - \widehat{\mathbf{o}}_{t+1})} \quad (6.10)$$

Ο  $D$  ερμηνεύει την ομοιότητα μεταξύ των καταστάσεων του κάθε στόχου σε σχέση με κάθε ίχνος. Όσο πιο κοντά στο 0 τόσο πιο όμοια μεταξύ τους τα αντικείμενα.

#### 6.1.3.2 ΑΝΤΙΣΤΟΙΧΙΣΗ ΔΕΔΟΜΕΝΩΝ

Έστω ο πίνακας συγγένειας  $m$  εντοπισμών και  $n$  ιχνών:

$$\mathbf{D} = \begin{bmatrix} d_{11} & d_{12} & d_{13} & \dots & d_{1n} \\ d_{21} & d_{22} & d_{23} & \dots & d_{2n} \\ d_{31} & d_{32} & d_{33} & \dots & d_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ d_{m1} & d_{m2} & d_{m3} & \dots & d_{mn} \end{bmatrix} \quad (6.11)$$

Ο αλγόριθμος αντιστοίχισης πρώτα επιπεδοποιεί (flatten) τον  $D$  και στην συνέχεια τον ταξινομεί κατά αύξουσα σειρά:

$$\mathbf{a} = [d_{11}, d_{12}, \dots, d_{1n}, d_{21}, d_{22}, \dots, d_{mn}] \quad (6.12)$$

$$\mathbf{a}_\sigma = [a_{\sigma(1)}, a_{\sigma(2)}, \dots, a_{\sigma(mn)}] \quad (6.13)$$

Όπου  $\sigma(i)$  αντιπροσωπεύει την αρχική θέση του  $i$ -οστού μικρότερου στοιχείου στον  $a$ .

Στην συνέχεια κατασκευάζεται ένας πίνακας με τις δυάδες θέσεων του αρχικού πίνακα σε αύξουσα σειρά:

$$(i, j) = \left( \left\lfloor \frac{\sigma(i)}{n} \right\rfloor, \sigma(i) \bmod n \right) \quad (6.14)$$

Άρα:

$$\mathbf{A} = \begin{bmatrix} s_0 \\ s_1 \\ \vdots \\ s_v \end{bmatrix} \quad (6.15)$$

Όπου  $s_0 = (i, j)$  τέτοιο ώστε  $\min(d_{ij})$  στον  $\mathbf{D}$  και  $s_v = (i, j)$  η θέση του  $\max(d_{ij})$ .  $p$  στο  $[0, mn * mn]$  με κάθε  $s_p$  μοναδικό.

Ο  $\mathbf{A}$  περιέχει τους δείκτες όλων των συνδυασμών του  $\mathbf{D}$  κατά αύξουσα σειρά με βάση την συνάρτηση συγγένειας  $d$ . Στη συνέχεια μέσω μίας προσπέλασης των γραμμών του  $\mathbf{A}$  αντιστοιχούνται, κατά 1 προς 1,  $M \leq m$  στόχοι στα  $N \leq n$  ίχνη, για αυτό και η τεχνική αυτή θεωρείται άπληστη (greedy). Η έξοδος είναι μία λίστα με τις θέσεις των συσχετίσεων του πίνακα  $\mathbf{D}$ .

Τα ίχνη και οι στόχοι που δεν υπάρχουν στην λίστα αντιστοίχισης είναι μη σχετιζόμενα – αντιστοιχισμένα (unmatched), ενώ αυτά που υπάρχουν αποτελούν την λίστα αντιστοιχισμένων (matched). Εάν η απόσταση  $d_{ij}$  της συσχέτισης  $s_p(i, j)$  είναι μεγαλύτερη από ένα κατώφλι αποδοχής  $T$ , τότε θεωρείται ότι το ίχνος  $j$  και ο στόχος  $i$  είναι και αυτά μη σχετιζόμενα, καθώς απέχουν πιο μακριά από το αποδεκτό. Έτσι δημιουργούνται δύο λίστες, μία λίστα μη αντιστοιχισμένων ιχνών (unmatched detections) και μία στόχων (unmatched tracks).

#### 6.1.4 ΓΕΝΕΣΗ ΚΑΙ ΘΑΝΑΤΟΣ ΙΧΝΩΝ

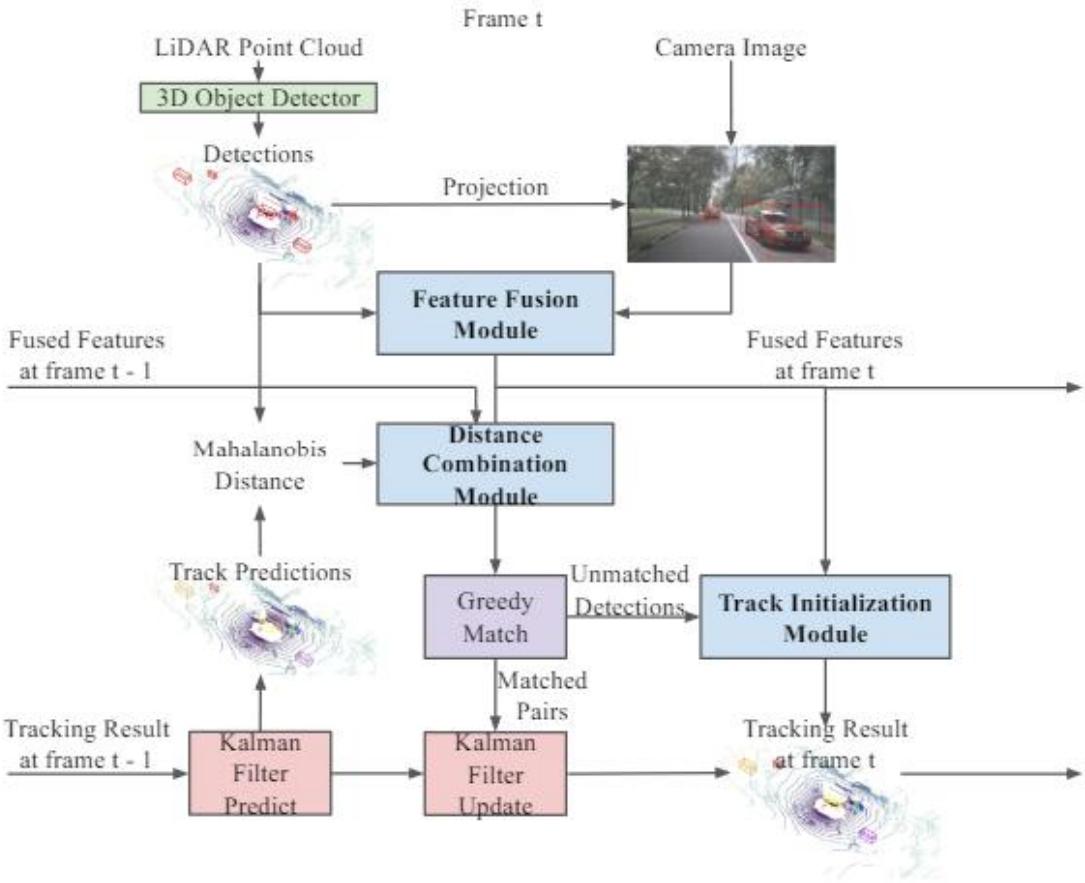
Για τους μη αντιστοιχισμένους στόχους αρχικοποιούνται καινούργια ίχνη ή αρχικοποιούνται ψευδή ίχνη (dummy – temporary) και μετά από έναν αριθμό βημάτων εντάσσονται ως πραγματικά. Όσα ίχνη δεν έχουν αντιστοιχηθεί μετά από αρκετά βήματα διαγράφονται.

## 6.2 ΤΡΙΣΔΙΑΣΤΑΣΤΗ ΠΟΛΥΤΡΟΠΙΚΗ ΠΑΡΑΚΟΛΟΥΘΗΣΗ ΠΟΛΛΑΠΛΩΝ ΣΤΟΧΩΝ ΜΕΣΩ ΓΕΩΜΕΤΡΙΚΩΝ ΚΑΙ ΕΜΦΑΝΙΣΙΑΚΩΝ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ

Η εργασία βασίστηκε σε δημοσίευση του Ινστιτούτου της Toyota (Toyota Research Institute, TRI) και το Stanford [36].

Τα δύο χαρακτηριστικά εισέρχονται στο πρώτο εκπαιδευόμενο δίκτυο, την μονάδα συγχώνευσης χαρακτηριστικών (Feature Fusion Module). Η βαθιά αυτή συγχώνευση έχει ως αποτέλεσμα το τελικό διάνυσμα χαρακτηριστικών να περιέχει και σημασιολογική (semantic-texture) πληροφορία όπου δίνει η κάμερα μέσω του Mask R-CNN και την γεωμορφολογική (morphological-texture) πληροφορία της περιοχής μέσω του CenterPoint από το νέφος σημείων του LiDAR.

Τα συγχωνευμένα χαρακτηριστικά (fused features) κάθε εντοπισμού μαζί με τα συγχωνευμένα χαρακτηριστικά των ιχνών (προηγούμενα συγχωνευμένα χαρακτηριστικά εντοπισμών) εισέρχονται στην δεύτερη μονάδα επεξεργασίας, την μονάδα μήτρας απόστασης συνδυασμών (Distance Combination Module). Ο τελικός αυτός πίνακας περιέχει την ομοιότητα των ενισχυμένων γεωμετρικών και εμφανισιακών χαρακτηριστικών μέσω των συγχωνευμένων χαρακτηριστικών κάθε πιθανής συσχέτισης. Η αντιστοίχιση ακολουθεί την ίδια λογική με την προηγούμενη υλοποίηση, καθώς και η μοντελοποίηση του φίλτρου Kalman μετά από ανακατασκευή των πινάκων συνδιακύμανσης παρατήρησης και κατάστασης με βάση τον CenterPoint. Δεν αναπτύσσεται μονάδα αρχικοποίησης ιχνών (Track Initialization Module).



Εικόνα 6-2: Βαθιά Συγχώνευση Χαρακτηριστικών για την Τρισδιάστατη Πολυτροπική Παρακολούθηση Πολλαπλών Στόχων

### 6.2.1 ΕΞΩΤΕΡΙΚΗ ΑΡΧΙΤΕΚΤΟΝΙΚΗ ΜΟΝΑΔΑΣ ΠΑΡΑΚΟΛΟΥΘΗΣΗΣ

Για κάθε εποχή, καλείται το πρώτο στιγμιότυπο της εκάστοτε σκηνής. Επαναληπτικά, κάθε στιγμιότυπο εισέρχεται στην μονάδα, μέχρι και το τελευταίο, όπου τότε καλείται το πρώτο στιγμιότυπο της επόμενης και ούτω καθεξής έως ότου τελειώσουν οι σκηνές του διαχωρισμού. Κάθε στιγμιότυπο αποτελεί κλειδί για το λεξικό εξόδου του εξαγωγέα χαρακτηριστικών. Σε κάθε στιγμιότυπο υπάρχουν  $N_i$ , με  $i \in [0,7)$ , στόχοι, ο καθένας με τις αντίστοιχες πληροφορίες που τον ακολουθούν. Έτσι, η μονάδα παρακολούθησης καλείται σε μέγιστο αριθμό 7 φορές για κάθε στιγμιότυπο. Εάν η εποχή είναι η τελική συν μία, τότε εισάγονται τα δεδομένα συμπερασμού και το σύστημα περνάει στην φάση συμπερασμού. Τα δίκτυα εκπαιδεύονται για όλες τις κλάσεις μαζί, δηλαδή δεν υπάρχει ξεχωριστό δίκτυο για κάθε κλάση, έτσι ώστε τα βάρη να είναι τέτοια που να γενικεύσουν το μοτίβο που απαιτείται να μάθουν. Όλα αυτά κατασκευάζονται σε μία γενική κλάση (μετά από πολλά πειράματα) η οποία περιέχει όλα τα δίκτυα και όλες τις αντίστοιχες συναρτήσεις τους, μαζί με την ολική μονάδα παρακολούθησης. Αποτέλεσμα είναι μία πολύπλοκη αρχιτεκτονική η οποία ωστόσο μειώνει τον όγκο του

αλγορίθμου (από περίπου 4500 γραμμές σε 1700), καθώς απαιτείται συγκεκριμένα αντικείμενα της (λίστες ιχνών) να είναι δυναμικά, δηλαδή να μπορούν μετά από κάθε σκηνή να διαγράφονται από την μνήμη ενώ σε κάθε στιγμιότυπο να αλλάζουν, παράλληλα με άλλα που είναι στατικά (κατώφλια αποδοχής, εναλλάκτες θέσεων), χωρίς να εμποδίζουν την λειτουργία των νευρωνικών δικτύων και σε επίπεδο μάθησης αλλά και συμπερασμού<sup>10</sup>.

### 6.2.2 ΜΟΝΑΔΑ ΣΥΓΧΩΝΕΥΣΗΣ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ

Σκοπός αυτής της μονάδας είναι η βαθιά συγχώνευση των χαρακτηριστικών νέφους του CenterPoint μαζί με τα χαρακτηριστικά εμφάνισης του Mask R-CNN και του κωδικοποιητή κάμερας κάθε αντικειμένου έτσι ώστε ο τελικός τανυστής να περιέχει την συνδυασμένη πληροφορία τους.

$$\mathbf{F}^{fused} = \mathbf{G}_1(\mathbf{F}^{2D}) + \mathbf{F}^{3D} \quad (6.16)$$

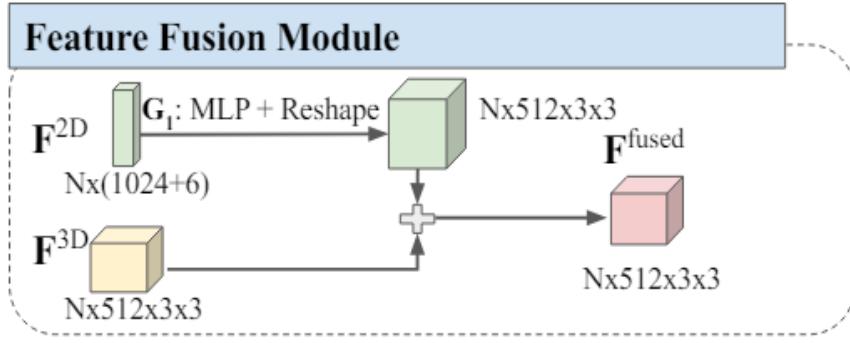
$\mathbf{F}^{2D} \in \mathbb{R}^{N \times (1024+6)}$  τα χαρακτηριστικά Mask R-CNN μαζί με τον κωδικοποιητή κάμερας και  $\mathbf{F}^{3D} \in \mathbb{R}^{N \times 512 \times 3 \times 3}$  τα χαρακτηριστικά του CenterPoint για τα  $N$  αντικείμενα της κλάσης.  $\mathbf{G}_1(\cdot)$  ένα πλήρως συνδεδεμένο πολυστρωματικό δίκτυο (MPL) με κρυμμένο επίπεδο μεγέθους 1536 μία μονάδα γραμμικής ανόρθωσης (Rectified Linear Unit, ReLU) και επίπεδο εξόδου διάστασης 4608, η οποία μετατρέπεται σε τανυστή μεγέθους  $512 \times 3 \times 3$ .  $\mathbf{F}^{fused} \in \mathbb{R}^{N \times 512 \times 3 \times 3}$  τα συγχωνευμένα χαρακτηριστικά των στόχων. Για κάθε αντιστοίχιση αναμιγνύονται με τα αντίστοιχα χαρακτηριστικά των ιχνών τους μέσω ενός απλού παράγοντα.

$$\mathbf{F}_t^{\text{tracks}} = \alpha \mathbf{F}_t^{\text{fused}} + (1 - \alpha) \mathbf{F}_{t-1}^{\text{tracks}} \quad (6.17)$$

Η μονάδα δεν έχει άμεση συνάρτηση σφάλματος, αλλά συνδέεται μαζί με τις υπόλοιπες για την σταδιακή εκπαίδευση της.

---

<sup>10</sup> [https://github.com/DimSpathoulas/GeomApp\\_3MOT.git](https://github.com/DimSpathoulas/GeomApp_3MOT.git)



Εικόνα 6-3: Μονάδα Συγχώνευσης Χαρακτηριστικών

### 6.2.3 ΧΑΡΤΗΣ ΣΥΝΕΝΩΜΕΝΩΝ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ

Από τα  $\mathbf{F}_{det}^{fused} \in \mathbb{R}^{N \times 512 \times 3 \times 3}$  και  $\mathbf{F}_{trk}^{fused} \in \mathbb{R}^{M \times 512 \times 3 \times 3}$  κατασκευάζεται ο πίνακας εισόδου της μονάδας απόστασης χαρακτηριστικών  $\mathbf{F}_{dxt}^{fused} \in \mathbb{R}^{N \times M \times 1024 \times 3 \times 3}$ . Αποτελεί έναν χάρτης διάστασης  $N \times M$  όπου σε κάθε θέση βρίσκονται τα συνενωμένα (concatenated) τους χαρακτηριστικά, δηλαδή κανάλια μεγέθους  $\mathbb{R}^{1024 \times 3 \times 3}$  για κάθε συνδυασμό. Έτσι στην πρώτη γραμμή και στήλη υπάρχουν τα χαρακτηριστικά του πρώτου εντοπισμού συνενωμένα με του πρώτου ίχνους, στην πρώτη γραμμή δεύτερη στήλη, του πρώτου στόχου και δεύτερου ίχνους και ούτω κάθε εξής.

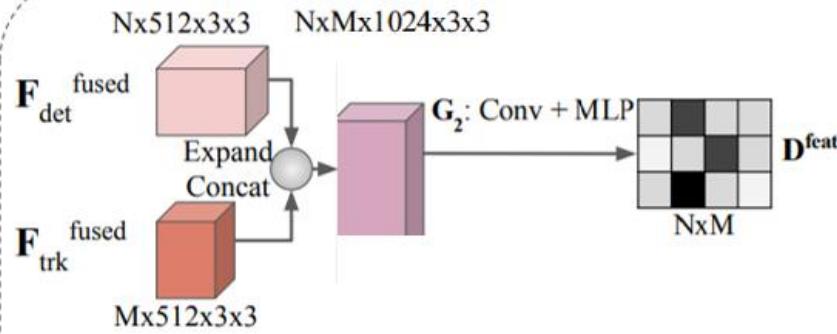
### 6.2.4 MHTPA ΒΑΘΙΑΣ ΑΠΟΣΤΑΣΗΣ ΣΥΝΔΥΑΣΜΩΝ

Ο πίνακας  $\mathbf{D}^{feat} \in \mathbb{R}^{N \times M}$  είναι η έξοδος του πρώτου μέρος της μονάδας συγγένειας συνδυασμών και έχει ως σκοπό να εκφράσει τα συνενωμένα χαρακτηριστικά κάθε στόχου με ίχνους σε αποστάσεις (μία τιμή) στο εύρος  $[0,1]$  όπου 0 η τέλεια αντιστοίχιση.

$$\mathbf{D}^{feat} = \mathbf{G}_2(\mathbf{F}_{det}^{fused}, \mathbf{F}_{trk}^{fused}) \quad (6.18)$$

Το  $\mathbf{G}_2(\cdot)$  αποτελείται από ένα συλλεκτικό επίπεδο, με μέγεθος πυρήνα  $3 \times 3$ , βήματος 1, γεμίσματος 0 και διάστασης εξόδου 256. Η έξοδος εισέρχεται σε ένα πλήρως συνδεδεμένο πολυστρωματικό δίκτυο με συναρτήσεις ενεργοποίησης ReLU και ένα κρυφό επίπεδο διάστασης 128. Το δίκτυο αυτό δεν μπορεί να αντικατασταθεί με μία απλή μετρική όπως η cosine.

## Distance Combination Module



Εικόνα 6-4: Χάρτης Συνενωμένων Χαρακτηριστικών και Μήτρα Βαθιάς Απόστασης Συνδυασμών

Η συνάρτηση σφάλματος του δικτύου είναι:

$$L^{dist} = \text{BCE}(\mathbf{D}^{feat}, \mathbf{K}) \quad (6.19)$$

Ο πίνακας εποπτείας δεικτών αντιστοίχισης (supervising matching indicator matrix)  $\mathbf{K}$  αναπαριστά τις πραγματικές συσχετίσεις μεταξύ των στόχων και των ιχνών με βάση το σύνολο δεδομένων. Έχει ίδιες διαστάσεις με τον πίνακα συγγένειας, με κάθε στοιχείο του να αντιστοιχεί στο ζεύγος εντοπισμού-ίχνους του  $\mathbf{D}^{feat}$ . Πρακτικά είναι μία δυαδική μάσκα.

Κατασκευάζεται λαμβάνοντας υπόψη δύο κριτήρια:

- Εγγύτητα: Η απόσταση μεταξύ του εντοπισμένου αντικειμένου και του αντίστοιχου πραγματικού αντικειμένου, καθώς και η απόσταση μεταξύ του ίχνους και του πραγματικού αντικειμένου από το προηγούμενο στιγμιότυπο, είναι μικρότερη από ένα συγκεκριμένο σταθερό κατώφλι (2 μέτρα).
- Ταυτότητα: Τα δύο αντίστοιχα πραγματικά αντικείμενα (το τρέχον και το προηγούμενο) έχουν το ίδιο συμβολικό στιγμιότυπο (instance token), το οποίο είναι μοναδικό αναγνωριστικό, σταθερό καθ' όλη τη διάρκεια ζωής του πραγματικού αντικειμένου στο σύνολο δεδομένων.

Έτσι, για κάθε ζεύγος εντοπισμού-ίχνους, ο  $\mathbf{K}$  θα έχει την τιμή 0 αν και μόνο αν:

- Η απόσταση μεταξύ του εντοπισμένου αντικειμένου και του αντίστοιχου πραγματικού αντικειμένου είναι μικρότερη από το κατώφλι.
- Η απόσταση μεταξύ του ίχνους και του αντίστοιχου πραγματικού αντικειμένου από το προηγούμενο στιγμιότυπο είναι επίσης μικρότερη από το κατώφλι.

- Το συμβολικό στιγμιότυπο του πραγματικού αντικειμένου που αντιστοιχεί στον τρέχοντα εντοπισμό είναι το ίδιο με το συμβολικό στιγμιότυπο του πραγματικού αντικειμένου που αντιστοιχεί στο ίχνος από το προηγούμενο στιγμιότυπο.

Αν οποιαδήποτε από αυτές τις συνθήκες δεν ικανοποιείται, η αντίστοιχη θέση του πίνακα θα έχει την τιμή 1. Στον  $\mathbf{K}$  μπορεί να υπάρχει το πολύ ένα μηδενικό ανά γραμμή και ανά στήλη του, δηλαδή το πολύ μία αντιστοίχιση για κάθε στόχο και το πολύ μία αντιστοίχιση για κάθε ίχνος. Η μονάδα συνδέεται με την μονάδα συγχώνευσης χαρακτηριστικών έτσι ώστε η ροή των κλίσεων (gradient flow) με βάση την συνάρτηση σφάλματος  $\mathbf{G}_2$  να περνάει στην συνέχεια και στο  $\mathbf{G}_1$ . Στην πραγματικότητα η συνάρτηση σφάλματος θεωρεί το 1 ως ορθό και το 0 ως λάθος, οπότε η μονάδα δεν μαθαίνει πως να υπολογίζει καλές αποστάσεις για τις καλές αντιστοιχίεις, αλλά πως να φτιάχνει όσο δυνατόν χειρότερες ώστε να εξασφαλίσει την μη αντιστοίχιση λανθασμένων συνδυασμών.

Κύριο πρόβλημα αποτελεί το γεγονός ότι ο  $\mathbf{K}$  είναι αραιός πίνακας, πχ στην καλύτερη περίπτωση εάν αποτελείται από 100 στοιχεία δηλαδή 10 στόχοι και 10 ίχνη, ο μέγιστος αριθμός μηδενικών είναι 10. Αυτό το πρόβλημα αντιμετωπίζεται με την τεχνική απώλεια εστίασης ή εστιακή απώλεια (focal loss) και επιβάρυνσης των μηδενικών μέσω μείωσης σημαντικότητας των μη συσχετίσεων.

$$FL(p_t) = -\alpha \cdot (1 - p_t)^{\gamma} \cdot \log(p_t) \quad (6.20)$$

Όπου α πηλίκο μεταξύ πλήθους αρνητικών και θετικών συσχετίσεων του  $\mathbf{K}$ , γ υπερπαράμετρος ελέγχου εστίασης,  $p_t$  πιθανότητες σωστών κλάδων. Στην παρούσα μοντελοποίηση η τιμή 1 είναι η αρνητική κλάση οπότε η εξίσωση ανισορροπίας δεν επιβαρύνει την κλάση 1 όπως είναι η γενική μορφή της συνάρτησης σφάλματος αλλά αντίθετα μειώνει την σημαντικότητα της έτσι ώστε έμμεσα να επιβαρύνει τα δείγματα 0.

### 6.2.5 ΟΛΙΚΟΣ ΣΧΕΔΙΑΣΜΟΣ ΕΚΠΑΙΔΕΥΣΗΣ ΚΑΙ ΣΥΜΠΕΡΑΣΜΟΥ

Για τον εντοπισμό και εξαγωγή νέφους χαρακτηριστικών χρειάζεται περίπου μία ημέρα, για το υπό-σύνολο μάθησης και μισή για το υπό-σύνολο συμπερασμού. Για την εξαγωγή χαρακτηριστικών μέσω Mask R-CNN χρειάζονται δύο ημέρες για το υπό-σύνολο μάθησης και μία για τον συμπερασμού. Και τα δύο συστήματα αποθηκεύονται σε στάδια τα δεδομένα για

βελτιστοποίηση χρήσης πόρων και μείωση συνεχών απαιτήσεων<sup>11</sup>.

Για την εκπαίδευση του G2, G1 χρειάζονται περίπου μία με δύο ώρες, Χρησιμοποιήθηκε ο βελτιστοποιητής (optimizer) Adam με αρχικό ρυθμό μάθησης 0.001 για 10 εποχές με σημείο αναβάθμισης το στιγμιότυπο. Το σύστημα λειτουργεί με δυνατότητα χρήσης cuda.

## 6.2.6 ΑΠΟΤΕΛΕΣΜΑΤΑ

### 6.2.6.1 ΜΕΤΡΙΚΕΣ

Οι μετρικές που χρησιμοποιούνται για την αποδοτικότητα και αποτελεσματικότητα της παρακολούθησης είναι:

**AMOTA** (Average Multi-Object Tracking Accuracy / Μέση Ακρίβεια Παρακολούθησης Πολλαπλών Αντικειμένων): Η συνολική ακρίβεια παρακολούθησης με βάση τη μέση βαθμολογία MOTA σε διαφορετικά επίπεδα ανάκλησης. Όσο υψηλότερη η τιμή, τόσο καλύτερη η απόδοση του συστήματος παρακολούθησης.

**AMOTP** (Average Multi-Object Tracking Precision / Μέση Ακρίβεια Τοποθέτησης Παρακολούθησης Πολλαπλών Αντικειμένων): Μέση ακρίβεια παρακολούθησης μέσω της απόστασης μεταξύ των προβλεπόμενων και των πραγματικών θέσεων των αντικειμένων. Χαμηλότερες τιμές υποδηλώνουν μεγαλύτερη ακρίβεια τοποθέτησης.

**RECALL** (Ανάκληση): Ποσοστό των αντικειμένων που ανιχνεύθηκαν σωστά σε σχέση με τα πραγματικά δεδομένα. Υψηλότερη ανάκληση σημαίνει λιγότερες χαμένες ανιχνεύσεις (false negatives).

**MOTAR** (Multi-Object Tracking Accuracy with Removal / Ακρίβεια Παρακολούθησης Πολλαπλών Αντικειμένων με Αφαίρεση Ταυτοποίησης): Παρόμοιο με το MOTA, αλλά υπολογίζεται αφού αφαιρεθούν οι αλλαγές ταυτότητας (identity switches). Όσο υψηλότερη η τιμή, τόσο μεγαλύτερη η ακρίβεια μετά την αφαίρεση των σφαλμάτων ταυτοποίησης.

**GT** (Ground Truth / Πραγματικά Δεδομένα): Συνολικός αριθμός των πραγματικών διαδρομών (trajectories) που περιλαμβάνονται στα δεδομένα αναφοράς.

**MOTA** (Multi-Object Tracking Accuracy / Ακρίβεια Παρακολούθησης Πολλαπλών Αντικειμένων): Ακρίβεια παρακολούθησης λαμβάνοντας υπόψη ψευδείς ανιχνεύσεις (FP), χαμένες ανιχνεύσεις (FN) και αλλαγές ταυτότητας (IDS). Όσο υψηλότερη η τιμή, τόσο

<sup>11</sup> [https://github.com/DimSpathoulas/GeomApp\\_3MOT.git](https://github.com/DimSpathoulas/GeomApp_3MOT.git)

καλύτερη η συνολική ακρίβεια.

**MOTP** (Multi-Object Tracking Precision / Ακρίβεια Τοποθέτησης Πολλαπλών Αντικειμένων): Ακρίβεια τοποθέτησης των αντικειμένων σε σχέση με τα πραγματικά δεδομένα, μέσω της απόστασης ανάμεσα σε προβλεπόμενες και πραγματικές θέσεις. Χαμηλότερες τιμές δείχνουν καλύτερη ακρίβεια.

**MT** (Mostly Tracked / Περισσότερο Παρακολουθούμενα): Ο αριθμός των αντικειμένων που παρακολουθούνται για περισσότερο από το 80% της διάρκειας της διαδρομής τους. Υψηλότερες τιμές δείχνουν καλύτερη παρακολούθηση.

**ML** (Mostly Lost / Περισσότερο Χαμένα): Ο αριθμός των αντικειμένων που παρακολουθούνται για λιγότερο από το 20% της διάρκειας της διαδρομής τους. Χαμηλότερες τιμές δείχνουν καλύτερη παρακολούθηση.

**FAF** (False Alarms per Frame / Ψευδείς Συναγερμοί ανά Καρέ): Αριθμός ψευδών ανιχνεύσεων ανά καρέ. Χαμηλότερες τιμές είναι επιθυμητές.

**TP** (True Positives / Αληθείς Θετικές): Αριθμός των σωστά ανιχνευμένων αντικειμένων σε σχέση με τα πραγματικά δεδομένα. Υψηλότερες τιμές είναι καλύτερες.

**FP** (False Positives / Ψευδείς Θετικές): Αριθμός των ψευδών ανιχνεύσεων, δηλαδή αντικείμενα που ανιχνεύθηκαν αλλά δεν υπάρχουν στην πραγματικότητα. Χαμηλότερες τιμές είναι επιθυμητές.

**FN** (False Negatives / Ψευδείς Αρνητικές): Αριθμός των αντικειμένων που υπάρχουν αλλά δεν ανιχνεύθηκαν από το σύστημα. Χαμηλότερες τιμές δείχνουν καλύτερη ανίχνευση.

**IDS** (Identity Switches / Αλλαγές Ταυτότητας): Ο αριθμός των φορών που η ταυτότητα ενός αντικειμένου αλλάζει κατά τη διάρκεια της παρακολούθησης. Χαμηλότερες τιμές υποδηλώνουν καλύτερη συνέπεια στην παρακολούθηση.

**FRAG** (Fragmentations / Κατακερματισμοί): Αριθμός των φορών που μια διαδρομή αντικειμένου διακόπτεται ή χάνεται προσωρινά. Χαμηλότερες τιμές είναι καλύτερες.

**TID** (Track Identity Duration / Διάρκεια Ταυτότητας Παρακολούθησης): Μέση διάρκεια κατά την οποία μια διαδρομή αντικειμένου παρακολουθείται σωστά. Χαμηλότερες τιμές δείχνουν καλύτερη διατήρηση της ταυτότητας.

**LGD** (Longest Gap Duration / Μέγιστη Διάρκεια Χαμένου Αντικειμένου): Το μέγιστο διάστημα που ένα αντικείμενο δεν παρακολουθείται πριν επανασυνδεθεί με το σωστό

αντικείμενο. Χαμηλότερες τιμές είναι επιθυμητές.

#### 6.2.6.2 ΣΥΓΚΡΙΣΗ ΑΠΟΤΕΛΕΣΜΑΤΩΝ ΔΙΑΦΟΡΕΤΙΚΩΝ ΔΙΑΚΥΜΑΝΣΕΩΝ ΒΑΣΗ ΑΝΙΧΝΕΥΤΩΝ

Παρακάτω (Εικόνα 6-5, Εικόνα 6-6) φαίνεται η διαφορά στις μετρικές παρακολούθησης των 7 διαφορετικών κλάσεων και οι συνολικές (Εικόνα 6-7). Είναι ξεκάθαρη η υπεροχή του CenterPoint με αύξηση των μετρικών παρακολούθησης περίπου 10-15% σχεδόν σε όλες τις κλάσεις με την κλάση ποδήλατο να υπάρχει 30% αύξηση. Τα αποτελέσματα μπορούν να δουν ακόμα μεγαλύτερη βελτίωση λαμβάνοντας υπόψιν την σχετική υποδειγματοληψία θορύβου η οποία μπορεί να μειωθεί εάν το κατώφλι εντοπισμών στο υποσύνολο εκπαίδευσης ήταν λιγότερο του 0.57 λόγω υπολογιστικών περιορισμών.

Class	AMOTA	AMOTP	RECALL	MOTAR	GT	MOTA	MOTP	MT	ML	FAF	TP	FP	FN	IDS	FRAG	TID	LGD
Bicycle	0.272	0.965	0.361	0.673	1993	0.243	0.179	27	96	15.2	719	235	1273	1	1	1.93	2.27
Bus	0.741	0.741	0.774	0.857	2112	0.664	0.507	64	23	14.0	1635	233	477	0	20	0.94	1.19
Car	0.735	0.534	0.738	0.851	58317	0.623	0.264	1978	1053	110.3	42660	6340	15290	367	341	0.79	1.08
Motorcycle	0.506	0.857	0.594	0.764	1977	0.449	0.284	42	48	18.6	1163	275	803	11	14	1.62	2.29
Pedestrian	0.755	0.525	0.800	0.824	25423	0.650	0.257	1052	313	78.2	20054	3531	5084	285	171	0.58	0.92
Trailer	0.337	1.227	0.385	0.800	2425	0.307	0.758	32	86	18.7	931	186	1492	2	21	1.75	2.34
Truck	0.580	0.754	0.589	0.764	9650	0.449	0.404	237	238	35.2	5669	1340	3968	13	38	0.95	1.32

Εικόνα 6-5: Αποτελέσματα παρακολούθησης ανά κλάση συνδιακύμανσης ανιχνευτή MEGVII (κατώφλι εντοπισμών εκπαίδευσης και συμπερασμού: 0.01)

Class	AMOTA	AMOTP	RECALL	MOTAR	GT	MOTA	MOTP	MT	ML	FAF	TP	FP	FN	IDS	FRAG	TID	LGD
Bicycle	0.377	0.711	0.478	0.732	1993	0.348	0.214	42	84	16.8	948	254	1040	5	5	1.58	1.77
Bus	0.819	0.603	0.824	0.898	2112	0.738	0.444	69	12	10.9	1736	177	372	4	16	0.83	1.18
Car	0.752	0.545	0.762	0.853	58317	0.643	0.278	2100	916	112.4	43977	6463	13852	488	458	0.68	1.02
Motorcycle	0.468	0.835	0.578	0.740	1977	0.422	0.379	42	46	20.5	1128	293	835	14	15	1.48	2.10
Pedestrian	0.681	0.586	0.736	0.804	25423	0.580	0.335	928	437	81.1	18317	3581	6716	390	197	0.72	1.17
Trailer	0.449	0.985	0.468	0.779	2425	0.363	0.547	39	71	24.7	1130	250	1291	4	5	1.06	1.69
Truck	0.630	0.674	0.659	0.763	9650	0.502	0.366	269	188	40.3	6350	1507	3287	13	36	0.85	1.14

Εικόνα 6-6: Αποτελέσματα παρακολούθησης ανά κλάση συνδιακύμανσης ανιχνευτή CenterPoint (κατώφλι εντοπισμών εκπαίδευσης 0.57 και συμπερασμού 0.03)

Ανιχνευτής	CenterPoint	MEGVII
AMOTA	0.597	0.561
AMOTP	0.706	0.800
RECALL	0.644	0.606

Εικόνα 6-7: Σύγκριση ολικών μετρικών (αριστερά CenterPoint, δεξιά MEGVII)

### 6.2.6.3 ΣΥΓΚΡΙΣΗ ΜΕ ΜΕΘΟΔΟΥΣ ΒΑΣΗΣ

Στην Εικόνα 6-8 και Εικόνα 6-9 φαίνονται οι μετρικές παρακολούθησης των στόχων με την χρήση πίνακα απόστασης Mahalanobis με κατώφλι αποδοχής εντοπισμών 0.57 και κατώφλι συσχέτισης 11 το οποίο υποστηρίζεται από την βιβλιογραφία [35-36], για σύγκριση.

Class	AMOTA	AMOTP	RECALL	MOTAR	GT	MOTA	MOTP	MT	ML	FAF	TP	FP	FN	IDS	FRAG	TID	LGD
Bicycle	0.297	1.284	0.336	0.845	1993	0.282	0.266	23	97	7.3	665	103	1323	5	6	1.88	2.20
Bus	0.658	0.866	0.708	0.912	2112	0.641	0.410	48	21	8.2	1483	130	617	12	22	1.26	2.24
Car	0.491	1.091	0.549	0.962	58317	0.519	0.211	790	1343	20.5	31422	1183	26315	580	563	1.89	2.59
Motorcycle	0.369	1.283	0.435	0.904	1977	0.389	0.254	16	54	6.1	851	82	1117	9	8	2.97	3.68
Pedestrian	0.573	0.960	0.645	0.907	25423	0.573	0.282	616	461	33.9	16051	1485	9017	355	285	1.32	1.89
Trailer	0.370	1.362	0.431	0.834	2425	0.357	0.545	31	72	16.9	1038	172	1379	8	13	1.91	2.57
Truck	0.342	1.359	0.430	0.844	9650	0.358	0.314	85	262	17.3	4096	640	5503	51	71	2.38	3.69

Εικόνα 6-8: Αποτελέσματα ανά κλάση αναφοράς μεθοδολογίας με Mahalanobis (κατώφλι αποδοχής 0.57)

Metric	Value
AMOTA	0.443
AMOTP	1.172
RECALL	0.505
MOTAR	0.887
GT	14556
MOTA	0.445
MOTP	0.326
MT	1609
ML	2310
FAF	15.7
TP	55606
FP	3795
FN	45271
IDS	1020
FRAG	968
TID	1.94
LGD	2.70

Εικόνα 6-9: Ολικά Αποτελέσματα μεθοδολογίας Mahalanobis

Στην Εικόνα 6-10 και Εικόνα 6-11 φαίνονται οι μετρικές μόνο με την χρήση του πίνακα απόστασης χαρακτηριστικών, με κατώφλι αποδοχής συσχέτισης 0.9 και παράγοντα ανάμειξης 0.45. Υπάρχει δυνατότητα ενίσχυσης αυτών των αποτελεσμάτων με πιο λεπτομερής και προσεκτική επιλογή παραμετροποίηση. Συμπερασματικά, το δίκτυο ανταποκρίνεται στον σκοπό, μαθαίνοντας να απορρίπτει λανθασμένες συσχετίσεις σε έναν πάρα πολύ αραιό χώρο, πλησιάζοντας αρκετά κοντά στην μήτρα απόστασης Mahalanobis.

Class	AMOTA	AMOTP	RECALL	MOTAR	GT	MOTA	MOTP	MT	ML	FAF	TP	FP	FN	IDS	FRAG	TID	LGD
Bicycle	0.226	1.350	0.360	0.616	1993	0.220	0.325	23	89	17.9	711	273	1276	6	13	1.87	2.22
Bus	0.617	0.940	0.688	0.899	2112	0.600	0.440	46	21	9.0	1410	142	659	43	37	1.32	2.30
Car	0.410	1.244	0.504	0.936	58317	0.441	0.258	642	1460	30.6	27491	1760	28913	1913	1105	1.91	2.68
Motorcycle	0.330	1.321	0.451	0.709	1977	0.308	0.364	14	51	18.1	858	250	1085	34	39	2.67	3.16
Pedestrian	0.483	1.126	0.602	0.850	25423	0.479	0.360	511	518	49.3	14323	2142	10115	985	658	1.27	1.93
Trailer	0.289	1.460	0.401	0.714	2425	0.273	0.584	24	72	25.9	929	266	1452	44	38	2.23	3.01
Truck	0.310	1.410	0.411	0.807	9650	0.324	0.342	72	275	20.1	3868	745	5683	99	95	2.38	3.72

Εικόνα 6-10: Αποτελέσματα ανά κλάση με Πίνακα Βαθιάς Απόστασης Συνδυασμών (κατώφλι αποδοχής 0.57, παράγοντας ανάμειξης 0.45 και κατώφλι αποδοχής συσχέτισης 0.9)

Metric	Value
AMOTA	0.381
AMOTP	1.264
RECALL	0.488
MOTAR	0.790
GT	14556
MOTA	0.378
MODP	0.382
MT	1332
ML	2486
FAF	24.4
TP	49590
FP	5578
FN	49183
IDS	3124
FRAG	1985
TID	1.95
LGD	2.72

Εικόνα 6-11: Ολικές Μετρικές Παρακολούθησης με Πίνακα Βαθιάς Απόστασης Συνδυασμών

Ο πίνακας χαρακτηριστικών δίνει πολύ καλύτερα αποτελέσματα σε σχέση με την χρήση IoU ή ευκλείδειας απόστασης<sup>12</sup> μεταξύ προβλέψεων και στόχων ως πίνακα συσχέτισης, όπως και φαίνεται στην Εικόνα 6-12. Επαναλαμβάνεται ότι όσο μεγαλύτερο το κατώφλι αποδοχής ανιχνεύσεων τόσο καλύτερες οι μετρικές σε οποιαδήποτε μεθολογία.

<sup>12</sup> <https://www.nuscenes.org/tracking?externalData=all&mapData=all&modalities=Any>

Object Class	AMOTA	AMOTP (m)	RECALL
bicycle	0	1.828	0.194
bus	0.408	1.127	0.43
car	0.278	1.325	0.464
motorcycle	0.081	1.703	0.201
pedestrian	0.141	1.698	0.257
trailer	0.136	1.479	0.263
truck	0.013	1.345	0.124

Εικόνα 6-12: Αποτελέσματα βάσης αναφοράς μεθοδολογίας (κατώφλι αποδοχής 0.01)

#### 6.2.6.4 ΑΝΑΛΥΣΗ ΕΥΑΙΣΘΗΣΙΑΣ ΠΑΡΑΓΟΝΤΑ ΑΝΑΜΕΙΞΗΣ

Αν και ο παράγοντας ανάμειξης είναι ζωτικής σημασίας για το δίκτυο, φαίνεται να μην διαφοροποιεί αισθητά τα αποτελέσματα. Φαινομενικά καλύτερη επιλογή περίπου 0.5 (Εικόνα 6-13). Η σύγκριση έγινε με κατώφλι αποδοχής 0.85.

Παράγοντας	AMOTA	AMOTP	RECALL
0.20	0.341	1.326	0.465
0.30	0.358	1.290	0.474
0.40	0.360	1.304	0.471
0.50	0.375	1.276	0.481
0.75	0.376	1.271	0.478

Εικόνα 6-13: Ανάλυση Ενασθησίας Παράγοντα Ανάμειξης (κατώφλι αποδοχής συσχέτισης 0.85)

#### 6.2.6.5 ΑΝΑΛΥΣΗ ΕΥΑΙΣΘΗΣΙΑΣ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ

Τέλος για ανάλυση σε επίπεδο χαρακτηριστικών συγκρίνονται τα αποτελέσματα για τυχαία χαρακτηριστικά, χαρακτηριστικά μόνο της κάμερας, χαρακτηριστικά μόνο του LiDAR με μήτρα συγγένειας την cosine και με το δίκτυο (Εικόνα 6-14). Όπως και επιβεβαιώνεται από την βιβλιογραφία τα γεωμετρικά χαρακτηριστικά είναι πιο ισχυρά από τα εμφανισιακά [36]. Το δίκτυο υπολογισμού συγγένειας όντως έχει νόημα αφού μόνο η χρήση της ομοιότητάς cosine φέρει σχετικά αδύναμα αποτελέσματα. Όσον αφορά τα αποτελέσματα με τυχαία χαρακτηριστικά εισόδου συμπεραίνεται ότι τα μοντέλα, οι τεχνικές και τα εργαλεία που

χρησιμοποιούνται μπορούν να εντοπίσουν μοτίβα ακόμα και στον ελεγχόμενο θόρυβο.

Χαρακτηριστικά	AMOTA	AMOTP	RECALL
Τυχαία	0.068	1.893	0.102
Εμφανισιακά	0.134	1.628	0.282
Γεωμετρικά (cosine)	0.290	1.405	0.381
Γεωμετρικά	0.351	1.316	0.457

Εικόνα 6-14: Ανάλυση Εναισθησίας Χαρακτηριστικών

## **7 ΑΝΑΛΥΤΙΚΟΣ ΥΠΟΛΟΓΙΣΜΟΣ ΡΙΣΚΟΥ**

### **ΣΥΓΚΡΟΥΣΗΣ ΠΟΛΛΑΠΛΩΝ ΣΤΟΧΩΝ ΓΙΑ ΤΗΝ**

### **ΑΥΤΟΝΟΜΗ ΠΛΟΗΓΗΣΗ**

---

#### **7.1 ΥΠΟΛΟΓΙΣΜΟΣ ΤΡΟΧΙΑΣ ΚΑΙ ΚΑΤΑΣΚΕΥΗ**

#### **ΔΙΕΥΡΥΝΣΗΣ ΠΕΡΙΟΧΗΣ ΑΣΦΑΛΕΙΑΣ**

Το ΑΟ θεωρείται ότι έχει ντετερμινιστική τροχιά (από τον σχεδιαστή τροχιάς), λαμβάνοντας έτσι τις μελλοντικές του καταστάσεις από το σύνολο δεδομένων. Για κάθε εμπόδιο-ίχνος γίνεται πρόβλεψη της επόμενης του θέσης μέσω δυναμικού συστήματος σταθερής γραμμικής ταχύτητας και σταθερού προσανατολισμού, δεδομένου θορύβου διαδικασίας, ο οποίος λαμβάνει υπόψιν την αβεβαιότητα θέσης και ταχύτητας μέσα από τους ήδη υπολογισμένους πίνακες συνδιακύμανσης βάση του ανιχνευτή που επιλέχθηκε. Δεν υπάρχει βήμα διόρθωσης, με αυτόν τον τρόπο η συνδιακύμανση του εμποδίου σταδιακά αυξάνεται.

Για το μοντέλο μετάβασης κατάστασης:

$$\mathbf{F} = \begin{bmatrix} 1 & 0 & t & 0 \\ 0 & 1 & 0 & t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (7.1)$$

Όπου  $t=0.5$  (με βάση τον συγχρονισμό αισθητήρων του συνόλου [33]). Το διάνυσμα κατάστασης ανανεώνεται:

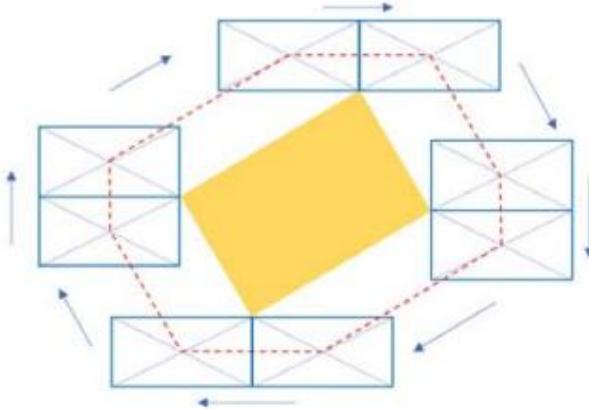
$$\mathbf{x}_t = \mathbf{F} \cdot \mathbf{x}_{t-1} \quad (7.2)$$

Και η αβεβαιότητα αυξάνεται ως:

$$\mathbf{P}_t = \mathbf{F} \cdot \mathbf{P}_{t-1} \cdot \mathbf{F}^\top + \mathbf{Q} \quad (7.3)$$

Όπου  $\mathbf{Q}$  και  $\mathbf{P}_{t=0}$  λαμβάνονται από τον προ-κατασκευασμένο πίνακα συνδιακύμανσης  $\Sigma_0$ .

Μετά από κάθε τέτοια πρόβλεψη, υπολογίζεται ένα νέο πολύγωνο για κάθε ζεύγος ΑΟ-εμποδίου με κέντρο αναφοράς τις ρομποκεντρικές συντεταγμένες, μέσω της μεταξύ τους συνέλιξης, δηλαδή του αθροίσματος Minkowski. Αποτέλεσμα του αθροίσματος Minkowski μεταξύ δύο ορθογωνίων είναι ένα οκτάγωνο (εάν δεν επικαλύπτονται).

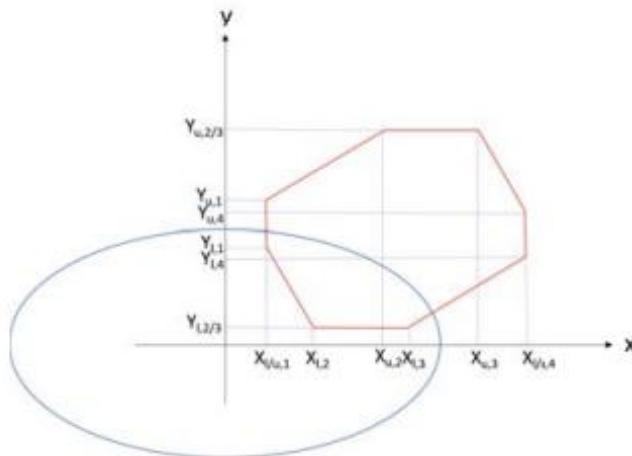


Εικόνα 7-1: Συνέλιξη μεταξύ Αυτόνομου Οχήματος και Εμποδίου-Τζνους

## 7.2 ΑΝΑΛΥΤΙΚΟΣ ΥΠΟΛΟΓΙΣΜΟΣ ΧΩΡΙΚΗΣ ΕΠΙΚΑΛΥΨΗΣ

Για τον υπολογισμό του CSP κάθε ζεύγους ΑΟ-εμποδίου αρχικά γίνεται μετασχηματισμός του συστήματος συντεταγμένων με βάση την θέση του εμποδίου στον χώρο και του προσανατολισμού του έτσι ώστε οι κύριοι άξονες της συνδιακύμανσης να βρίσκονται πάνω στους κάθετους άξονες της βάσης του χώρου.

Κατασκευάζονται όλα τα ευθύγραμμα τμήματα μεταξύ των δύο διαδοχικών κορυφών του, καθένα έχοντας την ιδιότητα  $-1 \leq 1$  εάν βρίσκεται κάτω ή πάνω από τον οριζόντιο άξονα. Τα ευθύγραμμα τμήματα σχεδόν παράλληλα του κάθετου άξονα δεν λαμβάνονται υπόψιν σύμφωνα με την μεθοδολογία.



Εικόνα 7-2: Περιοχή Τομής μεταξύ Οκταγώνου και Συνδιακύμανσης

Υπολογίζεται η τομή μεταξύ της Γκαουσιανής συνδιακύμανσης και του οκταγώνου και στην συνέχεια υπολογίζεται ο όγκος κάτω από την περιοχή αυτή. Αυτός ο όγκος είναι το CSP μεταξύ

του συγκεκριμένου εμποδίου και ΑΟ στον συγκεκριμένο χρόνο [31].

$$P(C, T_p) = \int_y \int_x I_C(x, y) \frac{1}{2\pi\sigma_x\sigma_y} e^{-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right)} dx dy \quad (7.4)$$

Καταλήγουμε στο απλοποιημένο ολοκλήρωμα με την διάσπαση του οκταγώνου σε  $C_n$  τμήματα:

$$P(C_i, T_p) = \int_{y=0}^{y=m_i x + b_i} \int_{x=x_{l,i}}^{x=x_{u,i}} \frac{1}{2\pi\sigma_x\sigma_y} e^{-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right)} dx dy \quad (7.5)$$

Τα  $x_{u,i}, x_{l,i}$  είναι τα όρια του  $C_i$  στον οριζόντιο άξονα και  $y = 0, y = m_i x + b_i$  τα όρια ολοκλήρωσης του κατακόρυφου άξονα.

Μόνο το πρώτο μέρος μπορεί να υπολογιστεί αναλυτικά έτσι:

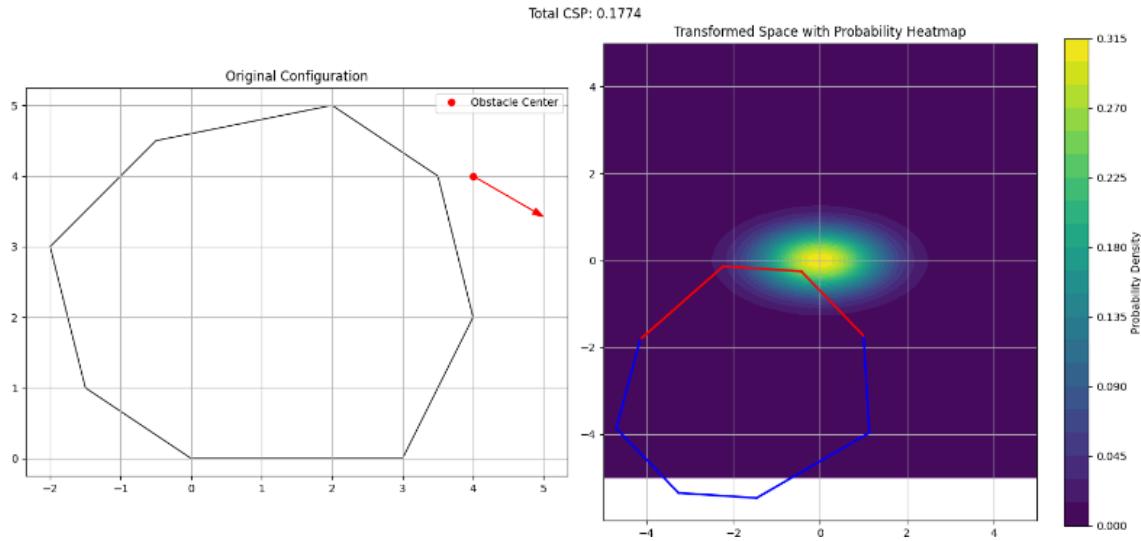
$$P(C_i, T_p) = \frac{1}{\sqrt{8\pi}\sigma_x} \int_{x=x_{l,i}}^{x=x_{u,i}} \operatorname{erf}\left(\frac{m_i x + b_i}{\sqrt{2}\sigma_x}\right) e^{-\frac{1}{2}\sigma_x^2} dx \quad (7.6)$$

Όπου  $\operatorname{erf}$  η συνάρτηση σφάλματος Gauss η οποία δίνει την πιθανότητα της Γκαουσιανής στα συγκεκριμένα όρια. Για τον υπολογισμό του ολοκληρώματος χρησιμοποιείται η συνάρτηση `quad` από το πακέτο `scipy.integrate`.

Τελικά η συνολική πιθανότητα σε αυτόν τον χρόνο για αυτό το ζεύγος Α.Ο-εμποδίου είναι:

$$P(C, T_p) = \sum_{i=0}^{i < m} P(C_i, T_p) - \sum_{i=m}^{i < 6} P(C_i, T_p) \quad (7.7)$$

Όπου  $i < m$  περιοχές θετικής συνεισφοράς. Εάν το οκτάγωνο βρίσκεται κάτω από τον άξονα x τότε οι θετικές και αρνητικές συνεισφορές αντιστρέφονται για αυτό και τα ευθύγραμμα τμήματα έχουν την προ-αναφερθέν ιδιότητα. Εάν η ανάλυση γινόταν προς τον άξονα y η συνεισφορά θα λάμβανε υπόψιν την σχετική θέση προς τον άξονα x, με θετικές να είναι αυτές δεξιά του. Υπάρχει περίπτωση να μην υπάρχει αρνητική συνεισφορά όταν τμήματα είναι πάνω από τον άξονα y και τα υπόλοιπα κάτω από τον άξονα.



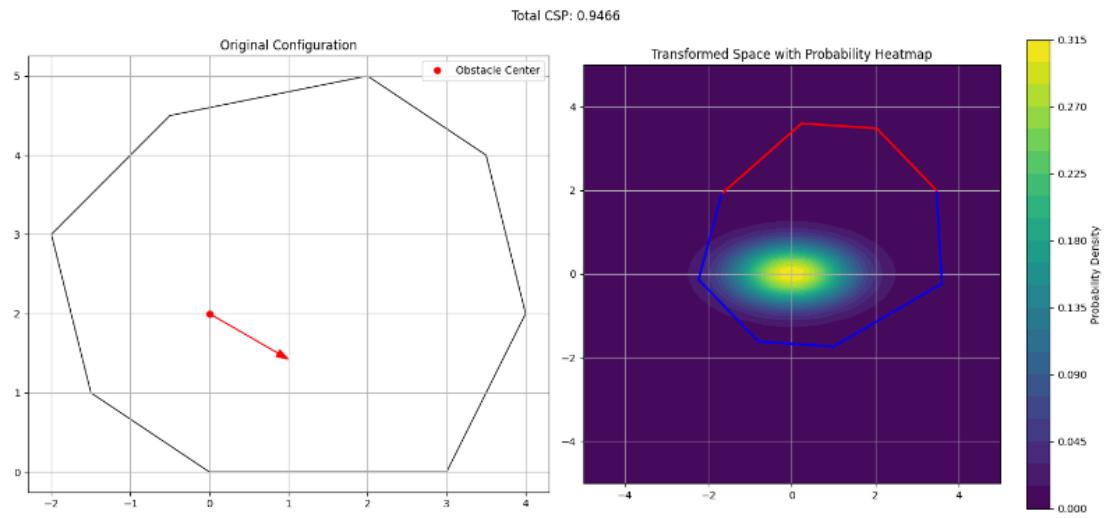
Εικόνα 7-3: Υπολογισμός κινδύνου σύγκρουσης (οκτάγωνο κάτω από τον άξονα  $x$ , αντιστροφή συνεισφορών)

Τα αποτελέσματα συγκρίνονται με την πιθανότητα κινδύνου μεταξύ της Γκαουσιανής και του ορθογωνίου που δημιουργείται αν ληφθεί υπόψιν το μέγιστο εμβαδό του οκτάγωνου, κάτι επιλύσιμου μέσω συναρτήσεων βιβλιοθήκης `scipy`.

Σημαντική σημείωση ότι καθώς η συνδιακύμανση του αντικειμένου-εμποδίου αυξάνεται και το οκτάγωνο πλησιάζει το κέντρο του, η πιθανότητα κινδύνου σύγκρουσης δεν αυξάνεται απαραίτητα γραμμικά. Αυτό οφείλεται στο γεγονός ότι, ενώ η κατανομή Gauss που αντιπροσωπεύει το αντικείμενο-εμπόδιο διευρύνεται (αυξάνεται η αβεβαιότητα θέσης του), το οκτάγωνο παραμένει σταθερό σε μέγεθος. Αυτή η σχέση δημιουργεί μια περίπλοκη αλληλεπίδραση μεταξύ του σταθερού μεγέθους του οκτάγωνου και της αύξησης της αβεβαιότητας τοποθέτησης του εμποδίου<sup>13</sup>.

---

<sup>13</sup> [https://github.com/DimSpathoulas/Collision\\_Risk\\_Calculation.git](https://github.com/DimSpathoulas/Collision_Risk_Calculation.git)



Εικόνα 7-4: Υπολογισμός κινδύνου σύγκρουσης (τμήματα οκταγώνου άνω και κάτω οριζόντιου άξονα)

## **8 ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΠΕΡΑΙΤΕΡΩ ΕΡΕΥΝΑ**

---

### **8.1 ΣΥΜΠΕΡΑΣΜΑΤΑ**

#### **8.1.1 ΜΟΝΑΔΑ ΜΑΘΗΣΗΣ ΠΟΛΛΑΠΛΟΤΗΤΩΝ**

Μεταξύ των διαθέσιμων τεχνικών μείωσης διαστάσεων, η UMAP αποδεικνύεται η καλύτερη επιλογή για τα συγκεκριμένα δεδομένα. Οι γραμμικές τεχνικές, δεν είναι επαρκώς αναπαραστατικές, κυρίως λόγω της υψηλής διαφορισμότητας των δεδομένων (έξι διαφορετικές κάμερες, πολύπλοκες πυκνές γεωμετρίες). Από τις απλές μετρικές, η cosine αναδεικνύεται ως η καλύτερη επιλογή για τη διατήρηση δομής. Παρατηρείται ότι τα εξαγόμενα χαρακτηριστικά είναι ιδιαίτερα ευαίσθητα στη στοχαστικότητα των αισθητήρων, ακόμη και για το ίδιο αντικείμενο, υπάρχουν μικρές αλλά εμφανείς διαφορές στα διανύσματα χαρακτηριστικών. Η πλειοψηφία των λανθασμένων εντοπισμών στα δεδομένα χωρίς κατώφλι αποδοχής είναι συντριπτική. Η ευαισθησία στη ρύθμιση του κατωφλίου δείχνει ότι ένα εύρος αποδοχής μεταξύ 0.5 και 0.6 αποτελεί σημείο ισορροπίας μεταξύ της δραστικής αύξησης λανθασμένων εντοπισμών και μείωσης χαμένων ανιχνεύσεων. Το μέγεθος της κυψέλης επηρεάζει εμφανώς τη μορφολογία των αντικειμένων, καθιστώντας την επιλογή του μεγέθους της κυψέλης κρίσιμο παράγοντα για την ακρίβεια της αναπαράστασης καθώς και την επικείμενη χρήση των χαρακτηριστικών. Τέλος, η ανάλυση μέσω SVD δείχνει ότι, σε λογικά πλαίσια, μπορεί να διατηρήσει τα πιο σημαντικά χαρακτηριστικά, τόσο τα γεωμετρικά όσο και τα εμφανισιακά. Ωστόσο, η χρήση της είναι κυρίως αποτελεσματική για σκοπούς συμπίεσης και ταχύτητας επεξεργασίας, παρά για την απόλυτη διατήρηση δομών.

#### **8.1.2 ΜΟΝΑΔΑ ΠΑΡΑΚΟΛΟΥΘΗΣΗΣ ΠΟΛΛΑΠΛΩΝ ΣΤΟΧΩΝ**

Η απλή μετρική cosine δεν είναι αρκετά αναπαραστατική στον υπολογισμό απόστασης μεταξύ των χαρακτηριστικών. Ο παράγοντας ανάμειξης είναι κρίσιμος στην υλοποίηση ακόμα και εάν δεν διαφοροποιεί σημαντικά τα αποτελέσματα όταν χρησιμοποιείται. Η μονάδα μήτρας εμφανισιακών χαρακτηριστικών έχει ως σκοπό την εκμάθηση εσωτερικών αποστάσεων των πολλαπλοτήτων. Οι πολλαπλότητες του συνόλου δεδομένων εκπαίδευσης και επικύρωσης είναι παρόμοιες αφού το σύστημα παρακολούθησης λειτουργεί. Τα γεωμετρικά χαρακτηριστικά είναι αυτά που ενισχύουν περισσότερο το σύστημα. Τα εμφανισιακά χαρακτηριστικά έχουν ως κύρια δύναμη το χρώμα. Μέσω κανονικοποίησης και μόνον

μπορούν να προστεθούν με ουσιαστικό τρόπο τα γεωμετρικά και εμφανισιακά χαρακτηριστικά μέσα στην μονάδα συγχώνευσης. Τεχνικές γραμμικής ελάττωσης διαστάσεων όπως η SVD μπορούν να πλησιάσουν τα τελικά αποτελέσματα, αλλά όχι πλήρως. Το σύστημα δίνει παρόμοια αποτελέσματα σε πολλαπλές επαναλήψεις, οπότε είναι στιβαρό. Η μονάδα παρακολούθησης μπορεί να ενισχυθεί πέραν της ψευδομετρικής mahalanobis εάν γίνει σωστή συγχώνευση με την μήτρα χαρακτηριστικών, καθώς οι δύο τους βρίσκονται σε διαφορικές κλίμακες.

### 8.1.3 MONADA ANIXNEYESHΣ KINΔΥΝΟΥ

Λόγω της φύσης του συνόλου δεδομένων οι τιμές κινδύνου άνω του 10% είναι σπάνιες. Για τον υπολογισμό του CSP δεν χρειάζεται η γνώση της σχετικής ταχύτητας των αντικειμένων. Η αύξηση της συνδιακύμανσης δεν σημαίνει αύξηση κινδύνου σύγκρουσης ακόμα και εάν τα αντικείμενα παραμένουν στάσιμα μεταξύ τους. Ο κίνδυνος σύγκρουσης αυξάνεται γεωμετρικά όταν υπάρχει διείσδυση. Σε συγκεκριμένες καταστάσεις προβολής (ανάλογα τα δευτερόλεπτα πρόβλεψης) το οκτάγωνο μπορεί να διαπεράσει πλήρως το ίχνος και η πιθανότητα σύγκρουσης να αυξηθεί και ύστερα να μειωθεί, αφού δεν υπάρχει τρόπος ελέγχου του Α.Ο. αλλά και το μοντέλο πρόβλεψης έχει όλο και μικρότερο βαθμό εμπιστοσύνης σε βάθος χρόνου. Η πιθανότητα σύγκρουσης μεταξύ οκτάγωνου και συνδιακύμανσης και αντίστοιχου ορθογωνίου και συνδιακύμανσης είναι παρόμοιες σε συγκεκριμένα σενάρια, ανάλογα της διαφοράς μεγέθους που προκύπτει από την γεωμετρική απλοποίηση. Δεν υπάρχει σχεδιασμός τροχιάς σε πραγματικό χρόνο για πλήρη μοντελοποίηση και αξιοποίηση του αλγόριθμου.

## 8.2 ΠΕΡΑΙΤΕΡΩ ΕΡΕΥΝΑ

Η παρούσα εργασία είχε ως στόχο την εξαγωγή χαρακτηριστικών από δύο πασίγνωστους και ευρέως πολυχρησιμοποιημένους ανιχνευτές, τον τρισδιάστατο ανιχνευτή CenterPoint και τον δισδιάστατο ανιχνευτή Mask R-CNN. Εν συνεχείᾳ την αναπαράσταση και ανάλυση των αυτών των χαρακτηριστικών, την χρονική τους σύνδεση με σκοπό την χρήση των μεταξύ τους αποστάσεων ως μήτρα συσχέτισης σε ένα περιβάλλον παρακολούθησης πολλαπλών στόχων και τέλος την χρήση αυτών των ιχνών σε μία αναλυτική επίλυση υπολογισμού κινδύνου σύγκρουσης. Η ολική μονάδα εξαγωγής χαρακτηριστικών μπορεί να εξάγει χρήσιμα και ουσιαστικά γεωμετρικά και εμφανισιακά μοτίβα. Η περιοχή πλέγματος είναι μείζονος σημασίας στον όγκο της πληροφορίας αλλά και στην πολυπλοκότητα της για μεγάλα

αντικείμενα ή ακόμα και κίνδυνο να ξεφεύγει από τον εντοπισμό μικρών αντικειμένων. Το ακριβώς επόμενο επίπεδο χαρακτηριστικών είναι διάστασης 64, αν και χρησιμοποιήθηκε σε πειράματα μπορεί μέσω πιο εκλεπτυσμένων μεθόδων να περιέχει περισσότερη χωρική συμπιεσμένη πληροφορία. Τα εμφανισιακά χαρακτηριστικά είναι κατά βάση πιο αδύναμα από τα γεωμετρικά. Σε ακραίες καταστάσεις όπως αντικείμενα να φεύγουν πολύ γρήγορα σε διαφορικά πεδία όρασης και διαφορετικές κάμερες μπορεί το διάνυσμα χαρακτηριστικών του ίδιου ίχνους να μην έχει σχεδόν καμία σχέση εκτός του χρώματος. Θα μπορούσε να αναπτυχθεί ένας ταχύτερος αλγόριθμος κατασκευής τελικών διανυσμάτων πληροφορίας μέσω μεθόδων πολύ-επεξεργασίας, ως βάση υπάρχει το αυτόνομο<sup>14</sup> αποθετήριο<sup>15</sup>. Η μονάδα μάθησης πολλαπλών συλλόγων παρέχει σχετικά σταθερές καμπύλες στις εξόδους της, ίσως ωστόσο η μετρική απόστασης cosine όπως αποδεικνύεται από την επόμενη μονάδα να μην είναι ιδανική για πραγματική αποτύπωση. Ως μελλοντική έρευνα η χρήση πιο σύγχρονων μεθόδων, όπως μεταβαλλόμενών αυτό-κωδικοποιητών ή αρχιτεκτονικές κωδικοποιητών-αποκωδικοποιητών με συναρτήσεις σφάλματος εμπνευσμένες από μη γραμμικές τεχνικές (πχ UMAP) θα ήταν πιο αρμόζουσες. Εύλογη θα ήταν επίσης μία παρόμοια μελέτη στο υποσύνολο μάθησης, αλλά και σύγκριση των δύο υποσυνόλων, με χρήση μεθόδων συσχέτισης των ακμών των παραγόμενων πολλαπλότητων. Η μονάδα παρακολούθησης πολλαπλών στόχων θα μπορούσε να βελτιωθεί συγχωνεύοντας τις δύο μήτρες χαρακτηριστικών και κατάστασης, αλλά και μέσω της μονάδας αρχικοποίησης ιχνών. Πειράματα και αρχιτεκτονικές έχουν ήδη υλοποιηθεί στο αυτόνομο αποθετήριο<sup>16</sup>. Η μονάδα ανίχνευσης κινδύνου<sup>17</sup> θα επωφελούταν από σενάρια προσομοίωσης, επιπρόσθετα θα ήταν χρήσιμη η υλοποίηση και εφαρμογή υπολογισμού πιθανότητας συμβάντος σύγκρουσης η οποία λαμβάνει υπόψιν την δυναμική κατάσταση των αντικειμένων. Τέλος η υλοποίηση του αλγόριθμου στις τρείς διαστάσεις θα ήταν ακόμα πιο ωφέλιμη σε εφαρμογές αέρα και ύδατος. Όλες οι μονάδες με αλλαγές μπορούν να λειτουργήσουν και για τον αέρα (π.χ. UAVs) και για το νερό (π.χ. υποβρύχια) ακόμα και για διαστημικές εφαρμογές (π.χ. αποφυγή διαστημικών συντριμμάτων) μέσω αντίστοιχων συνόλων δεδομένων αλλά και μέσω προσομοιώσεων.

<sup>14</sup> [https://github.com/DimSpathoulas/2D\\_FEATURE\\_EXTRACTOR.git](https://github.com/DimSpathoulas/2D_FEATURE_EXTRACTOR.git)

<sup>15</sup> [https://github.com/DimSpathoulas/Point\\_Cloud\\_Feature\\_Extractor.git](https://github.com/DimSpathoulas/Point_Cloud_Feature_Extractor.git)

<sup>16</sup> [https://github.com/DimSpathoulas/GeomApp\\_3MOT.git](https://github.com/DimSpathoulas/GeomApp_3MOT.git)

<sup>17</sup> [https://github.com/DimSpathoulas/Collision\\_Risk\\_Calculation.git](https://github.com/DimSpathoulas/Collision_Risk_Calculation.git)

## БІВЛІОГРАФІА

---

1. Yeong, D.J., Velasco-Hernandez, G., Barry, J. and Walsh, J., 2021. Sensor and sensor fusion technology in autonomous vehicles: A review. *Sensors*, 21(6), p.2140.
2. Woo, A., Fidan, B. and Melek, W.W., 2018. Localization for autonomous driving. *Handbook of position location: theory, practice, and advances, second edition*, pp.1051-1087.
3. Wu, Y., Wang, Y., Zhang, S. and Ogai, H., 2020. Deep 3D object detection networks using LiDAR data: A review. *IEEE Sensors Journal*, 21(2), pp.1152-1171.
4. Zhu, J., Li, H. and Zhang, T., 2023. Camera, LiDAR, and IMU based multi-sensor fusion SLAM: A survey. *Tsinghua Science and Technology*, 29(2), pp.415-429.
5. Gonzalez, R.C., 2009. *Digital image processing*. Pearson education india.
6. Hartley, R. and Zisserman, A., 2003. *Multiple view geometry in computer vision*. Cambridge university press.
7. Haralick, R.M. and Shapiro, L.G., 1993. *Computer and robot vision*. Computer and Robot Vision.
8. Hart, P.E., Stork, D.G. and Duda, R.O., 2000. *Pattern classification*. Hoboken: Wiley.
9. Bishop, C.M., 2006. *Pattern recognition and machine learning by Christopher M. Bishop*. Springer Science+ Business Media, LLC.
10. Thrun, S. and Burgard, W., Fox. D.(2005), Probabilistic Robotics.
11. Greg, W. and Gary, B., 2006. An introduction to the Kalman filter.
12. He, K., Gkioxari, G., Dollár, P. and Girshick, R., 2017. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 2961-2969).
13. Girshick, R., Donahue, J., Darrell, T. and Malik, J., 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 580-587).
14. Girshick, R., 2015. Fast r-cnn.
15. Ren, S., He, K., Girshick, R. and Sun, J., 2016. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE transactions on pattern analysis and machine intelligence*, 39(6), pp.1137-1149. *recognition* (pp. 2117-2125).
16. He, K., Zhang, X., Ren, S. and Sun, J., 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

17. Lin, T.Y., Dollár, P., Girshick, R., He, K., Hariharan, B. and Belongie, S., 2017. Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern*
18. Zhou, Y. and Tuzel, O., 2018. Voxelnet: End-to-end learning for point cloud based 3d object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4490-4499).
19. Yan, Y., Mao, Y. and Li, B., 2018. Second: Sparsely embedded convolutional detection. *Sensors*, 18(10), p.3337.
20. Graham, B. and Van der Maaten, L., 2017. Submanifold sparse convolutional networks. *arXiv preprint arXiv:1706.01307*.
21. Yin, T., Zhou, X. and Krahenbuhl, P., 2021. Center-based 3d object detection and tracking. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 11784-11793).
22. McInnes, L., Healy, J. and Melville, J., 2018. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*.
23. Maćkiewicz, A. and Ratajczak, W., 1993. Principal components analysis (PCA). *Computers & Geosciences*, 19(3), pp.303-342.
24. Gerbrands, J.J., 1981. On the relationships between SVD, KLT and PCA. *Pattern recognition*, 14(1-6), pp.375-381.
25. Van der Maaten, L. and Hinton, G., 2008. Visualizing data using t-SNE. *Journal of machine learning research*, 9(11).
26. Arora, S., Hu, W. and Kothari, P.K., 2018, July. An analysis of the t-sne algorithm for data visualization. In *Conference on learning theory* (pp. 1455-1462). PMLR.
27. Weng, X., Wang, J., Held, D. and Kitani, K., 2020, October. 3d multi-object tracking: A baseline and new evaluation metrics. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (pp. 10359-10366). IEEE.
28. Kuhn, H.W., 1955. The Hungarian method for the assignment problem. *Naval research logistics quarterly*, 2(1-2), pp.83-97.
29. A. Houenou, P. Bonnifait, and V. Cherfaoui, "Risk assessment for collision avoidance systems," in Intelligent Transportation Systems (ITSC), 2014 IEEE 17th International Conference on. IEEE, 2014, pp. 386–391
30. Altendorfer, R. and Wilkmann, C., 2021. A new approach to estimate the collision probability for automotive applications. *Automatica*, 127, p.109497.
31. Philipp, A. and Goehring, D., 2019, May. Analytic collision risk calculation for autonomous vehicle navigation. In *2019 International Conference on Robotics and Automation (ICRA)* (pp. 1744-1750). IEEE.
32. Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P. and Zitnick, C.L., 2014. Microsoft coco: Common objects in context. In *Computer Vision–ECCV 2014: 13th*

*European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V* 13 (pp. 740-755). Springer International Publishing.

33. Caesar, H., Bankiti, V., Lang, A.H., Vora, S., Liong, V.E., Xu, Q., Krishnan, A., Pan, Y., Baldan, G. and Beijbom, O., 2020. nuscenes: A multimodal dataset for autonomous driving. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp.11621-11631).
34. Zhu, B., Jiang, Z., Zhou, X., Li, Z. and Yu, G., 2019. Class-balanced grouping and sampling for point cloud 3d object detection. *arXiv preprint arXiv:1908.09492*.
35. Chiu, H.K., Prioletti, A., Li, J. and Bohg, J., Probabilistic 3D multi-object tracking for autonomous driving. *arXiv* 2020. *arXiv preprint arXiv:2001.05673*.
36. Chiu, H.K., Li, J., Ambruş, R. and Bohg, J., 2021, May. Probabilistic 3D multi-modal, multi-object tracking for autonomous driving. In *2021 IEEE international conference on robotics and automation (ICRA)* (pp. 14227-14233). IEEE.