

Ανοικτό Πανεπιστήμιο Κύπρου

Σχολή Θετικών και Εφαρμοσμένων Επιστημών

**Μεταπτυχιακό Πρόγραμμα Σπουδών
Πληροφορικά και Επικοινωνιακά Συστήματα**

Μεταπτυχιακή Διατριβή



Computer Vision for Aerial Applications

Κωνσταντίνος Παππάς

**Επιβλέπων Καθηγητής
Αντρέας Παναγίδης**

Μάιος 2018

Ανοικτό Πανεπιστήμιο Κύπρου

Σχολή Θετικών και Εφαρμοσμένων Επιστημών

**Μεταπτυχιακό Πρόγραμμα Σπουδών
Πληροφοριακά και Επικοινωνιακά Συστήματα**

Μεταπτυχιακή Διατριβή

Computer Vision for Aerial Applications

Κωνσταντίνος Παππάς

**Επιβλέπων Καθηγητής
Αντρέας Παναγίδης**

Η παρούσα μεταπτυχιακή διατριβή υποβλήθηκε προς μερική εκπλήρωση των απαιτήσεων για απόκτηση μεταπτυχιακού τίτλου σπουδών στα Πληροφοριακά και Επικοινωνιακά Συστήματα από τη Σχολή Θετικών και Εφαρμοσμένων Επιστημών του Ανοικτού Πανεπιστημίου Κύπρου.

Μάιος 2018

ΛΕΥΚΗ ΣΕΛΙΔΑ

Περίληψη

Το αντικείμενο της επιτήρησης περιοχών για λόγους ασφαλείας αποτελεί ένα κεφαλαιώδους σημασίας ζήτημα για κάθε κρατική οντότητα. Μία από τις βασικές αρχές εκμετάλλευσης πληροφοριών, η συλλογή των οποίων αποτελεί και το κύριο αντικείμενο επιτήρησης, είναι η δυνατότητα έγκαιρης μεταβίβασής τους στα αντίστοιχα όργανα λήψης αποφάσεων. Έχοντας ως φάρο τα παραπάνω, η παρούσα μεταπτυχιακή διατριβή ασχολείται με την ανάπτυξη μίας διάταξης η οποία έχει ως στόχο την αυτόματη ανίχνευση και ιχνηλάτιση οχημάτων, εφαρμόζοντας τεχνικές μηχανικής όρασης σε ροή video που μεταδίδεται σε σχεδόν πραγματικό χρόνο από εναέριες πλατφόρμες επιτήρησης, με εκμετάλλευση των τεχνολογιών 3G/4G της κινητής τηλεφωνίας.

Για την υλοποίηση της συγκεκριμένης διάταξης πραγματοποιήθηκε η εκπαίδευση δύο τύπων κατηγοριοποιητών σε εικόνες υψηλής ανάλυσης. Ο πρώτος κατηγοριοποιητής εκπαιδεύτηκε με μηχανή υποστήριξης διανυσμάτων η οποία τροφοδοτήθηκε με απεικονίσεις χαρακτηριστικών σημείων των οχημάτων, που προέκυψαν από τη χρήση Ιστογραμμάτων Βάθμωσης. Ο δεύτερος στηρίζεται στην κατασκευή μίας αλληλουχίας κατηγοριοποιητών που εκπαιδεύονται με εφαρμογή τεχνικών ενίσχυσης αξιοποιώντας απεικονίσεις χαρακτηριστικών σημείων που εξήχθησαν μέσω Τοπικών Δυαδικών Μοτίβων, μετασχηματισμών Haar και Ιστογραμμάτων Βάθμωσης. Στα μοντέλα που προέκυψαν από τις παραπάνω εκπαιδεύσεις, αφού αξιολογήθηκαν, διαπιστώθηκε ότι αυτό με τον κατηγοριοποιητή που εκπαιδεύτηκε με μηχανή υποστήριξης διανυσμάτων κάνοντας χρήση Ιστογραμμάτων Βάθμωσης παρουσίασε τις καλύτερες επιδόσεις σε όλες τις κατηγορίες πλην αυτής του απαιτούμενου χρόνου για την πραγματοποίηση της ανίχνευσης. Λόγω της εφαρμογής του συγκεκριμένου μοντέλου σε video, για την επιτάχυνση της διαδικασίας, συνδυάστηκε με αλγόριθμο ιχνηλάτισης οχημάτων. Επίσης, μετά την πραγματοποίηση μετρήσεων της ταχύτητας μετάδοσης δεδομένων του δικτύου 3G/4G της κινητής τηλεφωνίας σε διαφορετικά σημεία- ύψη - ταχύτητες πτήσης, διαπιστώθηκε ότι αποτελεί μια ρεαλιστική προσέγγιση ο σχεδιασμός εφαρμογών που θέτει τα 512Kbps ως ελάχιστη ταχύτητα δικτύου.

Τέλος, αφού λήφθησαν υπόψη οι παραπάνω διαπιστώσεις, πραγματοποιήθηκε επιτυχής δοκιμή ανίχνευσης και ιχνηλάτισης οχημάτων σε ροή video, η οποία μεταδιδόταν σε σχεδόν πραγματικό χρόνο από μία διάταξη Raspberry/κάμερα/USB 4G Stick, προσαρμοσμένη σε E/Π και drone. Η συγκεκριμένη δοκιμή αποδεικνύει ότι είναι εφικτή η ανάπτυξη ευέλικτων και χαμηλού κόστους διατάξεων για την επιτήρηση περιοχών από απόσταση.

Λέξεις Κλειδιά: Μηχανική όραση, Ροή video, 3G/4G δίκτυα κινητής τηλεφωνίας, Ανίχνευση, Ιχνηλάτιση, Ιστογράμματα Βάθμωσης, Τοπικό Δυαδικό Μοτίβο, Haar.

Summary

The subject of area surveillance for security reasons, is an issue of vital importance for every state. One of the main principles regarding the exploitation of information, keeping in mind that the intelligence is the main purpose of surveillance, is the capability to transmit these information to decision makers, as soon as possible. In the view of the above, this master thesis deals with the development of a system whose main purpose is the near-real-time automatic detection and tracking of vehicles. This task is accomplished with the implementation of computer vision techniques to video streaming that is transmitted from aerial surveillance platforms, which make use of the 3G/4G mobile services technology.

In order to fulfill this purpose, two different classifiers were trained, using high definition pictures. The first one was trained with Support Vector Machine (SVM), using Histograms of Oriented Gradients (HOG) as feature representation. The second one is based on boosted cascade classifiers which use the Gentle AdaBoost algorithm in conjunction with three different feature descriptors, Local Binary Patterns (LBP), Haar wavelets and HOG. Regarding the models that originated from this training procedure, after their evaluation, it was found that the HOG/SVM model outperformed the rest algorithms with the exception of the time needed to scan the image in order to detect the presence of vehicles. Due to the fact that the intention was to implement the HOG/SVM model on video streaming, it was combined with a correlation tracker, making the whole system considerably quicker. Also, after measuring the data transfer speed of 3G/4G cellular networks, in different height-speed and areas of flight, it was found that it may be considered a realistic approach the design of systems which establish the 512Kbps as a minimum network speed.

Finally, taking into consideration the fore mentioned findings, it was carried out a successful test of vehicle detection and tracking using near - real - time video streaming which was transmitted by a Raspberry/camera/USB 4G Stick system, mounted both on helicopter and drone. This particular test proves that it is possible the development of agile and low cost systems that can be used for remote surveillance of a distant areas.

Keywords: Computer vision, Video streaming, 3G/4G cellular network, vehicle detection and tracking, Histograms of Oriented Gradients (HOG), Local Binary Patterns (LBP), Haar.

Ευχαριστίες

Για την εκπόνηση της παρούσας μεταπτυχιακής διατριβής θα ήθελα να εκφράσω τις θερμές μου ευχαριστίες στον επιβλέποντα καθηγητή μου Δρ. Αντρέα Παναγίδη ο οποίος με εμπιστεύτηκε αναθέτοντάς μου τη συγκεκριμένη εργασία, με καθοδήγησε, μου πρόσφερε ουσιαστική βοήθεια σε σημεία που συναντούσα σοβαρές δυσκολίες και μου παρείχε τις απαραίτητες συμβολές για τη βελτίωσή της.

Ιδιαίτερες ευχαριστίες θα ήθελα να εκφράσω στο Γενικό Επιτελείο Στρατού που μου επέτρεψε την συνεκμετάλλευση προγραμματισμένων πτήσεων για τους σκοπούς της έρευνας. Χωρίς την άδειά του δεν θα ήταν εφικτή τόσο η άντληση πολύτιμων συμπερασμάτων σχετικά με την ταχύτητα ροής δεδομένων του δικτύου κινητής τηλεφωνίας, όσο και η υλοποίηση της δοκιμής μετάδοσης video σε πραγματικό χρόνο. Πολύτιμη ήταν και η συμβολή του Διοικητή του 1ου Τάγματος Ε/Π Στρατού, ο οποίος στάθηκε αρωγός στην προσπάθειά μου, με τη διάθεση των αναγκαίων Ε/Π για την υλοποίηση των μετρήσεων.

Επιπλέον θα ήθελα να ευχαριστήσω τους κυρίους Νικόλαο Βορνίτη, Στράτο Κερανίδη, Νίκο Μακρή και Ιωάννη Καζνταρίδη για τη διάθεση μέσων για τους σκοπούς της εργασίας (drone, υλικά κλπ), καθώς επίσης και τη βοήθεια που παρείχαν στην πτήση των drone, προκειμένου να δημιουργηθεί η βάση δεδομένων με τις εναέριες βιντεοσκοπήσεις οχημάτων.

Σε αυτό το σημείο θα ήθελα να ευχαριστήσω όλους τους καθηγητές μου, στο συγκεκριμένο μεταπτυχιακό πρόγραμμα σπουδών, οι οποίοι αποτέλεσαν τους φάρους που με καθοδήγησαν τα τελευταία τέσσερα έτη σε αυτό το δύσκολο και συνάμα υπέροχο ταξίδι, σε έναν ωκεανό γνώσεων.

Τέλος ευχαριστώ την οικογένειά μου, την Λιάννα, τον Απόστολο, τον Χριστόδουλο, τον Στυλιανό και τη Φαίδρα, που με στήριξαν όλα αυτά τα χρόνια, χωρίς να εκφράσουν ούτε ένα παράπονο για τις συχνές απουσίες μου από σημαντικές στιγμές της καθημερινότητάς τους.

Περιεχόμενα

1	Εισαγωγή	1
1.1	Πρόλογος.....	1
1.2	Σκοπός – Ερευνητικά Ερωτήματα.....	4
1.3	Κίνητρο - Αιτιολόγηση του Θέματος.....	6
1.4	Γενική Περιγραφή των Αντικειμένων στα οποία Στηρίχθηκε η Μεταπτυχιακή Διατριβή.....	8
1.4.1	Εναέρια Μέσα Επιτήρησης (Ελικόπτερα - Μη Επανδρωμένα Εναέρια Συστήματα).....	8
1.4.2	Computer Vision.....	10
1.5	Δομή Εργασίας.....	16
2	Ανιχνευτές Αντικειμένων και Κατηγοριοποιητές	18
2.1	Αναγνώριση και Ανίχνευση.....	18
2.2	Επεξεργασία Ψηφιακής Εικόνας.....	19
2.2.1	Βασικά Χαρακτηριστικά Εικόνας.....	20
2.2.2	Θόρυβος.....	21
2.2.3	Features – Descriptors.....	22
2.2.4	Corner Detector.....	23
2.3	Descriptors.....	25
2.3.1	Τοπικό Δυαδικό Μοτίβο (Local Binary Patterns - LBP).....	26
2.3.2	Μετασχηματισμοί Haar.....	29
2.3.3	Ιστογράμματα της Βάθμωσης (Histogram of Oriented Gradients-HOG).....	32
2.4	Κατηγοριοποιητές (Classifiers).....	36
2.4.1	Ada Boost (Gentle Ada Boost – GAB)	37
2.4.2	Μηχανή Διανυσμάτων Υποστήριξης (Support Vector Machines – SVM).....	41
2.5	Αξιολόγηση των Αποτελεσμάτων της Ανίχνευσης.....	43
2.5.1	Πραγματικά Δεδομένα (Ground Truth)	43
2.5.2	Ακρίβεια Προσδιορισμού Θέσης.....	44
2.5.3	Μήτρα Σύγχυσης.....	45
2.5.4	Μέτρα Απόδοσης.....	46
3	Ιχνηλάτηση Αντικειμένων μέσω Video Streaming	48
3.1	Ιχνηλάτηση.....	48
3.1.1	Ανάλυση της Τεχνικής της Ιχνηλάτησης.....	48
3.1.2	Αλγόριθμος Συσχετιστικής Ιχνηλάτησης (Correlation Tracker).....	50
3.2	Video.....	52
3.2.1	Ανάλυση (Video Resolution)	53
3.2.2	Συμπίεση Video.....	54
3.2.3	Video Streaming.....	57
3.3	Μεταφορά Δεδομένων Μέσω του Δικτύου Κινητής Τηλεφωνίας.....	59
3.3.1	Συστήματα Κινητής Τηλεφωνίας.....	60
3.3.2	Συστήματα Δεύτερης Γενιάς (2G).....	63
3.3.3	Συστήματα Τρίτης Γενιάς (3G).....	64
3.3.4	LTE και Συστήματα Τέταρτης Γενιάς (4G).....	66
3.3.5	Το Internet μέσω Κινητής Τηλεφωνίας στην Ελλάδα.....	66
4	Βάσεις Δεδομένων	69
4.1	Διαδικτυακές Βάσεις Δεδομένων Οχημάτων.....	69
4.1.1	Βάση Δεδομένων του California Institute of Technology.....	69
4.1.2	Βάση Δεδομένων CBCL StreetScenes Challenge Framework.....	69
4.1.3	UIUC Image Database	70
4.1.4	LISA – Q Front FOV Database	70
4.1.5	Βάση Δεδομένων του University of Central Florida (UCF)	71
4.1.6	DLR Munich Vehicle Dataset	71
4.1.7	Dutch UAS Dataset 001	71
4.1.8	Overhead Imagery Research Data Set (OIRDS)	72
4.1.9	Stanford Drone Dataset	73

4.1.10	INRIA Aerial Image Labeling.....	73
4.1.11	Vehicle Detection in Aerial Imagery (VEDAI)	74
4.2	Μία Νέα Βάση Δεδομένων.....	74
4.2.1	Μέσα που Χρησιμοποιήθηκαν για τις Εναέριες Λήψεις.....	75
4.2.2	Σχεδίαση των Προδιαγραφών της Νέας Βάσης Δεδομένων.....	76
5	Μεθοδολογία	82
5.1	Εκπαίδευση Κατηγοριοποιητών.....	82
5.1.1	Δημιουργία Δειγμάτων Εκπαίδευσης και Ελέγχου.....	82
5.1.2	Δοκιμαστική Φάση Εκπαίδευσης των Αλγορίθμων.....	85
5.1.3	Εκπαίδευση των Κατηγοριοποιητών.....	88
5.2	Χρησιμοποιούμενο Λογισμικό – Αλγόριθμοι Ανίχνευσης/ Ιχνηλάτησης – Αξιολόγηση Μοντέλων.....	92
5.3	Μέτρηση Ταχύτητας Δικτύου.....	102
5.4	Video Streaming από Ε/Π.....	103
5.5	Ανίχνευση και Ιχνηλάτηση σε Video Streaming Μεταδιδόμενο από UAV.....	106
6	Αποτελέσματα	108
6.1	1ο Στάδιο Αξιολόγησης: Απόδοση των Μοντέλων Ανίχνευσης Οχημάτων....	108
6.1.1	Διάρκεια Χρόνου Εκπαίδευσης.....	108
6.1.2	Αξιολόγηση της Ακρίβειας Εντοπισμού (Localization) των Οχημάτων.....	111
6.1.3	Ποιοτικά Χαρακτηριστικά της Απόδοσης κάθε Μοντέλου.....	113
6.1.4	Ποσοτική Σύγκριση της Απόδοσης κάθε Μοντέλου Χρησιμοποιώντας Μετρικούς Δείκτες.....	121
6.1.5	Σύγκριση Απόδοσης κάθε Μοντέλου μεταξύ των Περιοχών 1-2 και 3-4.....	126
6.1.6	Χρόνος Ανίχνευσης.....	127
6.2	2ο Στάδιο Αξιολόγησης: Περαιτέρω Μελέτη των Μοντέλων που Διακρίθηκαν.....	130
6.2.1	Αξιολόγηση της Απόδοσης σε Συμπιεσμένες Εικόνες.....	130
6.2.2	Εφαρμογή των HOG_GAB/SVM σε Διαφορετικά Datasets.....	132
6.3	3ο Στάδιο Αξιολόγησης: Απόδοση Αλγόριθμου Ιχνηλάτησης.....	135
6.4	Μετρήσεις Ταχύτητας Δικτύου.....	137
6.5	Near Real Time Video Streaming.....	138
6.5.1	Video Streaming από Ε/Π προς Σταθμό Εδάφους.....	139
6.5.2	Εφαρμογή Μοντέλου Ανίχνευσης και Ιχνηλάτησης Οχημάτων με Video Streaming από Ε/Π.....	140
6.5.3	Εφαρμογή Μοντέλου Ανίχνευσης και Ιχνηλάτησης Οχημάτων με Video Streaming από UAS.....	141
7	Συμπεράσματα	142
8	Επίλογος	149
	Βιβλιογραφία	152

Παραρτήματα

A Annotation Systems

A.1	Imglab.....	A-1
A.2	Object Marker.....	A-2
A.3	Cascade Trainer GUI.....	A-3

B Μετρήσεις Ταχύτητας Δικτύου

B.1	Μετρήσεις της Ταχύτητας Δικτύου (UAS).....	B-1
B.2	Μετρήσεις της Ταχύτητας Δικτύου (Ε/Π).....	B-1

Λίστα Εικόνων

Εικόνα 1. :	Επίδραση φωτισμού στην τιμή των pixel.....	15
Εικόνα 2. :	Η συνάρτηση f μιας εικόνας.....	19
Εικόνα 3. :	Οι τρεις τύποι ειόνων:δυναδική, κλίμακας του γκρι και έγχρωμη.....	20
Εικόνα 4. :	Είδη θορύβου.....	21
Εικόνα 5	Σημαντικότητα των γωνιών.....	23
Εικόνα 6. :	Κατηγοριοποίηση των σημείων μιας εικόνας, χρησιμοποιώντας τις τιμές ιδιοτιμές λ_1 και λ_2 (eigenvalues).....	24
Εικόνα 7. :	Οπτική απεικόνιση της εφαρμογής του Corner Response Function	25
Εικόνα 8. :	Αλγόριθμος LBP.....	27
Εικόνα 9. :	LBP histogram.....	28
Εικόνα 10. :	Κυκλικά συμμετρικοί γείτονες για διάφορες τιμές του (p, r)	28
Εικόνα 11. :	Η συνάρτηση ∇f για την κλίση (gradient) της εικόνας.....	29
Εικόνα 12. :	Haar – like features.....	30
Εικόνα 13. :	Αναπαράσταση εφαρμογής των Haar Features σε μία εικόνα.....	31
Εικόνα 14. :	Έλεγχος των κατηγοριοποιητών κατά στάδια.....	32
Εικόνα 15. :	Οπτικοποίηση της τιμής της κλίσης.....	33
Εικόνα 16. :	Μέγεθος και κατεύθυνση σε μοίρες της κλίσης	34
Εικόνα 17. :	Απεικόνιση των Oriented Gradients.....	35
Εικόνα 18. :	Επίλυση του προβλήματος κατηγοριοποίησης.....	36
Εικόνα 19. :	Γραφική απεικόνιση του αλγορίθμου AdaBoost.....	38
Εικόνα 20. :	Διαχωρισμός δυαδικού συνόλου δεδομένων.....	41
Εικόνα 21. :	Μη γραμμικά διαχωριζόμενα σύνολα.....	42
Εικόνα 22. :	Υπολογισμός IoU.....	44
Εικόνα 23. :	Αναλύσεις video.....	53
Εικόνα 24. :	Μέσος όρος κέρδους Bit Rate μεταξύ διαφορετικών προτύπων συμπίεσης.....	54
Εικόνα 25. :	Δειγματοληψία των στοιχείων YUV κατά τη συμπίεση.....	55
Εικόνα 26. :	Απεικόνιση της χρονικής συμπίεσης.....	56
Εικόνα 27. :	Απεικόνιση του video streaming.....	59
Εικόνα 28. :	Εφαρμογή της τεχνικής επαναχρησιμοποίησης συχνοτήτων σε κυψελοειδή δίκτυα επικοινωνιών.....	61
Εικόνα 29. :	Καθ' ύψος κάλυψη ενός BS.....	62
Εικόνα 30. :	Χάρτες κάλυψης παρόχων κινητής τηλεφωνίας της περιοχής της Θεσσαλίας.....	67

Εικόνα 31.:	Βάση δεδομένων του California Institute of Technology.....	69
Εικόνα 32.:	Βάση δεδομένων CBCL StreetScenes Challenge Framework.....	70
Εικόνα 33.:	UIUC Image Database.....	70
Εικόνα 34.:	LISA – Q Front FOV Database.....	70
Εικόνα 35.:	Βάση δεδομένων του University of Central Florida (UCF).....	71
Εικόνα 36.:	DLR Munich Vehicle dataset.....	71
Εικόνα 37.:	Dutch UAS Dataset 001.....	72
Εικόνα 38.:	Overhead Imagery Research Data Set (OIRDS).....	72
Εικόνα 39.:	Stanford Drone Dataset.....	73
Εικόνα 40.:	INRIA Aerial Image Labeling.....	73
Εικόνα 41.:	VEDAI Database.....	74
Εικόνα 42.:	DJI Phantom 3 Pro.....	75
Εικόνα 43.:	Περιοχές βιντεοσκόπησης για τη δημιουργία της VERVwUAS DB....	76
Εικόνα 44.:	Στιγμιότυπα από τα video της VERVwUAS DB.....	77
Εικόνα 45.:	Στιγμιότυπα από τα video της VERVwUAS DB.....	77
Εικόνα 46.:	Στιγμιότυπα από τα video της VERVwUAS DB.....	78
Εικόνα 47.:	Στιγμιότυπα από τα video της VERVwUAS DB.....	79
Εικόνα 48.:	Θετικά δείγματα οχημάτων 4 διαφορετικών μεγεθών με τη διαδικασία της περικοπής.....	83
Εικόνα 49.:	Δείγματα αρνητικών εικόνων.....	84
Εικόνα 50.:	Παράδειγμα annotated εικόνας από το imglab.....	84
Εικόνα 51.:	Αλγόριθμος ανίχνευσης.....	99
Εικόνα 52.:	Αλγόριθμος ιχνηλάτησης οχημάτων.....	100
Εικόνα 53.:	Διαδικασία εκπαίδευσης, ελέγχου και δημιουργίας της VERVwUAS DB.....	101
Εικόνα 54.:	Περιοχές πτήσης για τη μέτρηση της ταχύτητας του δικτύου 3G/4G.....	103
Εικόνα 55.:	Διάταξη VPN για τη ροή video μέσω του δικτύου κινητής τηλεφωνίας.....	104
Εικόνα 56.:	Τοποθέτηση Raspberry Pi3 B+ επί του Ε/Π.....	105
Εικόνα 57.:	Τοποθέτηση του Raspberry στο DJI Phantom.....	106
Εικόνα 58.:	Raspberry Pi Camera Module v2 IMX219.....	107
Εικόνα 59.:	Παραδείγματα από το μέγεθος των Bounding Boxes για κάθε ένα από τα μοντέλα.....	112
Εικόνα 60.:	Παράδειγμα εφαρμογής NMS.....	113

Εικόνα 61.:	Χαρακτηριστικά παραδείγματα της απόδοσης του Haar/BASIC_20stages.....	114
Εικόνα 62.:	Χαρακτηριστικά παραδείγματα της απόδοσης του Haar/CORE_20stages.....	114
Εικόνα 63.:	Χαρακτηριστικά παραδείγματα της απόδοσης του Haar/ALL_20stages.....	115
Εικόνα 64.:	Χαρακτηριστικά παραδείγματα της απόδοσης του Haar/ALL_25stages.....	116
Εικόνα 65.:	Χαρακτηριστικά παραδείγματα της απόδοσης του LBP_20stages..	116
Εικόνα 66.:	Χαρακτηριστικά παραδείγματα της απόδοσης του LBP_25stages..	117
Εικόνα 67.:	Χαρακτηριστικά παραδείγματα της απόδοσης του LBP_25stages_0,45MFAR.....	118
Εικόνα 68.:	Χαρακτηριστικά παραδείγματα της απόδοσης του HOG_GAB_20stages.....	119
Εικόνα 69.:	Χαρακτηριστικά παραδείγματα της απόδοσης του HOG_SVM	119
Εικόνα 70.:	Ανίχνευση σε συνθηκες χαμηλού φωτισμού.....	120
Εικόνα 71.:	Ανίχνευση οχημάτων σε διαφορετική κλίμακα.....	120
Εικόνα 72.:	Συσχέτιση μεταξύ χρόνου επεξεργασίας και του αθροίσματος TP και FP των μοντέλων.....	128
Εικόνα 73.:	Εφαρμογή του μοντέλου HOG_GAB σε φωτογραφία της DLR-MVD	133
Εικόνα 74.:	Εφαρμογή του μοντέλου HOG_SVM σε φωτογραφία της DLR-MVD	133
Εικόνα 75.:	Εφαρμογή του μοντέλου HOG_GAB σε φωτογραφίες της OIRDS dataset.....	134
Εικόνα 76.:	Εφαρμογή του μοντέλου HOG_SVM σε φωτογραφίες της OIRDS dataset.....	134
Εικόνα 77.:	Εφαρμογή του μοντέλου HOG_GAB σε φωτογραφία της VEDAI dataset.....	134
Εικόνα 78.:	Εφαρμογή του μοντέλου HOG_SVM σε φωτογραφία της VEDAI dataset.....	135
Εικόνα 79.:	Ανίχνευση και παρακολούθηση οχήματος μεσω video streaming από E/Π.....	140
Εικόνα 80.:	Ανίχνευση και παρακολούθηση οχήματος μεσω video streaming από UAS.....	141
Εικόνα A-1.	Γραφικό περιβάλλον του imglab.....	A-1
Εικόνα A-2.	Γραφικό περιβάλλον του ObjectMarker.....	A-2
Εικόνα A-3.	Γραφικό περιβάλλον του Cascade Trainer GUI.....	A-3

Λίστα Πινάκων

Πίνακας 1.:	Κατηγοριοποίηση των UAS.....	10
Πίνακας 2.:	Δυναμική Μήτρα Σύγχυσης.....	45
Πίνακας 3.:	Μέγεθος πλαισίου και video σε διαφορετικές αναλύσεις αυτού...	63
Πίνακας 4.:	Προδιαγραφές Phantom 3 Pro.....	75
Πίνακας 5.:	Προδιαγραφές Κάμερας.....	76
Πίνακας 6.:	Βασικά χαρακτηριστικά των Dataset οχημάτων.....	81
Πίνακας 7.:	Στοιχεία του δείγματος εκπαίδευσης στη φάση της δοκιμής.....	85
Πίνακας 8.:	Παράμετροι εκπαίδευσης των Cascades κατά τη φάση των δοκιμών.....	86
Πίνακας 9.:	Στοιχεία εκπαίδευσης των Cascade αλγορίθμων.....	87
Πίνακας 10.:	Αρχικές παράμετροι εκπαίδευσης των LBP_GAB και HAAR_GAB...	91
Πίνακας 11.:	Στοιχεία εκπαίδευσης του LBP_GAB.....	93
Πίνακας 12.:	Στοιχεία εκπαίδευσης του HAAR_GAB.....	94
Πίνακας 13.:	Στοιχεία εκπαίδευσης του HOG_GAB.....	95
Πίνακας 14.:	Χρησιμοποιούμενες παράμετροι για την εκπαίδευση των HAAR/GAB, LBP/GAB και HOG/GAB.....	97
Πίνακας 15.:	Χρησιμοποιούμενες παράμετροι για την εκπαίδευση των HAAR/GAB, LBP/GAB και HOG/GAB.....	98
Πίνακας 16.:	Χρησιμοποιούμενες παράμετροι για την εκπαίδευση του HOG/SVM.....	98
Πίνακας 17.:	Τεχνικά χαρακτηριστικά HUAWAI 4G USB Stick E3372h – 153....	103
Πίνακας 18.:	Μετρήσεις δεικτών απόδοσης των μοντέλων (ανίχνευση οχημάτων στις εικόνες του test set διαστάσεων 1920x1080).....	121
Πίνακας 19.:	Σύγκριση των τιμών TP, FP και FN μεταξύ των μοντέλων (εικόνες 1920x1080).....	122
Πίνακας 20.:	Μετρήσεις δεικτών απόδοσης των μοντέλων (ανίχνευση οχημάτων στις εικόνες του test set διαστάσεων 960x540).....	123
Πίνακας 21.:	Σύγκριση των τιμών TP, FP και FN μεταξύ των μοντέλων (εικόνες 960x540).....	124
Πίνακας 22.:	Διάρκεια χρόνου επεξεργασίας του test set.....	127
Πίνακας 23.:	Συσχέτιση χρόνου επεξεργασίας εικόνας με αριθμό ανιχνεύσεων	128
Πίνακας 24.:	Χρόνος επεξεργασίας εικόνων (1920x1080).....	129
Πίνακας 25.:	Μετρικοί δείκτες του HOG_GAB στο σύνολο των εικόνων του dataset.....	131

Πίνακας 26.:	Μετρικοί δείκτες του HOG_SVM στο σύνολο των εικόνων του dataset.....	131
Πίνακας 27.:	Αποτελέσματα αξιολόγησης ανίχνευσης – παρακολούθησης του μοντέλου HOG_SVM στα test video.....	136
Πίνακας 28.:	Στατιστικά δεδομένα ρυθμού ροής δεδομένων δικτύου κινητής τηλεφωνίας (χρήση UAS).....	137
Πίνακας 29.:	Στατιστικά δεδομένα ρυθμού ροής δεδομένων δικτύου κινητής τηλεφωνίας από Ε/Π (περιοχή πτήσης : Κιλελέρ-Αρμένιο).....	138
Πίνακας 30.:	Στατιστικά δεδομένα ρυθμού ροής δεδομένων δικτύου κινητής τηλεφωνίας από Ε/Π (περιοχή πτήσης : Βόλος).....	138
Πίνακας 31.:	Χαρακτηριστικά αποστολής και λήψης video streaming.....	139
Πίνακας B-1.	Ρυθμός ροής δεδομένων (UAS).....	B-1
Πίνακας B-2.	Ταχύτητα ροής δεδομένων (Download/Upload) εντός του Α/Δ...	B-1
Πίνακας B-3.	Ταχύτητα ροής δεδομένων (Download) στην περιοχή Κιλελέρ – Αρμενίου.....	B-2
Πίνακας B-4.	Ταχύτητα ροής δεδομένων (Upload) στην περιοχή Κιλελέρ – Αρμενίου.....	B-3
Πίνακας B-5.	Ταχύτητα ροής δεδομένων (Download) στην περιοχή Βόλου.....	B-3
Πίνακας B-6.	Ταχύτητα ροής δεδομένων (Upload) στην περιοχή Βόλου.....	B-3

Λίστα Διαγραμμάτων

Διάγραμμα 1.: Διάρκεια εκπαίδευσης του συνόλου των μοντέλων που εκπαιδεύτηκαν.....	109
Διάγραμμα 2.: Διάρκεια εκπαίδευσης των μοντέλων της Φάσης ΙΙβ.....	110
Διάγραμμα 3.: Ακρίβεια προσδιορισμού της θέσης του οχήματος (διαστάσεις εικόνας 1920x1080).....	111
Διάγραμμα 4.: Ακρίβεια προσδιορισμού της θέσης του οχήματος (διαστάσεις εικόνας 960x540).....	112
Διάγραμμα 5.: Μετρήσεις δεικτών απόδοσης των μοντέλων (ανίχνευση οχημάτων στις εικόνες του test set διαστάσεων 1920x1080).....	121
Διάγραμμα 6.: Μετρήσεις δεικτών απόδοσης των μοντέλων (ανίχνευση οχημάτων στις εικόνες του test set διαστάσεων 960x540).....	123
Διάγραμμα 7.: Ιστόγραμμα σύγκρισης τιμών F1.....	125
Διάγραμμα 8.: Ιστόγραμμα απεικόνισης του ποσοστού μείωσης της τιμής F1.....	125
Διάγραμμα 9.: Ιστόγραμμα τιμών F1 στις περιοχές 1-2, 3-4 και 1-2-3-4.....	126
Διάγραμμα 10.: Ιστογράμμο σύγκρισης τιμής δείκτη F1 μεταξύ HOG_GAB και HOG_SVM.....	132

Κεφάλαιο 1 Εισαγωγή

1.1 Πρόλογος

Η επιτήρηση από αέρος μέσω της εναέριας εικονοληψίας, είτε με τη μορφή εικόνας είτε με τη μορφή της βιντεοσκοπήσης, αν και αποτελεί ένα σημαντικό τμήμα των αποστολών που ανατίθενται σε στρατιωτικό περιβάλλον, για τον εντοπισμό και πρόσκτηση των δυνάμεων του αντιπάλου, εν τούτοις, πλέον διαθέτει και ισχυρή παρουσία σε ένα πλήθος εμπορικών εφαρμογών. Τέτοιες εφαρμογές σχετίζονται με την παρακολούθηση της κατάστασης σημαντικών πόρων για την επιβίωση της ανθρωπότητας (π.χ. φυτείες, δάση), την έρευνα και διάσωση, την ζωντανή μετάδοση γεγονότων (live streaming video) από την καθημερινότητα (Kumar et al., 2001), τον έλεγχο της κυκλοφορίας των οχημάτων στον αστικό ιστό και στις εθνικές οδούς (Puri, 2005; Srinivasan, Latchman, Shea, Wong, & McNair, 2004), την αστυνόμευση κλπ.

Ακολουθώντας τα χνάρια της ιστορίας φαίνεται ότι η πρώτη εναέρια λήψη από φωτογραφικό φακό πραγματοποιήθηκε το 1858 από τον Gaspar Felix Tournachon ο οποίος, εκμεταλλευόμενος το αερόστατο, προχώρησε σε αεροφωτογράφιση για σκοπούς που άπτονταν του αντικειμένου της χαρτογράφησης (WordPress, 2018). Η χρήση της εναέριας φωτογράφισης για στρατιωτικούς σκοπούς¹ άρχισε να ωριμάζει με την έναρξη του Α΄ Παγκοσμίου Πολέμου λόγω της δυνατότητας επιχειρησιακής εκμετάλλευση της τρίτης διάστασης που προσέδιδε η χρήση των αεροσκαφών (Α/Φ) σταθερών πτερύγων². Το επόμενο βήμα της ανθρωπότητας στο μονοπάτι της εναέριας

¹ Οι πρώτες αποστολές αφορούσαν στην αναγνώριση, την επιτήρηση και τη συλλογή πληροφοριών.

² Οι Βρετανοί προχώρησαν στη συγκρότηση ειδικού σώματος (Royal Flying Corps) με αποστολή την εναέρια επιτήρηση. Χαρακτηριστικό παράδειγμα είναι ότι το 1915, ως μέρος των

εικονοληψίας πραγματοποιήθηκε με τη λήψη φωτογραφιών από τον δορυφόρο Explorer VI Earth, την 14 Αυγούστου 1959, σε απόσταση 17.000 μιλίων από την επιφάνεια της γης (Great Images in NASA, 2009).

Η εναέρια επιτήρηση πραγματοποιήθηκε αρχικά με λήψη αναλογικών φωτογραφιών η ευκρίνεια των οποίων βελτιώνονταν με την πάροδο των ετών. Η αξιολόγηση των καταγραφών δεν γινόταν σε πραγματικό χρόνο αλλά λάμβανε χώρα μετά την προσγείωση των αεροσκαφών, οπότε και εξειδικευμένο προσωπικό πραγματοποιούσε μελέτη και ανάλυση των εικόνων, προκειμένου να εξάγει τις πληροφορίες ενδιαφέροντος (Kumar et al., 2001). Το επόμενο βήμα στην εναέρια εικονοληψία πραγματοποιήθηκε με την εισαγωγή των καμερών λήψης video ενώ η πλέον σημαντική εξέλιξη σηματοδοτήθηκε με την αντικατάσταση του φιλμ από αισθητήρες³ οι οποίοι μετέτρεπαν το φως σε ηλεκτρικό σήμα. Έτσι, η επαγόμενη ψηφιοποίηση της εικονοληψίας, ήταν αυτή που ουσιαστικά άνοιξε το Κουτί της Πανδώρας για την επεξεργασία της εικόνας με τη χρήση ηλεκτρονικών υπολογιστών.

Η εφεύρεση των μικροεπεξεργαστών τη δεκαετία του 1970 (F. L. Xiao, 2013), η εξέλιξη τους σύμφωνα με το νόμο του Moore⁴, αλλά και η συνεπαγόμενη μείωση του κόστους⁵ και του μεγέθους αυτών, κατέστησε εφικτή, σε μικρό χρονικό διάστημα, τη μαζική πρόσβαση του κοινού στους ηλεκτρονικούς υπολογιστές⁶. Οι μικροεπεξεργαστές όμως δεν παρέμειναν "καθηλωμένοι" εντός του σκληρού περιβλήματος των υπολογιστικών συστημάτων αλλά κατέκλυσαν κάθε συσκευή της καθημερινότητας με αποτέλεσμα να εισαχθούν στο λεξιλόγιο νέοι όροι όπως αυτοί της κινητής και διάχυτης υπολογιστικής⁷. Η εμπορική εφαρμογή αυτών των τάσεων δεν άφησε ανεπηρέαστο το πεδίο της εικονοληψίας με τις κάμερες να έχουν ενσωματωθεί σε κινητά τηλέφωνα ή να

προπαρασκευαστικών εργασιών για την μάχη της Neuve Chapelle, πέτυχαν την καταγραφή με φωτογραφικό φακό του συνόλου των Γερμανικών χαρακωμάτων (Northstar Gallery, 1998).

³ CMOS ή CCD

⁴ Σύμφωνα με το νόμο του Moore, ο ρυθμός αύξησης των τρανζιστορ ενός ολοκληρωμένου κυκλώματος διπλασιάζεται κάθε 2 έτη

⁵ Το 1993 η τιμή ενός εκ των κορυφαίων μοντέλων της Dell (ο H/Y 486 με επεξεργαστή 66MHz, 8 MB RAM και σκληρό δίσκο χωρητικότητας 320-megabyte) ανέρχονταν στα 4.400\$ (<http://www.relativelyinteresting.com/comparing-todays-computers-to-1995s/>). Πρόσβαση:27/12/17

⁶ Μέχρι το τέλος του 1998 περισσότερα από το 40% των νοικοκυριών στις ΗΠΑ διέθετε προσωπικό υπολογιστή (TELEC, L 1999:xv)

⁷ Οι όροι Ubiquitous Computing and Pervasive Computing χρησιμοποιήθηκαν για πρώτη φορά από τις Xerox PARC και την IBM αντίστοιχα προκειμένου να αποτυπώσουν το γεγονός ότι η υπολογιστική ξέφευγε από το στενό περιβάλλον των προσωπικών υπολογιστών και επεκτείνονταν σε κινητές συσκευές (Want & Pering, 2005).

κατασκευάζονται σε τέτοια μεγέθη φορητότητας (πχ κάμερες GoPro), ώστε να είναι εύκολη η προσαρμογή τους στον εξοπλισμό που χρησιμοποιείται κατά τη διεξαγωγή κάθε μορφής ανθρώπινης δραστηριότητας (κατάδυση, ποδηλασία, ορειβασία κλπ).

Από την άλλη, η ανάπτυξη από τον Tim Bernes – Lee του παγκόσμιου ιστού (world wide web) στις αρχές του 1990, με σκοπό να διασυνδέσει τα ερευνητικά πανεπιστημιακά κέντρα (Berners-Lee, 1992), ήταν η σπίθα για αυτό που θα ξεσπούσε στη συνέχεια. Σε λιγότερο από 10 έτη⁸ ο πληθυσμός που έκανε χρήση του διαδικτύου έφτασε τα 300 εκατ. (Miniwatts Marketing Group, 2017) ενώ ο αριθμός των ιστοσελίδων ξεπέρασε τα 17 εκατ. (W3C, 2018). Παρά την αντιπαράθεση για την ισχύ ή όχι του νόμου του Metcalfe⁹ (Briscoe, Odlyzko, & Tilly, 2006), το βέβαιο είναι ότι η αύξηση του αριθμού των χρηστών του διαδικτύου επηρεάστηκε τόσο από την ανάπτυξη ασύρματων πρωτοκόλλων δικτύωσης και επικοινωνίας¹⁰, όσο και από την εξέλιξη του τρόπου χρήσης των υπολογιστών. Οι τελευταίοι, από μια συσκευή γραφείου (mainframe), την οποία διαμοιράζονταν πολλοί χρήστες, μετατράπηκαν σε διάχυτα (ubiquitous) συστήματα με αποτέλεσμα ένα άτομο να κατέχει πληθώρα κινητών συσκευών (Want et al., 2010) με δυνατότητες δικτύωσης, επικοινωνίας και επεξεργασίας δεδομένων, οπουδήποτε και οποτεδήποτε (ibid).

Όλα τα παραπάνω, που αποτελούν βασικά στοιχεία της ψηφιακής και πληροφοριακής - επικοινωνιακής¹¹ επανάστασης που βιώνεται από το τέλος του προηγούμενου αιώνα, επηρέασαν με τον πιο δραστικό τρόπο το πεδίο της εναέριας υπολογιστικής όρασης (aerial computer vision), που αποτελεί και το αντικείμενο της συγκεκριμένης μεταπτυχιακής διατριβής. Αυτό εξηγείται από το γεγονός ότι βασικοί πυλώνες για την έρευνα στο εν λόγω πεδίο είναι υλοποιήσιμοι πλέον με την καταβολή ενός σχετικά μικρού αντιτίμου, καθόσον (1) η εναέρια εικονοληψία δεν προϋποθέτει αποκλειστική χρήση Α/Φ ή Ε/Π αλλά αυτή είναι εφικτή και από drone τα οποία δύνανται να φέρουν

⁸ Από τον Αύγουστο του 1991 έως το 2000

⁹ Ο Robert Metcalfe, συνεφευρέτης του Ethernet, θεώρησε ότι η αξία του διαδικτύου αυξάνει αναλογικά με τον αριθμό των χρηστών.

¹⁰ WLAN, 3G/4G. Στα συγκεκριμένα δίκτυα, η αύξηση του εύρους ζώνης ακολούθησε τον νόμο του Moore (Fettweis & Alamouti, 2014).

¹¹ Ο χρησιμοποιούμενος όρος είναι Information and Communication Technology (ICT). Αφορά ένα ευρύ πεδίο εφαρμογών το οποίο σχετίζεται με τους υπολογιστές, τον επικοινωνιακό εξοπλισμό και τις υπηρεσίες που σχετίζονται με αυτούς τους δύο τομείς. Ενδεικτικά περιλαμβάνει την τηλεφωνία, τα κυψελοειδή δίκτυα, τις δορυφορικές επικοινωνίες κλπ. (https://en.wikibooks.org/wiki/The_Information_Age/The_Digital_and_ICT_Revolutions).

Πρόσβαση : 12/12/17

κάμερες αντίστοιχου μεγέθους (2) υπάρχει η δυνατότητα προμήθειας φορητών Η/Υ, για την επεξεργασία του λαμβανόμενου video, η υπολογιστική ισχύς των οποίων ξεπερνά αυτή των υπερυπολογιστών της δεκαετίας του 1990 και τέλος (3) η δυνατότητα δικτύωσης των παραπάνω συσκευών είτε μέσω WLAN είτε με εκμετάλλευση των κυψελοειδών δικτύων κινητής τηλεφωνίας 3G/4G/5G δίνει τη δυνατότητα μετάδοσης της εικόνας σε πραγματικό χρόνο.

1.2 Σκοπός – Ερευνητικά Ερωτήματα

Ο σκοπός της παρούσης μεταπτυχιακής διατριβής είναι διττός και στοχεύει:

- Σε πρώτη φάση, να μελετήσει τις δυνατότητες χρησιμοποίησης εικόνας video που λαμβάνεται από μη επανδρωμένα εναέρια μέσα (UAV) για σκοπούς αναγνώρισης και παρακολούθησης αντικειμένων
- Σε δεύτερη φάση να διερευνήσει τη δυνατότητα μετάδοσης δεδομένων, με τη μορφή ροής video από την εναέρια πλατφόρμα σε επίγειο σταθμό εδάφους, για περεταίρω επεξεργασία και εκμετάλλευση.

Τα βασικά ερευνητικά ερωτήματα προέρχονται από την μελέτη (1) των υφιστάμενων τεχνικών αναγνώρισης αντικειμένων¹², και (2) την αντιμετώπιση των προκλήσεων που ενέχει η μετάδοση εικόνας video από εναέρια πλατφόρμες σε επίγειους σταθμούς Συγκεκριμένα:

- Ποιος από τους παραδοσιακούς αλγόριθμους ανίχνευσης αντικειμένων (HAAR – LBP – HOG) σε συνδυασμό με τους αντίστοιχους κατηγοριοποιητές (AdaBoost – SVM) αποδίδει καλύτερα από πλευράς αναγνώρισης;
- Ποιος από τους παραπάνω αλγόριθμους είναι ταχύτερος ώστε να δύναται να υιοθετηθεί σε εφαρμογές σχεδόν πραγματικού χρόνου;
- Επαρκεί η χρήση αλγορίθμου ανίχνευσης ή αυτός πρέπει να συνδυαστεί και με αλγόριθμο ιχνηλάτησης αντικειμένου προκειμένου να εξοικονομηθούν υπολογιστικοί πόροι;
- Υποστηρίζονται οι απαιτούμενοι ρυθμοί μετάδοσης video από το δίκτυο 3G για την ολοκληρωμένη (end-to-end) εφαρμογή του προτεινόμενου μοντέλου ή εφαρμόζεται μόνο σε περιοχές που υπάρχει κάλυψη 4G(LTE) από τον πάροχο κινητής τηλεφωνίας;

¹² Το εν λόγω ζήτημα άπτεται του ερευνητικού πεδίου του computer vision.

Το ερευνητικό ενδιαφέρον της μεταπτυχιακής διατριβής εστιάζεται στην ανίχνευση και αυτόματη ιχνηλάτηση οχημάτων, τα οποία βιντεοσκοποούνται από αέρος, με τη χρήση τεχνικών και αλγορίθμων του computer vision. Το γενικό πλαίσιο της εργασίας συνίσταται στην καταγραφή εικόνας video από κάμερα η οποία είναι προσαρμοσμένη σε UAV ή Ε/Π. Η εν λόγω καταγραφή αποτελεί ταυτόχρονα και τα δεδομένα εισόδου σε ένα λογισμικό το οποίο:

- Μεταβιβάζει τα δεδομένα ανίχνευσης στο λογισμικό ιχνηλάτισης (tracker)
- Αναλόγως των υφιστάμενων συνθηκών, μπορεί να επιλεγθεί να λειτουργεί είτε on-line σε σχεδόν πραγματικό χρόνο (near real time)¹³ είτε off-line.

Οι "υφιστάμενες συνθήκες" για τις οποίες έγινε αναφορά στην προηγούμενη παράγραφο χαρακτηρίζονται:

- Από την αδυναμία των UAV, τα οποία χρησιμοποιήθηκαν, να φέρουν ενσωματωμένο πληροφοριακό εξοπλισμό ο οποίος διαθέτει την απαιτούμενη υπολογιστική ισχύ, για την επίτευξη onboard-online επεξεργασίας.
- Από την ικανότητα των UAV να μεταδίδουν σε επίγειο σταθμό video υψηλής ανάλυσης (συνήθως μέσω Wi-Fi σύνδεσης) σε συνδυασμό με την δυνατότητα χρησιμοποίησης στο έδαφος πληροφοριακού εξοπλισμού με μεγάλη υπολογιστική ισχύ.
- Από την απαίτηση τα UAV να ίπτανται εντός συγκεκριμένων χωρικών ορίων, σε ότι αφορά στην απόστασή τους από το σταθμό ελέγχου¹⁴, ώστε να είναι δυνατή η μετάδοση τόσο των εντολών κατεύθυνσης όσο και του καταγραφόμενου video (αφορά περιπτώσεις near real time ανίχνευσης).
- Από τη δυνατότητα των UAV να ίπτανται αυτόνομα σε προκαθορισμένη διαδρομή πτήσης, χωρίς την απαίτηση επικοινωνίας με τον επίγειο σταθμό εδάφους. Κατά τη διάρκεια της πτήσης πραγματοποιείται καταγραφή και αποθήκευση του video το οποίο στη συνέχεια, μετά την προσγείωση του UAV, ανακτάται και επεξεργάζεται (περίπτωση off line επεξεργασίας).
- Από την δυνατότητα των Ε/Π (1) να φέρουν πληροφοριακά συστήματα υψηλών υπολογιστικών δυνατοτήτων στα οποία καθίσταται ευχερής η on Board - OnLine επεξεργασία του video που καταγράφουν και (2) να

¹³ Ουσιαστικά πρόκειται για σχεδόν πραγματικό χρόνο (near real time) λόγω υστέρησης

¹⁴ Συνήθως επιδιώκεται η ύπαρξη οπτικής επαφής (Line of Sight – LOS) μεταξύ UAV και σταθμού ελέγχου.

μεταδώσουν σε επίγειο σταθμό εδάφους, σε σχεδόν πραγματικό χρόνο και χρησιμοποιώντας πληθώρα διαφορετικών ασύρματων επικοινωνιακών συστημάτων (κυψελοειδή συστήματα GSM - 3/4G, ασύρματοι VHF - UHF, δορυφορικοί ασύρματοι κλπ) του καταγραφόμενου video που σχετίζεται με την ανίχνευση και ιχνηλάτηση αντικειμένων ενδιαφέροντος. Ένα σημαντικό στοιχείο που σχετίζεται με την εναέρια επιτήρηση από Ε/Π, είναι η καταλυτική προσθήκη του ανθρώπινου παράγοντα (human in the loop) στην όλη διαδικασία (λήψη video, χρησιμοποιούμενοι αλγόριθμοι, επιλογή μεθόδου αποστολής του video σε επίγειο σταθμό σύμφωνα με τις τρέχουσες κάθε φορά δυνατότητες του κυψελοειδούς δικτύου κινητής τηλεφωνίας, εκμετάλλευση δορυφορικών επικοινωνιών κλπ).

1.3 Κίνητρο - Αιτιολόγηση του Θέματος

Οι Shapiro και Stockman, στο βιβλίο τους με τίτλο "Computer Vision" που εκδόθηκε πριν από δεκαεπτά έτη με σκοπό να εισάγει το ευρύ κοινό στο ομώνυμο επιστημονικό πεδίο, αναφέρει ότι: «οι εφαρμογές των υπολογιστών στην ανάλυση εικόνας είναι ουσιαστικά απεριόριστες» (Shapiro & Stockman, 2001). Μετά την παρέλευση 13 ετών και ο Klette συντονίστηκε στο ίδιο μήκος κύματος επισημαίνοντας ότι «η πρόοδος που έχει προσφάτως επιτευχθεί στην τεχνολογία των καμερών, των υπολογιστών καθώς επίσης και στο θεωρητικό υπόβαθρο του computer vision έχει καταστήσει το τελευταίο, σημαντικό μέρος των τεχνολογιών οι οποίες για την υλοποίησή τους χρησιμοποιούν ένα πλήθος διαφορετικών πεδίων» (Klette, 2014).

Έχοντας ως σημείο εκκίνησης το αυξημένο ερευνητικό ενδιαφέρον που παρουσιάζει τα τελευταία έτη το πεδίο του Computer Vision και αφού μελετήθηκε η αντίστοιχη βιβλιογραφία διαπιστώθηκε ότι:

- Μεγάλο μέρος αυτής πραγματεύεται το συγκεκριμένο αντικείμενο από τη σκοπιά της εναέριας εικονοληψίας. Ο Xiao και άλλοι προτείνουν μία πιθανολογική γραφική προσέγγιση για την ταυτόχρονη ανίχνευση και ιχνηλάτηση μεγάλου αριθμού οχημάτων σε video με χαμηλό ρυθμό πλαισίων (J. Xiao, Cheng, Sawhney, & Han, 2010). Η Gaszczak και άλλοι προσέγγισαν το θέμα της ανίχνευσης και ιχνηλάτισης οχημάτων και ανθρώπων με τη χρησιμοποίηση κατηγοριοποιητή Haar, ενώ για την μείωση των θετικών ή αρνητικών σφαλμάτων ανίχνευσης εκμεταλλεύονται το αποτέλεσμα της σύγκρισης

εικόνων στο οπτικό και υπέρυθρο φάσμα (Gaszczak, Breckon, & Han, 2011). Ο Breckon και άλλοι χρησιμοποίησαν επίσης κατηγοριοποιητή Haar για την ανίχνευση οχημάτων αλλά με ρυθμό εικονοληψίας 1Hz (Breckon, Barnes, Eichner, & Wahren, 2009). Τέλος, ο Gleason και άλλοι μελέτησαν την ανίχνευση οχημάτων στην ύπαιθρο μέσω της σύγκρισης δύο διαφορετικών μεθόδων εξαγωγής χαρακτηριστικών, του Histogram of Oriented Gradients και του Histogram of Gabor Coefficients (Gleason, Nefian, Bouysounousse, Fong, & Bebis, 2011).

- Δίδεται μεγαλύτερη βαρύτητα στην ανίχνευση και ιχνηλάτηση οχημάτων εντός του αστικού ιστού, για λόγους που σχετίζονται με τη ρύθμιση του κυκλοφοριακού προβλήματος. Ο Yang και άλλοι προτείνουν μία μέθοδο η οποία για την ανίχνευση των κινούμενων οχημάτων χρησιμοποιεί από κοινού τους αλγορίθμους Scale Invariant Feature Transform (SIFT) and Kanada-Lucas-Tomasi (KLT), ενώ η ανίχνευση των στατικών οχημάτων επιτυγχάνεται κάνοντας χρήση των πληροφοριών blob (Yang, Liu, Wang, Luo, & Liu, 2012). Ο Cao και άλλοι προσεγγίζουν το ίδιο θέμα μέσω της ανάλυσης των πολλαπλών επιπέδων κίνησης, όπως αυτά εξάγονται από την τροχιά που ακολουθεί το κάθε όχημα (Cao, Lan, Yan, & Li, 2012).
- Σε ότι αφορά στην ασύρματη μετάδοση εικόνας από εναέρια πλατφόρμα σε επίγειο σταθμό, ο Puri, προσεγγίζοντας το θέμα της επιτήρησης του κυκλοφοριακού με τη βοήθεια μη επανδρωμένου οχήματος, προτείνει την μετάδοση δεδομένων και εικόνας με τη χρήση ασύρματου δικτύου που λειτουργεί στα 2,4 GHz¹⁵ (Puri, 2005). Από την άλλη, ο Dong και άλλοι μελέτησαν την ασύρματη μετάδοση εικόνας από UAV εκμεταλλευόμενοι το δίκτυο της κινητής τηλεφωνίας (Dong, Chen, Cai, & Peng, 2007).

Υπό το πρίσμα όσων αναφέρθηκαν, η βιβλιογραφία είναι αρκετά πλούσια αναφορικά με την ανίχνευση και ιχνηλάτηση οχημάτων ή την ασύρματη μετάδοση video από εναέριες πλατφόρμες. Όμως, όλες οι προσεγγίσεις μελετούν τα συγκεκριμένα ζητήματα ξεχωριστά και με συγκεκριμένες προϋποθέσεις ως προς τον εξοπλισμό, τη μέθοδο κλπ. Στη συγκεκριμένη μεταπτυχιακή διατριβή επιχειρείται μια πιο ευέλικτη προσέγγιση που εξετάζει συνδυαστικά τα παρακάτω ζητήματα:

¹⁵ Εκμεταλλεύεται τις υφιστάμενες επίγειες κεραίες ιδιοκτησίας του Τμήματος Μεταφορών της Πολιτείας της Φλώριδας

- Μελέτη των υφιστάμενων δυνατοτήτων μετάδοσης εικόνας (video streaming), από τις εναέριες πλατφόρμες προς επίγειο σταθμό, με τη χρήση του δικτύου κινητής τηλεφωνίας, σύμφωνα πάντα με το διαθέσιμο σε πραγματικό χρόνο bandwidth που προσφέρεται άνωθεν του ηπειρωτικού ή του θαλάσσιου χώρου.
- Επεξεργασίας εικόνας video σε σχεδόν πραγματικό χρόνο για την μελέτη των υφιστάμενων τεχνικών αυτόματης ανίχνευσης και ιχνηλάτισης στατικών ή κινούμενων οχημάτων από εναέρια μέσα, για σκοπούς επιτήρησης περιοχών (πχ εντοπισμό οχημάτων που εισέρχονται σε προστατευόμενη δασική έκταση).

1.4 Γενική Περιγραφή των Αντικειμένων στα οποία Στηρίχθηκε η Μεταπτυχιακή Διατριβή

Για τη διεξαγωγή της συγκεκριμένης μεταπτυχιακής διατριβής πραγματοποιήθηκε επεξεργασία video, που συλλέγεται από κάμερα τοποθετημένη σε εναέριο μέσο επιτήρησης, χρησιμοποιώντας τεχνικές του computer vision. Η επεξεργασία του video πραγματοποιείται σε επίγειο σταθμό ο οποίος λαμβάνει live streaming, μέσω του δικτύου της κινητής τηλεφωνίας. Για τη μείωση του κόστους, χρησιμοποιήθηκε μη επανδρωμένο αερο-οχημα προκειμένου να λάβουν χώρα οι εναέριες λήψεις οι οποίες στη συνέχεια χρησιμοποιήθηκαν για την εκπαίδευση του αλγόριθμου αυτόματης ανίχνευσης οχημάτων. Στη συνέχεια θα παρουσιαστούν υπό τύπο εισαγωγικού σχολιασμού τα δύο αντικείμενα στα οποία στηρίχθηκε η μεταπτυχιακή διατριβή και τα οποία ουσιαστικά συγκροτούν και τον τίτλο αυτής (Computer Vision for Aerial applications).

1.4.1 Εναέρια Μέσα Επιτήρησης (Ελικόπτερα - Μη Επανδρωμένα Εναέρια Συστήματα)

Το ελικόπτερο (Ε/Π) είναι ένα Α/Φ περιστρεφόμενων πτερύγων¹⁶ στην κλίση των οποίων στηρίζεται η κίνησή του στις τρεις διαστάσεις. Οι πρώτες πτήσεις Ε/Π εντοπίζονται στις αρχές του προηγούμενου αιώνα με το πρώτο επιχειρησιακό Ε/Π να παράγεται το 1936 (Goebel, 2017). Η πρώτη ευρεία χρησιμοποίηση Ε/Π πραγματοποιήθηκε στον πόλεμο της Κορέας όπου ανέλαβε αποστολές έρευνας και διάσωσης, διακομιδής απωλειών υγείας και αερομεταφοράς προσωπικού και υλικού (Heikell, 2005).

¹⁶ Σε αντιδιαστολή με τα αεροσκάφη (Α/Φ) σταθερών πτερύγων.

Πλέον τα Ε/Π χρησιμοποιούνται μαζικά στο στρατιωτικό περιβάλλον, σε δημόσιες υπηρεσίες καθώς επίσης και στον ιδιωτικό τομέα. Μία κοινή αποστολή που αναλαμβάνουν και στις τρεις αυτές περιπτώσεις είναι η επιτήρηση. Σύμφωνα με τον Διακλαδικό Κανονισμό Ορολογίας των Ενόπλων Δυνάμεων των ΗΠΑ, «επιτήρηση είναι η συστηματική παρατήρηση του εναέριου χώρου, της επιφάνειας (επίγειας και θαλάσσιας) ή υπό αυτής περιοχής, χώρων, προσώπων ή αντικειμένων, με τη χρήση οπτικών, ακουστικών, ηλεκτρονικών, φωτογραφικών ή άλλων μέσων» (US Joint Chiefs of Staff, 15 February 2016).

Από την άλλη, τα μη Επανδρωμένα Εναέρια Συστήματα (Unmanned Aerial Systems – UAS) αποτελούν τα τελευταία έτη ένα ενεργό ερευνητικό αντικείμενο (Van Blyenburgh, 1999) καθώς επίσης και έναν ταχέως αναπτυσσόμενο εμπορικό τομέα (Divya, 2017). Είναι γνωστά με ένα πλήθος διαφορετικών ονομασιών όπως “Unmanned Aerial Vehicles (UAV)”, “drones” και “aerial robots” (Colomina & Molina, 2014). Ο Διεθνής Οργανισμός Πολιτικής Αεροπλοΐας (International Civil Aviation Organization – ICAO) έχει εισάγει την έννοια του τηλεχειριζόμενου αερο-συστήματος (RPAS - Remotely - Piloted Aerial System) με το σκεπτικό ότι μόνο τα συγκεκριμένα συστήματα θα δύνανται να ενταχθούν στο σύστημα της διεθνούς πολιτικής αεροπλοΐας (Circular, 2011).

Σύμφωνα με τον ορισμό του UAV που παρέχεται από το λεξικό της Οξφόρδης: «πρόκειται για αερο-οχήματα των οποίων ο έλεγχος ασκείται είτε με τηλεχειρισμό είτε από ενσωματωμένο υπολογιστή» (Oxford University Press, 2017). Ο Austin θεωρεί το συγκεκριμένο ορισμό υπεραπλουστευμένο και αναφέρεται στο UAV «ως σύστημα το οποίο απαρτίζεται από αριθμό υποσυστημάτων που περιλαμβάνουν το αεροσκάφος (συχνά αναφέρεται ως UAS ή μη επανδρωμένο αερόχημα), το ωφέλιμο φορτίο του (payload), το σύστημα ή συστήματα ελέγχου (συμπεριλαμβάνονται και οι σταθμοί τηλεχειρισμού), τα συστήματα/πλατφόρμες εκτόξευσης - επανάκτησης, το υποσύστημα επικοινωνίας κλπ» (Austin, 2011). Επίσης, διαχωρίζει τα drone από τα UAV υποστηρίζοντας ότι τα πρώτα διαθέτουν μηδενική “νοημοσύνη” εκτελώντας απλώς ένα προκαθορισμένο ίχνος πτήσης, σε αντίθεση με τα UAV που διαθέτουν κατ’ ελάχιστο δυνατότητα επικοινωνίας με το σταθμό εδάφους, μεταφέροντας και λαμβάνοντας πληθώρα δεδομένων (ibid).

Αναλόγως της χρήσης του UAV ή του drone καθορίζεται και το ωφέλιμο φορτίο του ή ο εξοπλισμός που θα διαθέτει. Ο μόνος περιορισμός σχετίζεται με το μέγεθος και το βάρος αυτού (Vergouw, Nagel, Bondt, & Custers, 2016). Η χρήση των UAS για σκοπούς πέραν των στρατιωτικών εφαρμογών δύναται να ενταχθεί σε έξι μεγάλες κατηγορίες (1) στην επιστημονική έρευνα, (2) στην προστασία από καταστροφές και διαχείριση αυτών, (3) στην προστασία του περιβάλλοντος, (4) στην αποκατάσταση επικοινωνιών μεταξύ σημείων χωρίς οπτική επαφή, (5) στην προστασία κρίσιμων εγκαταστάσεων και (6) στην αστυνόμευση (Skrzypietz, 2012). Ουσιαστικά τα UAV αναλαμβάνουν πλέον ένα μεγάλο μέρος των αποστολών στις οποίες μέχρι και τα τέλη του προηγούμενου αιώνα χρησιμοποιούνταν επανδρωμένες πλατφόρμες (Α/Φ ή Ε/Π)

ΚΑΤΗΓΟΡΙΑ	ΒΕΛΗΝΕΚΕΣ (KM)	ΥΨΟΣ ΠΤΗΣΗΣ (m)	ΑΥΤΟΝΟΜΙΑ (h)	ΜΕΓΙΣΤΟ ΒΑΡΟΣ ΑΠΟΓΕΙΩΣΗΣ (Kg)
Micro & Mini UAV (MUAV)	<10	300	<2	<30
Medium Altitude Long Endurance (MALE)	>500	15.000	24 - 48	1.500 - 7000
High Altitude Long Endurance (HALE)	>2000	20.000	24 - 48	4.500 - 15.000
Vertical Take -off and Landing (VTOL) UAV	x - 204	x-6.100	0.18 - 8	0.019 - 1.400

Πίνακας 1. Κατηγοριοποίηση των UAS (Skrzypietz, 2012)

1.4.2 Computer Vision

Στη βιβλιογραφία, ο όρος computer vision χρησιμοποιείται αρκετές φορές εναλλακτικά του machine vision (Davies, 2004; Jain, Kasturi, & Schunck, 1995), ενώ από την μελέτη της, δε βρέθηκε να υπάρχει ένας τυποποιημένος ορισμός. Σύμφωνα με τους Shapiro και Stockman, «στόχος του computer vision είναι να εξάγει χρήσιμες αποφάσεις για τα πραγματικά φυσικά αντικείμενα και σκηνές, βάσει των δεδομένων που γίνονται αντιληπτά μέσω μιας εικόνας» (Shapiro & Stockman, 2001). Ο Klette θεωρεί ότι «το αντικείμενο του computer vision αποσκοπεί στην χρήση καμερών για την ανάλυση και κατανόηση των σκηνών του πραγματικού κόσμου, ενώ ο αντίστοιχος επιστημονικός κλάδος μελετά μεθοδολογικά και αλγοριθμικά προβλήματα καθώς επίσης και θέματα που σχετίζονται με την εφαρμογή των σχεδιαζόμενων λύσεων (Klette, 2014). Ο Jain και άλλοι αναφέρουν ότι «ο σκοπός του machine vision είναι η δημιουργία ενός μοντέλου του πραγματικού κόσμου μέσα από τις εικόνες» και πιο συγκεκριμένα «η ανάκτηση χρήσιμων πληροφοριών για μία σκηνή, από τις δισδιάστατες προβολές της» (Jain et al., 1995). Ο Szeliski υποστηρίζει ότι «με τη χρήση των τεχνικών του computer vision

επιδιώκουμε να περιγράψουμε τον "κόσμο" που βλέπουμε σε μία ή περισσότερες εικόνες και να αναδομήσουμε – αναπαραστήσουμε ιδιότητές του, όπως το σχήμα, η φωτεινότητα και η χρωματική κατανομή (Szeliski, 2010). Τέλος ο Davies, αναφερόμενος στο διαχωρισμό των εννοιών του computer και machine vision σημειώνει ότι «computer vision είναι η επιστήμη της όρασης ενώ machine vision η μελέτη των μεθόδων, τεχνικών και υλικού μέσω της οποίας μπορεί να επιτευχθεί τεχνητή (artificial) όραση για πρακτικές εφαρμογές» και ολοκληρώνει το συλλογισμό του θεωρώντας ότι «πρόσφατα, η διάθεση ταχύτερων υπολογιστών έχει οδηγήσει στη σύγκλιση της χρήσης των δύο παραπάνω όρων» (Davies, 2012) .

Σύμφωνα με τον Klette, τα τελευταία χρόνια το computer vision αποτελεί τεχνολογία κλειδί για πολλά πεδία» (Klette, 2014), παρά το γεγονός ότι πρόκειται για ένα σχετικά νέο πεδίο έρευνας (Davies, 2012). Αυτό αποδεικνύεται και από την καθημερινότητα όπου πολλές φορές, χωρίς να γίνεται αντιληπτό, συναντάται πλήθος εφαρμογών οι οποίες κάνουν χρήση του υπόψη επιστημονικού πεδίου. Χαρακτηριστικά παραδείγματα αποτελούν η αυτόματη αναγνώριση προσώπου που διαθέτει η φωτογραφική μηχανή των κινητών τηλεφώνων, τα συστήματα αναγνώρισης γραμμωτού κώδικα (bar code readers), η ικανότητα αυτόματης αναγνώρισης της λωρίδας κυκλοφορίας που διαθέτουν τα σύγχρονα οχήματα, ο προσδιορισμός της κατάστασης υγείας ενός ηλικιωμένου ατόμου που παρέχεται από τα "έξυπνα" συστήματα οικιακής ιατρικής φροντίδας κλπ. Σημαντική είναι όμως και η εκμετάλλευση των δεδομένων που δύναται να εξάγει ένα σύστημα, μέσω της εφαρμογής των τεχνικών του computer vision, τα οποία μπορεί στη συνέχεια να δρομολογηθούν για στρατιωτική χρήση (πχ επιτήρηση ευαίσθητων χώρων), αστυνόμευση δημοσίων χώρων (πχ αναγνώριση συμπεριφοράς πλήθους), αύξηση της απόδοσης της βιομηχανικής παραγωγής (πχ αυτοματοποιημένος έλεγχος) κλπ.

Οι ρίζες του computer vision εντοπίζονται στις αρχές του 1970 οπότε ερευνητές στους τομείς της τεχνητής νοημοσύνης και της ρομποτικής, στα πανεπιστήμια του MIT, Stanford και CMU, προσπάθησαν να επιλύσουν το πρόβλημα της αυτόματης αναγνώρισης αντικειμένων μέσω της εικόνας που καταγράφει μια κάμερα (Szeliski, 2010). Την συγκεκριμένη εποχή βρισκόταν σε πλήρη ανάπτυξη το πεδίο της επεξεργασίας εικόνας ενώ για να διαχωριστεί ο τομέας του computer vision από αυτή, ως διακριτό όριο τέθηκε η στόχευση του computer vision στην προσπάθεια για τρισδιάστατη

αναπαραγωγή του χώρου με πληροφορίες που αντλούνταν από εικόνες¹⁷(ibid).

Αφού οι ερευνητές μελέτησαν χαμηλού επιπέδου αντικείμενα όπως η τμηματοποίηση της εικόνας και η ανίχνευση άκρων (Huang, 1996), στη συνέχεια και εντός της δεκαετίας του 1980, η έρευνα γύρω από το αντικείμενο του computer vision αρχίζει να σημειώνει εκθετική πρόοδο με την εμφάνιση νέων θεωριών και εννοιών. Μία από τις πρώτες εφαρμογές σε αυτή την κατεύθυνση ήταν η οπτική αναγνώριση χαρακτήρων (Optical character recognition - OCR) η οποία χρησιμοποιήθηκε για την αυτόματη αναγνώριση γραμμάτων, συμβόλων και αριθμών.

Η δεκαετία του 1990 βρήκε το πεδίο του machine vision να συνδέεται στενά με την αντίστοιχη βιομηχανία¹⁸ ενώ ταυτόχρονα η πρόοδος που σημειώνονταν στην λειτουργία των αισθητήρων και την αρχιτεκτονική ελέγχου, επέτεινε περαιτέρω τις δυνατότητες των υπόψη συστημάτων, μειώνοντας ταυτόχρονα το κόστος τους (EPIC Systems, 2017). Επίσης, εντός της ίδιας δεκαετίας εμφανίστηκαν οι πρώτες στατιστικές μέθοδοι αναγνώρισης προσώπων σε εικόνες (Szeliski, 2010) ενώ το πλέον αξιοσημείωτο στοιχείο προόδου για αυτή τη δεκαετία θεωρήθηκε η αλληλεπίδραση με τα γραφικά των υπολογιστών¹⁹ (Seitz & Szeliski, 1999).

Τη δεκαετία του 2000 μία σημαντική τάση στον τομέα του computer vision, στην οποία στηρίζεται και μέρος της συγκεκριμένης μεταπτυχιακής διατριβής, αποτέλεσε η εφαρμογή τεχνικών βασισμένων στα χαρακτηριστικά ενός αντικειμένου²⁰ η οποία, συνδυαζόμενη με αλγόριθμους εκμάθησης, οδήγησε στην αυτόματη αναγνώριση αντικειμένων (Ponce, Hebert, Schmid, & Zisserman, 2007). Σύμφωνα με τον Szeliski «η τελευταία τάση σε αυτόν τον τομέα, η οποία και κυριαρχεί στην έρευνα για την οπτική αναγνώριση, είναι η εφαρμογή εξελιγμένων τεχνικών μηχανικής εκμάθησης²¹, για την επίλυση των αντίστοιχων προβλημάτων» (Szeliski, 2010). Είναι σημαντικό να αναφερθεί ότι η όλη διαδικασία ανάπτυξης αλγορίθμων για την αναγνώριση αντικειμένων, έλαβε ουσιαστική ώθηση από τη σταδιακή ανάπτυξη και αύξηση χρήσης του internet σε συνδυασμό με την επίτευξη όλο και μεγαλύτερων ταχυτήτων

¹⁷ Πατέρας του computer vision θεωρείται ο Larry Roberts ο οποίος μελέτησε τις δυνατότητες εξαγωγής τρισδιάστατων σχημάτων μέσω της δισδιάστατης προοπτικής (Huang, 1996).

¹⁸ Πάνω από 100 εταιρίες άρχισαν να προμηθεύουν machine vision συστήματα

¹⁹ Αφορά στις τεχνικές image morphing

²⁰ Feature-Based Techniques

²¹ Machine/Deep Learning

μεταφοράς δεδομένων, ως αποτέλεσμα της συνεχούς αύξησης του εύρους ζώνης. Ο συνδυασμός των παραπάνω γεγονότων, επέτρεψε την ευχερή διακίνηση βάσεων δεδομένων με εικόνες (EPIC Systems, 2017). Τέλος μία νέα τάση σε ότι αφορά σε αυτές τις βάσεις, είναι η επισήμανση της θέσης διαφόρων αντικειμένων στις φωτογραφίες με αποτέλεσμα να καθίσταται πιο εύκολη η διαδικασία εκμάθησης, χωρίς να απαιτείται ταυτόχρονα η συνεχής ανθρώπινη παρουσία (Szeliski, 2010).

Γιατί όμως παρουσιάζει δυσκολίες το computer vision; Ποιες είναι οι προκλήσεις που αντιμετωπίζει ένας ερευνητής; Ο Szeliski θεωρεί ότι αυτό οφείλεται, εν μέρη, στο γεγονός ότι πρόκειται για ένα αντίστροφο πρόβλημα στο οποίο:

- «Αναζητείται η ανάκτηση ορισμένων άγνωστων στοιχείων έχοντας όμως δεδομένα τα οποία δεν επαρκούν για την πλήρη επίλυση του τιθέμενου προβλήματος» (ibid).
- «Μέσω του computer vision προσπαθούμε να περιγράψουμε την σκηνή που απεικονίζεται σε μία ή περισσότερες εικόνες και να αναδομήσουμε τα χαρακτηριστικά της όπως το σχήμα, το φωτισμό ή την χρωματική κατανομή» (ibid).

Με άλλα λόγια και αναλύοντας την παραπάνω θέση του Szeliski, μέσω του computer vision γίνεται προσπάθεια να δοθεί απάντηση σε τρεις βασικές ερωτήσεις. Η πρώτη αφορά στην αναγνώριση ή κατηγοριοποίηση των αντικειμένων, δηλαδή δίδεται η απάντηση στο ερώτημα: "τι αντικείμενα υπάρχουν σε μια εικόνα;" (detection). Η δεύτερη σχετίζεται με τον καθορισμό της θέσης ενός αντικειμένου σε μια εικόνα, δηλαδή "που βρίσκεται το αντικείμενο;" (localization). Συνήθως η απάντηση αναπαρίσταται τοποθετώντας ένα περίγραμμα, συνήθως ορθογώνιο, γύρω από το αντικείμενο ενδιαφέροντος. Η τελευταία ερώτηση επιδιώκει να αντλήσει πληροφορίες για τη σκηνή ώστε, αφού καθοριστεί η θέση ενός αντικειμένου εντός αυτής, να δοθεί απάντηση στο "πώς αυτό θα μεταβεί από μία θέση A σε μία θέση B" (πχ συστήματα αυτόματης ναυτιλίας για ρομπότ, αυτοκίνητα, UAV κλπ).

Οι ερευνητές, στις απαρχές της τεχνητής νοημοσύνης, εδραζόμενοι στον τρόπο με τον οποίο λειτουργεί η ανθρώπινη όραση και η διαδικασία επεξεργασίας της εικόνας από

τον εγκέφαλο²², θεωρούσαν ότι το γνωστικό μέρος της νοημοσύνης (cognitive) ήταν εγγενώς δυσκολότερο από το αντιληπτικό (perceptual)»(Boden, 2006). Έτσι όμως, ίσως υποτιμήθηκαν οι δυσχέρειες που παρουσιάζει το πεδίο του computer vision²³. Αυτές οι δυσκολίες εστιάζονται:

- Στο ζήτημα των διαστάσεων και του όγκου των δεδομένων που προκύπτουν. Αυτές αποτυπώνονται από τις διαστάσεις μιας εικόνας και τα δεδομένα της τιμής του κάθε pixel που φέρει. Ενώ σε μια μικρής ανάλυσης εικόνα μπορεί να χρησιμοποιηθεί η ευκλείδεια απόσταση για την εύρεση της ομοιότητας μεταξύ δύο αντικειμένων, η ίδια μέθοδος δεν θα αποδώσει σε μία υψηλής ανάλυσης εικόνα (1920 X 1080)²⁴. Το πρόβλημα γίνεται ακόμη δυσκολότερο αν η τιμή των pixel της συγκεκριμένης εικόνας δεν είναι δυαδική (άσπρο – μαύρο) αλλά δίδεται σε διαβαθμίσεις του τόνου μεταξύ άσπρου και μαύρου (συνήθως αναπαρίσταται με 256 τιμές) ή ακόμα περισσότερο αν η φωτογραφία δεν είναι ασπρόμαυρη αλλά έγχρωμη (οπότε μπορεί να έχουμε 256 τιμές για κάθε ένα από τα τρία βασικά χρώματα).
- Στη διαφορετικότητα κάθε εικόνας. Συνεχίζοντας την παραπάνω συλλογιστική, ένα στοιχείο που δυσκολεύει ακόμα περισσότερο το πρόβλημα είναι ότι το ίδιο αντικείμενο ενδιαφέροντος “περιγράφεται” με διαφορετικές τιμές pixel σε κάθε φωτογραφία. Έστω ότι γίνεται προσπάθεια “αναγνώρισης” του προσώπου της Selena Gomez. Μία απλή αναζήτηση στο διαδίκτυο θα δείξει την ύπαρξη πλήθους “διαφορετικών προσώπων” που οφείλονται στη διαφορετική κόμμωση, το make-up, το καλλιτεχνικό ύφος της φωτογράφισης, τη θέση του

²² «Ο ανθρώπινος εγκέφαλος διαθέτει πάνω από 10^{11} νευρώνες, με το μέσο όρο συνάψεων μεταξύ τους να ξεπερνά τις 7.000 (Drachman, 2005). Αν κάθε νευρώνας λειτουργεί ως μικροεπεξεργαστής τότε προκύπτει ένας “υπερυπολογιστής” του οποίου τα στοιχεία επεξεργασίας λειτουργούν ταυτόχρονα» (Davies, 2012).

²³ Μία “ματιά” περιλαμβάνει δεδομένα της τάξεως των megabits ενώ το “εύρος ζώνης” μεταξύ οφθαλμού και εγκεφάλου πιθανώς υπερβαίνει τα 10 Mbit/s (Davies, 2012). Ο εγκέφαλος, προσπαθώντας να ερμηνεύσει τις εικόνες που δέχεται από το σύστημα όρασης, λειτουργεί αφαιρετικά, διαθέτοντας ρουτίνες για την επεξεργασία των δεδομένων που λαμβάνει και κάνοντας χρήση συγκεκριμένων υποθέσεων, οι οποίες όμως είναι σε μεγάλο βαθμό υπεύθυνες για τη δημιουργία ψευδαισθήσεων (ibid).

²⁴ Για να κατανοηθεί το πρόβλημα της διάστασης των δεδομένων, αν έχουμε μία εικόνα διαστάσεων $m \times n$ τα pixel της οποίας παίρνουν τιμές 0 και 1 τότε το πλήθος των πιθανών τιμών που προκύπτουν από αυτήν την εικόνα είναι: $2^{m \times n}$. Οπότε το να ψάξουμε για τις πιθανές απαντήσεις, διερευνώντας συστηματικά μία – μία αυτές τις τιμές, είναι κάτι πρακτικά αδύνατο. Αλλά ακόμα και ο απλός χειρισμός των δεδομένων είναι δύσκολος λόγω του όγκου τους καθώς μία εικόνα υψηλής ανάλυσης αποτελείται από περισσότερα από 2 εκατ. pixel (1920 * 1080 = 2.073.600)

προσώπου ως προς το φακό κλπ. Αυτά τα χαρακτηριστικά, παρά το γεγονός ότι θα εντοπιστούν από την ανθρώπινη όραση, εν τούτοις, λόγω της αφαιρετικής λειτουργίας του, θα αγνοηθούν ως ένα βαθμό από τον ανθρώπινο εγκέφαλο ο οποίος θα αναγνωρίσει την τραγουδίστρια βάσει συγκεκριμένων και γενικών χαρακτηριστικών του προσώπου της. Το γεγονός ότι οι υπολογιστές δεν έχουν την ίδια μορφή αφαιρετικότητας με τον εγκέφαλο αλλά “διαβάζουν” τιμές από pixel τις οποίες επεξεργάζονται στη συνέχεια με διάφορους αλγόριθμους, έχει ως αποτέλεσμα να δυσκολεύονται πολύ περισσότερο και από ένα 5χρονο κοριτσάκι στο να δώσουν τη σωστή απάντηση. Σύμφωνα με τον Moses και άλλους, οι διαφορές που προκύπτουν για έναν υπολογιστή σε φωτογραφίες του ίδιου προσώπου, οι οποίες όμως λήφθηκαν με διαφορετική ένταση και θέση φωτισμού, είναι πολλές φορές περισσότερες από το αν άλλαζε το ίδιο το πρόσωπο (Moses, Adini, & Ullman, 1994).



Εικόνα 1. Επίδραση φωτισμού στην τιμή των pixel (Moses, Adini, & Ullman, 1994)

- Σε μη ελεγχόμενους παράγοντες όπως η παραμόρφωση ή ο θόρυβος αλλά και περιορισμούς που σχετίζονται με (1) την απώλεια πληροφοριών λόγω του ότι με μια εικόνα μεταβαίνουμε από τις 3-διαστάσεις (σκηνή πραγματικού κόσμου) στις 2-διαστάσεις (φωτογραφία), (2) τη συμπίεση της φωτογραφίας, (3) τη χρησιμοποιούμενη τεχνική, (4) τα λαμβανόμενα δείγματα, (5) την ποικιλομορφία και αντιπροσωπευτικότητα αυτών, (6) την υπολογιστική ισχύ κλπ (Davies, 2012).

Οι δυσκολίες που παρουσιάζει η επίλυση των προβλημάτων στο πεδίο του computer vision, σε συνδυασμό με την αμεσότητα και την πολλαπλότητα των εφαρμογών που έχει σε κάθε τομέα της ανθρώπινης δραστηριότητας, έχει καταστήσει το συγκεκριμένο αντικείμενο μία ενεργή περιοχή από απόψεως ερευνητικού ενδιαφέροντος. Εντός αυτού του πλαισίου, η υπόψη μεταπτυχιακή διατριβή θα μελετήσει τη χρήση αλγορίθμων για την αναγνώριση, εντοπισμό και ιχνηλάτηση οχημάτων, μέσω video το οποίο λαμβάνεται από εναέριο μέσο με την κάμερα να διοπτρεύει κάθετα το έδαφος.

1.5 Δομή Εργασίας

Η μεταπτυχιακή διατριβή είναι οργανωμένη στα παρακάτω κεφάλαια:

- **Κεφάλαιο 2ο : Ανιχνευτές Αντικειμένων και Κατηγοριοποιητές.** Σε αυτό το κεφάλαιο παρέχεται το θεωρητικό υπόβαθρο της ανίχνευσης αντικειμένων, μέσω της ανάλυσης των αλγορίθμων που συνθέτουν τα μοντέλα ανίχνευσης που θα εξεταστούν. Συγκεκριμένα, αναλύονται 3 τεχνικές παρουσίασης χαρακτηριστικών σημείων (feature representations: Haar, Local Binary Patterns και Histogram of Oriented Gradients) και 2 κατηγοριοποιητές που συνδυάζονται κατά περίπτωση με τις εν λόγω τεχνικές (Gentle AdaBoost και Support Vector Machine). Στο τέλος του κεφαλαίου παρουσιάζονται οι μετρικοί δείκτες που χρησιμοποιούνται για την αξιολόγηση της απόδοσης των μοντέλων ανίχνευσης.
- **Κεφάλαιο 3ο : Ιχνηλάτηση Αντικειμένων μέσω Video Streaming.** Σε αυτό το κεφάλαιο γίνεται εισαγωγή στην έννοια της ιχνηλάτησης και παρουσιάζεται ο αλγόριθμος που θα χρησιμοποιηθεί. Στη συνέχεια παρέχονται βασικά στοιχεία που σχετίζονται με τη λειτουργία του video και του video streaming. Τέλος γίνεται αναφορά στα δίκτυα κινητής τηλεφωνίας και στους ρυθμούς μετάδοσης δεδομένων που παρέχουν.
- **Κεφάλαιο 4ο : Βάσεις Δεδομένων.** Στο τέταρτο κεφάλαιο παρουσιάζονται οι βάσεις δεδομένων που υπάρχουν στο διαδίκτυο και οι οποίες σχετίζονται με την ανίχνευση οχημάτων. Επίσης γίνεται παρουσίαση της βάσης δεδομένων με εναέριες λήψεις οχημάτων που δημιουργήθηκε για τους σκοπούς της συγκεκριμένης μεταπτυχιακής διατριβής.
- **Κεφάλαιο 5ο : Μεθοδολογία.** Στο συγκεκριμένο κεφάλαιο αναλύεται η διαδικασία που ακολουθήθηκε για τη δημιουργία των μοντέλων ανίχνευσης οχημάτων. Επίσης γίνεται αναφορά και στα εργαλεία που θα χρησιμοποιηθούν για την αξιολόγηση των μοντέλων ανίχνευσης και ιχνηλάτησης οχημάτων.

- **Κεφάλαιο 6ο : Αποτελέσματα.** Σε αυτό το κεφάλαιο αναφέρονται τα πειραματικά αποτελέσματα που σχετίζονται με την απόδοση των μοντέλων στις διαδικασίες ανίχνευσης και ιχνηλάτισης οχημάτων.
- **Κεφάλαιο 7ο : Συμπεράσματα.** Στο έβδομο κεφάλαιο αναφέρονται τα συμπεράσματα της έρευνας.
- **Κεφάλαιο 8ο :** Το συγκεκριμένο κεφάλαιο αποτελεί τον επίλογο και γίνονται κάποιες προτάσεις για μελλοντική επέκταση της παρούσας μεταπτυχιακής διατριβής.

Κεφάλαιο

2

Ανιχνευτές Αντικειμένων και Κατηγοριοποιητές

2.1 Αναγνώριση και Ανίχνευση

Η έρευνα των Viola και Jones που αφορά στην ανίχνευση με τη χρήση κατηγοριοποιητών ακολουθίας (cascade), ουσιαστικά αποτελεί μία τεχνική κατηγοριοποίησης (categorization) αντικειμένων και ως εργασία διαφέρει από αυτή της αναγνώρισης (recognition) ενός αντικειμένου. Σύμφωνα με τους Vedaldi και Zisserman «Ο σκοπός της αναγνώρισης αντικειμένου είναι η αντιστοίχιση (αναγνώριση) ενός συγκεκριμένου αντικειμένου σε μια εικόνα/σκηνή (Howse, Puttemans, Hua, & Sinha, 2015). Παραδείγματα περιλαμβάνουν την αναγνώριση ενός συγκεκριμένου κτιρίου, όπως του πύργου της Πίζα, ή ενός πίνακα ζωγραφικής, για παράδειγμα της Mona Liza. Το αντικείμενο αναγνωρίζεται παρά τις μεταβολές στην κλίμακα απεικόνισης, τη γωνία λήψης, τις συνθήκες φωτισμού και τις περιπτώσεις μερικής απόκρυψής του» (ibid). Αυτό σημαίνει ότι τα συγκεκριμένα αντικείμενα έχουν ακριβή περιγραφή. Όταν άλλα αντικείμενα επιδιώκεται να αντιστοιχηθούν με αυτά τότε συνήθως τίθεται ένα ελάχιστο κατώφλι χαρακτηριστικών τα οποία εάν δε ικανοποιηθούν το αντικείμενο δεν αντιστοιχίζεται.

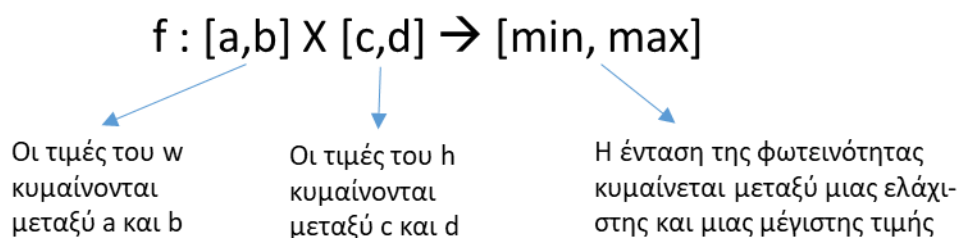
Η εργασία της ανίχνευσης (detection) «πηγαίνει την όλη διαδικασία ένα βήμα περαιτέρω, καθώς επιδιώκεται η εύρεση ενός αντικειμένου, σε ποικίλες διατάξεις αυτού, μέσω της εκμάθησης μιας πιο αντικειμενοστραφούς περιγραφής, αντί μιας απλής περιγραφής της ίδιας της εικόνας που περιλαμβάνει το αντικείμενο. Στις περιπτώσεις που το προς ανίχνευση αντικείμενο καταστεί πιο πολύπλοκο (η ποικιλομορφία στις διάφορες εκφάνσεις του αντικειμένου μπορεί να είναι αρκετά μεγάλη και να μεγαλώνει

ακόμα περισσότερο όσο προστίθενται νέες εικόνες του), τότε δε γίνεται λόγος πλέον για την ανίχνευση ενός καθ' εαυτού αντικειμένου αλλά για την ανίχνευση της κλάσης που αντιπροσωπεύουν τα υπόψη αντικείμενα. Έτσι, μέσω της κατηγοριοποίησης επιδιώκεται η εκμάθηση ενός γενικού μοντέλου, για την κλάση στην οποία εντάσσεται ένα αντικείμενο, λαμβάνοντας υπόψη την ποικιλομορφία των αντικειμένων της κλάσης» (ibid). Για παράδειγμα στην περίπτωση της ανίχνευσης οχημάτων, η κλάση είναι τα αυτοκίνητα και στιγμιότυπα ή εκφάνσεις είναι τα διαφορετικά είδη αυτών (station, sedan κλπ) ή το διαφορετικό σχήμα τους.

Οι διαφορετικοί αλγόριθμοι που θα χρησιμοποιηθούν, και οι οποίοι αναλύονται στις επόμενες ενότητες, αποσκοπούν στην δημιουργία ενός μοντέλου το οποίο θα προέρχεται από διαφορετικά στιγμιότυπα, αναφορικά με το σχήμα, χρωματισμό, φωτεινότητα κλπ. της ίδιας κλάσης. Η εφαρμογή των συγκεκριμένων μοντέλων σε μια εικόνα καταλήγει να δίνει απάντηση στο ερώτημα "Που είναι το αντικείμενο ενδιαφέροντος εντός της εικόνας;". Συνήθως η ανίχνευση του αντικειμένου απεικονίζεται με το να περικλείεται το αντικείμενο ενδιαφέροντος εντός ενός τετράπλευρου ή ενός κύκλου.

2.2 Επεξεργασία Ψηφιακής Εικόνας

Η εικόνα αποτελεί μια συνάρτηση $I(w, h)$ όπου το I σχετίζεται με την ένταση της φωτεινότητας (image intensity). Το βασικό στοιχείο από το οποίο συντίθεται μια εικόνα είναι το pixel. Κάθε pixel μπορεί να περιγραφεί με τρεις τιμές από τις οποίες οι δύο αφορούν στην θέση του επάνω στη φωτογραφία και η τρίτη στην φωτεινότητά του. Με γνώμονα τα παραπάνω και επειδή η εικόνα ορίζεται ως ένα ορθογώνιο παραλληλόγραμμο πεπερασμένων διαστάσεων, γίνεται κατανοητό ότι η συνάρτηση $I(w, h)$ δίνει την τιμή της φωτεινότητας του pixel στη δεδομένη θέση (w, h) της εικόνας. Όταν οι τιμές των w, h και έντασης I είναι πεπερασμένες και διακριτές, τότε η εικόνα ορίζεται ως ψηφιακή (Wasseem,).



Εικόνα 2. Η συνάρτηση f μιας εικόνας

2.2.1 Βασικά Χαρακτηριστικά Εικόνας

Οι εικόνες διακρίνονται σε τέσσερις βασικούς τύπους, τις δυαδικές (binary), αυτές που αποτελούν διαβαθμίσεις της κλίμακας του γκρι χρώματος (grayscale), τις έγχρωμες (color) και τις πολυφασματικές (multispectral)²⁵ (Wasseem). Οι πρώτες αποτελούν την πιο απλή μορφή ψηφιακής εικόνας καθώς το κάθε pixel μπορεί να λάβει την τιμή 0 ή 1²⁶. Στη δεύτερη περίπτωση (grayscale), το κάθε pixel μπορεί να λάβει ένα εύρος τιμών το οποίο εξαρτάται από τις ορισθείσες διαβαθμίσεις της κλίμακας του γκρι όπου στο ένα άκρο αυτής είναι το μαύρο χρώμα και στο άλλο άκρο στο άσπρο. Το μέγεθος της κλίμακας εξαρτάται από το αριθμό των bits που θα διατεθεί για την αναπαράσταση του χρωματικού κώδικα. Για παράδειγμα, κάθε pixel μίας grayscale εικόνας 8-bit, αναφέρεται σε 256 τιμές διαβάθμισης ($2^8=256$) της κλίμακας του γκρι (το μαύρο λαμβάνει την τιμή 0 και στην άλλη άκρη της κλίμακας το άσπρο την τιμή 255). Για τις έγχρωμες εικόνες, η αντίστοιχη συνάρτηση λαμβάνει τη μορφή:

$$I(x, y) = \begin{bmatrix} r(x, y) \\ g(x, y) \\ b(x, y) \end{bmatrix}$$

με το κάθε pixel να αποτελεί άνυσμα τριών αριθμών όπου κάθε αριθμός αντιστοιχεί σε ένα από τα τρία κανάλια: κόκκινο (red), πράσινο (green) και μπλε (blue).



Εικόνα 3. Οι τρεις τύποι εικόνων:δυαδική, κλίμακας του γκρι και έγχρωμη (κατά σειρά από τα αριστερά προς τα δεξιά)

Το φάσμα της χρωματικής διαβάθμισης που λαμβάνει το κάθε pixel εξαρτάται από τον αριθμό των bits που καταχωρείται σε αυτό (bpp – bits per pixel) και ισούται με 2^{bpp} . Με βάση τα παραπάνω, το μέγεθος S μιας εικόνας δίνεται από τη σχέση $S = w * h * bpp$, όπου w και h ο αριθμός στηλών και γραμμών αντίστοιχα. Η τιμή του bpp εξαρτάται από

²⁵ Οι πολυφασματικές περιλαμβάνουν πληροφορίες οι οποίες, αν και είναι εκτός του ανθρώπινου οπτικού φάσματος (αφορούν στο IR, UV κλπ), εν τούτοις, απεικονίζονται με έναν από τους άλλους τρεις τύπους

²⁶ Η τιμή 0 αποδίδεται για το μαύρο χρώμα και 1 για το άσπρο.

το αν η φωτογραφία είναι δυαδική, κλίμακας του γκρι ή έγχρωμη καθώς επίσης και από το χρωματικό μοντέλο που χρησιμοποιείται (RGB, HSI, CMYK κλπ).

Όπως αναφέρθηκε, το pixel ορίζεται ως το μικρότερο "δομικό" στοιχείο μιας εικόνας, στο οποίο αποδίδεται μία τιμή φωτεινότητας. Ο όρος ανάλυση της εικόνας (pixel resolution) αναφέρεται στο συνολικό αριθμό των pixel ο οποίος προκύπτει από το γινόμενο του πλάτους w της εικόνας (αριθμός των pixel που σχηματίζουν τις στήλες) με το ύψος h αυτής (αριθμός των pixel που σχηματίζουν τις γραμμές). Η ανάλυση της εικόνας διατηρεί ανάλογη σχέση με την ποιότητα αυτής αλλά και με το μέγεθος S της μνήμης που καταλαμβάνει. Μία επιπλέον έννοια η οποία σχετίζεται με την ανάλυση της εικόνας είναι η λόγος του πλάτους w προς το ύψος h (aspect ratio). Οι πιο συνηθισμένες αναλογίες είναι οι 1.33:1 και 1,77:1.

2.2.2 Θόρυβος

Ο θόρυβος (noise) αποτελεί επίσης μία συνάρτηση η οποία συνδυαζόμενη με αυτή της εικόνας δίνει την ακόλουθη σχέση : $\vec{I}'(w, h) = \vec{I}(w, h) + \vec{n}(w, h)$. Υπάρχουν διάφορα είδη θορύβου. Ο Salt and Pepper noise είναι χαρακτηριστικός καθώς "προσθέτει" στην εικόνα ασπρόμαυρες κουκίδες, ο Impulse noise χαρακτηρίζεται από άσπρες κουκίδες και ο Gaussian noise διακρίνεται από μεταβολές στη φωτεινότητα.



Εικόνα 4. Είδη θορύβου (Bastian)

Πολλές φορές ο θόρυβος "καταστρέφει" την πληροφορία ιδιαίτερα όταν προστίθεται με

pixel οι τιμές των οποίων βρίσκονται πλησίον του ανώτερου ορίου φωτεινότητας (όταν το άθροισμα είναι εκτός του καθοριζόμενου ορίου η τιμή περιορίζεται στο ανώτερο όριο). Το παραπάνω, σε συνδυασμό με το γεγονός ότι για την αφαίρεση του θορύβου πρέπει να είναι γνωστή η συνάρτηση δημιουργίας του, έχει ως αποτέλεσμα η υπόψη διαδικασία να καθίσταται πρακτικά ανέφικτη. Από την άλλη, επειδή όμως είναι γνωστή η στατιστική κατανομή του θορύβου είναι δυνατή η εφαρμογή κάποιων γραμμικών φίλτρων που στοχεύουν στην μείωσή του. Έτσι, στην περίπτωση του Gaussian noise είναι δυνατή η εφαρμογή ενός Gaussian averaging φίλτρου για την αφαίρεσή του ενώ αντίστοιχα για το θόρυβο Salt and Pepper εφαρμόζονται Median φίλτρα.

2.2.3 Features – Descriptors

Ένα σημαντικό στοιχείο το οποίο, αφενός μεν σχετίζεται με την επεξεργασία της εικόνας, αφετέρου δε, αποτελεί ερευνητικό αντικείμενο της μηχανικής όρασης, είναι τα "χαρακτηριστικά σημεία (features)". Η υιοθέτηση του συγκεκριμένου όρου προέκυψε από την απαίτηση να εντοπιστούν σε μία εικόνα εκείνα τα σημεία του αντικείμενου ενδιαφέροντος τα οποία θα δύναται να διακρίνονται και να ανιχνεύονται, χρησιμοποιώντας μία μέθοδο η οποία θα προσφέρει αξιοπιστία αναφορικά με την ακρίβεια εντοπισμού τους. Η αξιοπιστία έγγυται στο γεγονός ότι τα ίδια σημεία, για το ίδιο αντικείμενο, θα πρέπει να εντοπιστούν και σε μία διαφορετική εικόνα.

Το παραπάνω πρόβλημα, της ανίχνευσης των ίδιων σημείων και στις δύο εικόνες, επιλύεται με τη βοήθεια ενός επαναλαμβανόμενου ανιχνευτή (repeatable detector) και στη συνέχεια με αντιστοίχιση αυτών των σημείων μεταξύ των δύο εικόνων. Η δεύτερη ενέργεια πραγματοποιείται με τη χρήση ενός συστήματος περιγραφής (descriptor). Τα features χρησιμοποιούνται για την ευθυγράμμιση εικόνων, την 3D ανακατασκευή του χώρου αλλά και για την ανίχνευση αντικειμένων (το τελευταίο αποτελεί αντικείμενο που πραγματεύεται η συγκεκριμένη μεταπτυχιακή διατριβή). Τα βασικά χαρακτηριστικά που πρέπει να πληροί ένα feature είναι (Bobick, 2011–2018).

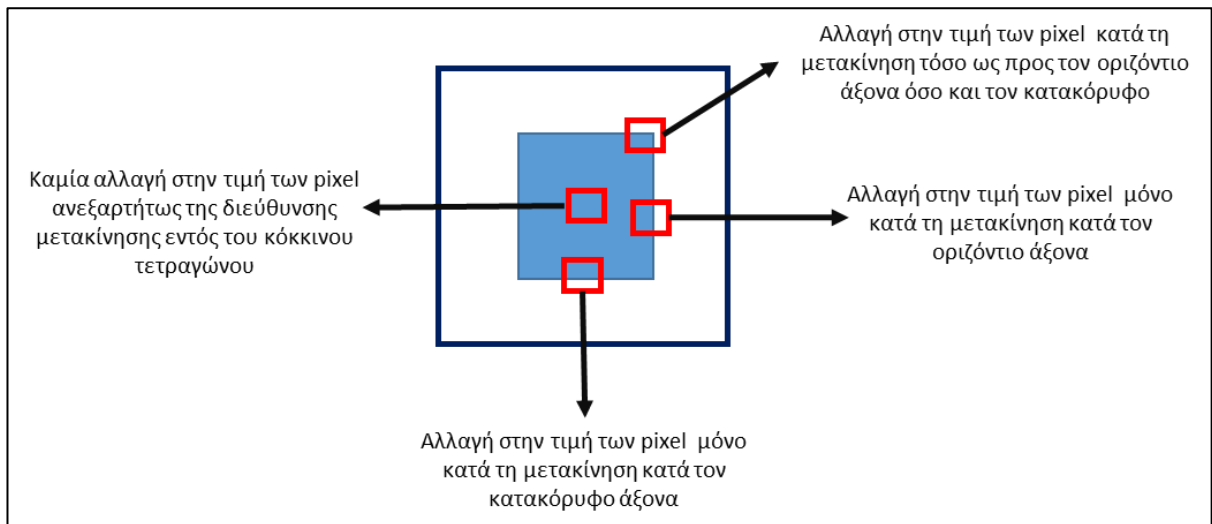
- Επαναληψιμότητα/Ακρίβεια : Το αντικείμενο ενδιαφέροντος μιας εικόνας θα πρέπει να δύναται να εντοπιστεί και σε άλλες εικόνες ανεξαρτήτως γεωμετρικών ή φωτομετρικών μετασχηματισμών.
- Χαρακτηριστικότητα : Κάθε feature θα πρέπει να "διαθέτει" ξεχωριστή και διακριτή "υπογραφή" ώστε να δύναται να περιγραφεί με μοναδιαίο τρόπο
- Αποδοτικότητα : Πρέπει τα χαρακτηριστικά σημεία να είναι πολύ λιγότερα σε

αριθμό, σε σχέση με τον αριθμό των pixel της εικόνας

- Τοπικά Χαρακτηριστικά : Το feature πρέπει να καταλαμβάνει μικρή περιοχή στην εικόνα ώστε να παρέχει ευρωστία στα παρασιτικά σήματα (clutter) και στις περιπτώσεις απόκρυψης του αντικειμένου.

2.2.4 Corner Detector

Η βασική οντότητα πάνω στην οποία στηρίζεται η εύρεση των features είναι η ανίχνευση γωνιών (corner detection) που αποτέλεσε και το ερευνητικό αντικείμενο των Harris και Stephens. Αυτό συμβαίνει επειδή, όπως φαίνεται στην εικόνα 5, στις γωνίες των αντικειμένων λαμβάνει χώρα σημαντική αλλαγή στην τιμή των pixel, ανεξαρτήτως της διεύθυνσης μετακίνησης.



Εικόνα 5. Σημαντικότητα των γωνιών

Η μεθοδολογία των Harris Corners, λαμβάνει υπόψη τη σχέση (Harris & Stephens, 1988):

$$E(u, v) = \sum_{x,y} w(x, y) [I(x + u, y + v) - I(x, y)]^2$$

όπου E το σφάλμα, I είναι η ένταση της φωτεινότητας (image intensity), ενώ το $w(x, y)$ αποτελεί ένα παράθυρο το οποίο μετακινείται κατά μία απόσταση u, v . Έτσι η $I(x, y)$ εκφράζει την ένταση της φωτεινότητας στο αρχικό παράθυρο και η $I(x+u, y+v)$ την αντίστοιχη τιμή στο μετακινούμενο παράθυρο. Από τη συγκεκριμένη σχέση, και προκειμένου να εξαχθεί το συμπέρασμα ότι το παραπάνω παράθυρο βρίσκεται σε μια περιοχή που υπάρχει γωνία, πρέπει να αναμένεται σημαντική μεταβολή στην τιμή της ανεξαρτήτως κατεύθυνσης μετατόπισης (ακόμα και για μικρές αλλαγές). Προκειμένου να προβλεφθεί η τιμή της παραπάνω συνάρτησης για μικρές μετακινήσεις λαμβάνεται

η δεύτερη σειρά Taylor του $E(u, v)$ με τα (u, v) να τείνουν στο μηδέν :

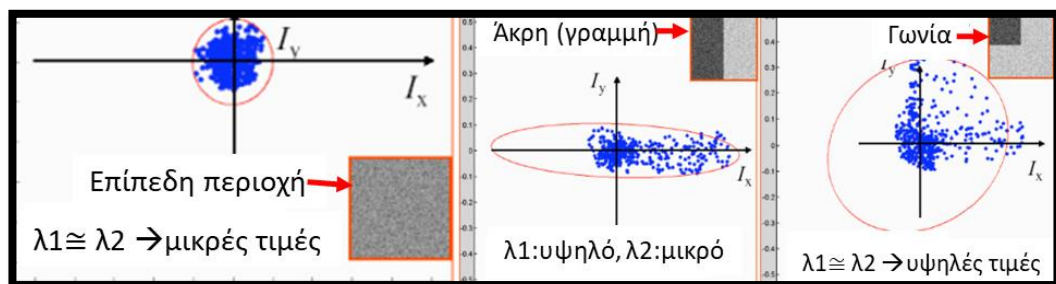
$$E(u, v) \cong E(0,0) + [u, v] \begin{bmatrix} E_{u(0,0)} \\ E_{v(0,0)} \end{bmatrix} + \frac{1}{2} [u, v] \begin{bmatrix} E_{uu(0,0)} & E_{uv(0,0)} \\ E_{uv(0,0)} & E_{vv(0,0)} \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}$$

Λόγω του ότι τα $(u, v) \rightarrow 0$, τα $E(0,0)$, $E_{u(0,0)}$ και $E_{v(0,0)}$ επίσης θα τείνουν στο μηδέν. Οπότε, η παραπάνω γραμμική εξίσωση λαμβάνει τη μορφή:

$$E(u, v) \cong [u, v] \begin{bmatrix} E_{uu(0,0)} & E_{uv(0,0)} \\ E_{uv(0,0)} & E_{vv(0,0)} \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}$$

Με εφαρμογή της προσέγγισης τετραγωνικού βαθμού γίνεται $E(u, v) \cong [u, v]M \begin{bmatrix} u \\ v \end{bmatrix}$. Το M (second moment matrix) προκύπτει από το σταθμισμένο άθροισμα τετραγώνων καθώς και το γινόμενο των αντίστοιχων παραγώγων, κατά τους άξονες x και y , που προκύπτουν από την εικόνα που εξετάζεται. Έτσι:

$$M = \sum_{x,y} w(x, y) \begin{bmatrix} I_x^2 & I_x * I_y \\ I_x * I_y & I_y^2 \end{bmatrix} = \sum \begin{bmatrix} \Sigma I_x I_x & \Sigma I_x I_y \\ \Sigma I_x I_y & \Sigma I_y I_y \end{bmatrix} \\ = \sum \left(\begin{bmatrix} I_x \\ I_y \end{bmatrix} [I_x \quad I_y] \right) = \sum \nabla I * (\nabla I)^T$$



Εικόνα 6. Η κατηγοριοποίηση των σημείων μιας εικόνας, χρησιμοποιώντας τις τιμές ιδιοτιμές λ_1 και λ_2 (eigenvalues) του M καταδεικνύουν (1) περιοχή χωρίς γραμμές και γωνίες όταν τα λ_1 και λ_2 λαμβάνουν μικρές τιμές (2) γραμμικά όρια όταν $\lambda_2 \gg \lambda_1$ ή $\lambda_1 \gg \lambda_2$ και (3) γωνίες όταν τα λ_1 και λ_2 λαμβάνουν υψηλές τιμές (Collins,).

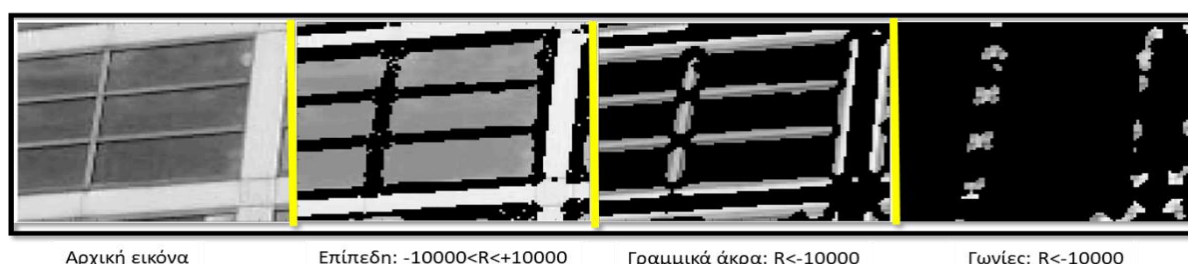
Η παραπάνω σχέση του M δύναται να κατανοηθεί με διαισθαντικό τρόπο αποτυπώνοντας τα ανύσματα της κλίσης (gradient vectors), ως ένα σύνολο (dx, dy)

σημείων, τοποθετημένα σε ένα καρτεσιανού τύπου σύστημα συντεταγμένων με σημείο εκκίνησης έστω το σημείο (0,0). Αναλόγως της τιμής της κλίσης λαμβάνονται τα αποτελέσματα της εικόνας 6.

Από τα παραπάνω προκύπτει η συνάρτηση Corner Response σύμφωνα με την οποία: η ορίζουσα (determinant-det) του M δίνεται από το γινόμενο των ιδιοτιμών της (eigenvalues) ενώ το ίχνος (trace) αυτού από το τετράγωνο του αθροίσματος αυτών:

$$R = \det(M) - \text{atrace}(M)^2 = \lambda_1 * \lambda_2 - \alpha(\lambda_1 + \lambda_2)^2 \text{ με } 0,04 < \alpha < 0,06$$

Ο τελεστής Harris χρησιμοποιείται για τον υπολογισμό της παραπάνω Response Function R . Για γωνίες, η τιμή του R είναι υψηλή, για γραμμικά όρια το R λαμβάνει τιμές μικρότερες του μηδενός ενώ μικρές τιμές του $|R|$ καταδεικνύουν επίπεδη περιοχή (Harris & Stephens, 1988) (εικόνα 7).



Εικόνα 7. Οπτική απεικόνιση της εφαρμογής του Corner Response Function (Collins)

2.3 Descriptors

Σε ότι αφορά στην αναγνώριση αντικειμένων, οι αρχικές προσπάθειες της επιστημονικής κοινότητας εστιάστηκαν στην ανίχνευση προσώπου. Το εν λόγω πρόβλημα μπορεί να προσεγγισθεί με διάφορες μεθόδους όπως η προσπάθεια εντοπισμού του χρώματος του δέρματος, τεχνική που αφορά κυρίως τις έγχρωμες εικόνες, η καταγραφή της κίνησης (πχ των χειλιών κατά την ομιλία), ο εντοπισμός του γενικού σχήματος της κεφαλής, συγκρινόμενου με το φόντο της εικόνας, προσπάθεια συνδυασμού όλων αυτών των μεθόδων κλπ. Το πρώτο αυτοματοποιημένο σύστημα αναγνώρισης (Kanade) υλοποιήθηκε το 1973 ενώ στη συνέχεια, το 1987, εφαρμόστηκε η μέθοδος της ανάλυσης των κύριων συνιστωσών (PCA – Principal Component Analysis)

των Sirovich και Kirby. Οι Turk και Pentland με το Eigenface το 1991²⁷, αναζωογόνησαν την έρευνα στο συγκεκριμένο τομέα, ενώ η μέθοδός τους επεκτάθηκε περαιτέρω από τους Etemad και Challepa με τα Fisherface. Το 2001, οι Viola και Jones εφάρμοσαν μία νέα μέθοδο ανίχνευσης προσώπων η οποία στηριζόταν στον συνδυασμό χρήσης του κατηγοριοποιητή AdaBoost με τα Haar wavelets, για την εξαγωγή των χαρακτηριστικών σημείων.

Αν επιχειρούνταν η κατηγοριοποίηση των χρησιμοποιούμενων αλγορίθμων τότε αυτοί δύναται να ενταχθούν σε δύο μεγάλες κατηγορίες, σε αυτούς που στηρίζονται στην εμφάνιση/σχήμα και σε αυτούς της "βαθιάς μάθησης (deep learning)" οι οποίοι στηρίζονται σε νευρωνικά δίκτυα. Στη συγκεκριμένη μεταπτυχιακή διατριβή θα μελετηθεί η πρώτη κατηγορία. Έτσι, οι ανιχνευτές αντικειμένων που θα περιγραφούν στη συνέχεια, στους οποίους και βασίζεται το κύριο πλαίσιο της μεταπτυχιακής διατριβής, στηρίζονται στον εντοπισμό συγκεκριμένων feature τα οποία εμπεριέχουν πληροφορίες που "χαρακτηρίζουν" το αντικείμενο ενδιαφέροντος. Συγκεκριμένα χρησιμοποιούνται "χαρακτηριστικά τύπου HAAR, Local Binary Patterns (LBP) και Histogram of Oriented Gradients, με σκοπό να αποτυπωθούν σημεία του αντικειμένου τα οποία ξεχωρίζουν σε ότι αφορά στην χωρική τους σχέση με τα υπόλοιπα αντικείμενα της εικόνας. Στη συνέχεια, αυτά τα χαρακτηριστικά χρησιμοποιούνται για την εκπαίδευση ενός κατηγοριοποιητή (GAB, SVM), προκειμένου να εντοπίζεται σε μια εικόνα το αντικείμενο ενδιαφέροντος (Luigi De Russis, Alberto Sacco, 2016).

2.3.1 Τοπικό Δυναμικό Μοτίβο (Local Binary Patterns - LBP)

Τα LBPs έχουν χρησιμοποιηθεί για ένα ευρύ φάσμα εφαρμογών όπως ανίχνευση προσώπου (Hadid, Pietikainen, & Ahonen, 2004), αναγνώριση προσώπου (Ahonen, Hadid, & Pietikainen, 2006), αναγνώριση έκφρασης (Shan, Gong, & McOwan, 2009) και κατηγοριοποίησης υφής/μοτίβων (L. Liu, Zhao, Long, Kuang, & Fieguth, 2012). Αν και ως έννοια εισήχθησαν τη δεκαετία του '90 (Ojala, Pietikainen, & Harwood, 1994), εν τούτοις, καθιερώθηκαν με την έρευνα του Ojala και άλλων το 2002 (Ojala, Pietikainen, & Maenpaa, 2002). Σε αντίθεση με τα χαρακτηριστικά υφής "Haralick", τα οποία λειτουργούν στηριζόμενα στην καθολική απεικόνιση της υφής (Haralick & Shanmugam, 1973), με τα LBPs επιδιώκεται ο υπολογισμός μίας τοπικής απεικόνιση της υφής.

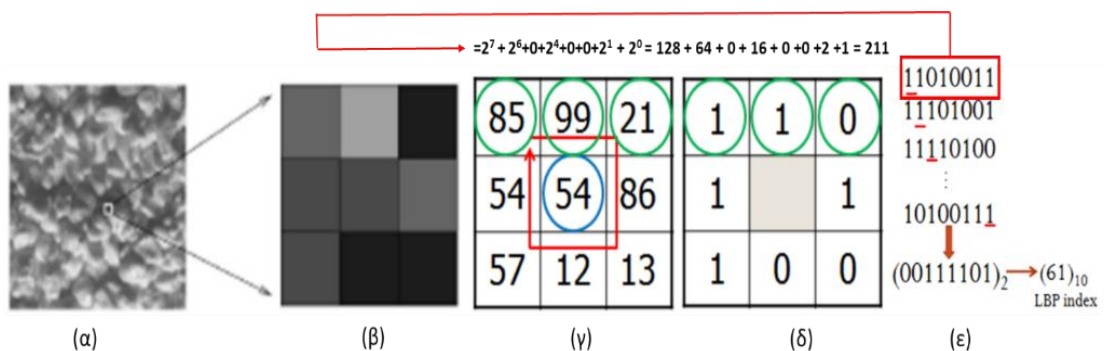
²⁷ Αποτέλεσαν σημαντική καμπή καθώς ήταν μια νέα τεχνική της οποίας ακόμα εφαρμόζονται αρκετές παραλλαγές

Μία συχνή υλοποίηση των LBPs είναι αυτή με τη χρήση ενός παραθύρου 3X3, με το οποίο σαρώνεται η εικόνα (Kyrgkou, 2017). Η διαδικασία ακολουθεί τα παρακάτω βήματα (Ojala, Pietikäinen, & Harwood, 1996):

- Μετασχηματίζεται η εικόνα στην κλίμακα του γκρι (grayscale).
- Για κάθε pixel επιλέγεται η γειτνιάζουσα περιοχή (neighborhood). Στην περίπτωση του παραθύρου 3X3, ο λόγος γίνεται για τα 8 pixel τα οποία περιβάλλουν το κεντρικό, αποτελώντας έτσι τη "γειτονιά του" (neighbor pixels).
- Για κάθε περιοχή 3X3 της εικόνας, αφού ληφθεί η τιμή των pixel, κανονικοποιούνται λαμβάνοντας τις τιμές 0 και 1 σύμφωνα με μια τιμή κατώφλιου. Το κατώφλι ορίζεται από το κεντρικό pixel. Τα γειτνιάζοντα pixel του παραθύρου λαμβάνουν την τιμή 0 αν η τρέχουσα τιμή τους είναι μικρότερη από αυτή του κεντρικού pixel και 1 αν η τιμή τους είναι ίση ή μεγαλύτερη.

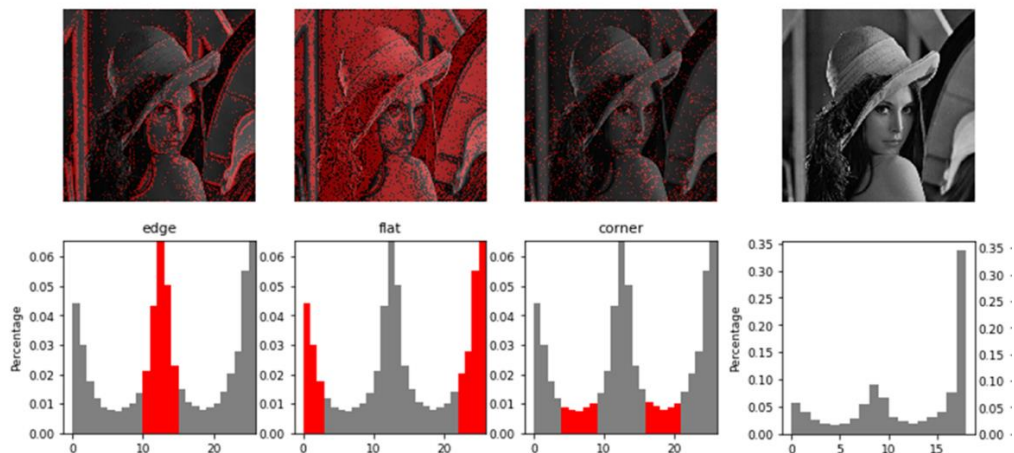
$$LBP_{p,r} = \sum_{n=0}^{p-1} S(x_{r,n} - x_{0,0})2^n, \quad S(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

- Ο κανονικοποιημένος πίνακας τιμών, που φαίνεται στην εικόνα 8 μετατρέπεται στη συνέχεια σε μία σειρά δυαδικών αριθμών ξεκινώντας από ένα ορισμένο γειτονικό pixel και εκτελώντας κίνηση σύμφωνα με τη φορά του ρολογιού ή αντίθετα αυτής. Δεν έχει σημασία ούτε η φορά κίνησης ούτε το pixel εκκίνησης. Αρκεί να διατηρούνται σταθερά όλα τα παράθυρα, κατά τη σάρωση της εικόνας.



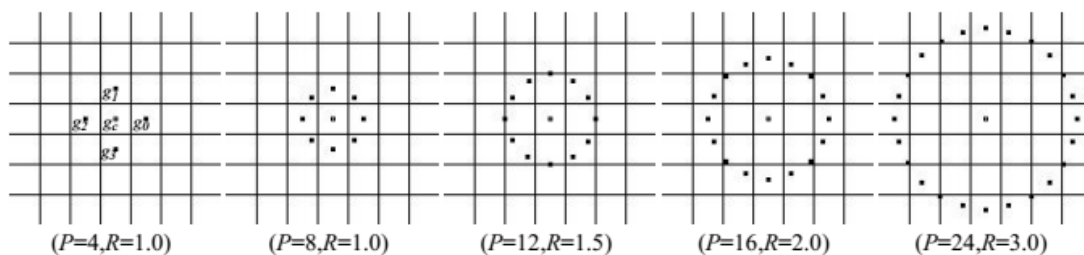
Εικόνα 8. Αλγόριθμος LBP (α) Αρχική εικόνα (β) Τυχαίο παράθυρο 3X3 από την εικόνα (γ) Τιμές των pixel του παραθύρου (δ) Κανονικοποίηση των τιμών βάσει αυτής του κεντρικού pixel (ε) Δυαδική και δεκαδική τιμή (Doost & Rahebi, 2012)

- Ο δυαδικός αριθμός (LBP code) των 8-bit που λαμβάνεται, μετατρέπεται σε δεκαδικό και αυτή η τιμή τοποθετείται στη θέση του κεντρικού pixel στον δυσδιάστατο πίνακα LBP που δημιουργείται (εικόνα 8).
- Αφού η παραπάνω διαδικασία ολοκληρωθεί για όλα τα pixel της εικόνας και συμπληρωθεί ο δυσδιάστατος πίνακας, τότε λαμβάνεται το αντίστοιχο ιστόγραμμα. Από τη στιγμή που λήφθηκε παράθυρο μεγέθους 3X3, το οποίο καταλήγει σε 8 γειτονικά pixel, το ιστόγραμμα μπορεί να λάβει $2^8 = 256$ τιμές (από 0 έως 255). Στην εικόνα 9 φαίνεται το αντίστοιχο ιστόγραμμα.



Εικόνα 9. LBP histogram

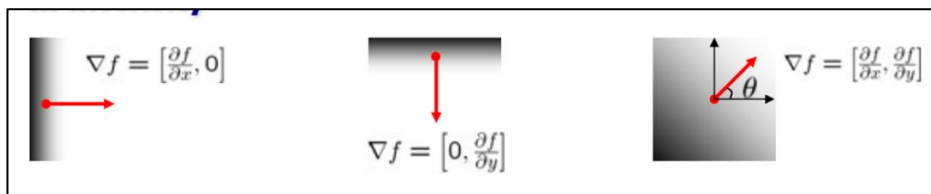
Η βασική εφαρμογή του LBP δίνει τη δυνατότητα να καταγραφούν και να ληφθούν υπόψη με εξαιρετική ακρίβεια οι λεπτομέρειες σε μια εικόνα. Ταυτόχρονα όμως αυτό αποτελεί και μειονέκτημα του αλγόριθμου καθώς δε λαμβάνονται υπόψη διαφορετικές κλίμακες. Αυτό επιλύθηκε με την εισαγωγή δύο παραμέτρων, τον αριθμό των σημείων p , σε ένα σύνολο pixel με συμμετρική κυκλική γειτνίαση και την ακτίνα r αυτού του κύκλου (Ojala 2002).



Εικόνα 10. Κυκλικά συμμετρικοί γείτονες για διάφορες τιμές του (p, r) (Ojala et al., 2002)

2.3.2 Μετασχηματισμοί Haar

Οι μετασχηματισμοί Haar χρησιμοποιούνται για την απεικόνιση των χαρακτηριστικών (features representation), των αντικειμένων ενδιαφέροντος. Αυτά τα χαρακτηριστικά χρησιμοποιούνται στη συνέχεια για την εκπαίδευση του κατηγοριοποιητή (Viola & Jones, 2001). Έχουν τη βάση τους στην ανίχνευση των περιγραμμάτων (edge detection) των αντικειμένων που εμφανίζονται σε μια εικόνα. Για να επιτευχθεί το edge detection, η εικόνα, η οποία είναι συνάρτηση των (x,y) , μετατρέπεται σε ένα set από pixel τα οποία κατά κάποιο τρόπο αποτελούν σημαντικά στοιχεία για αυτή. Η βασική ιδέα του edge detection στηρίζεται στην προσπάθεια εύρεσης "ισχυρών" ενδείξεων για "καθολική" αλλαγή της τιμής των pixel εντός της εξεταζόμενης "γειτνιαζουσας περιοχής". Αυτό επιτυγχάνεται με τη χρήση των παραγώγων (derivatives) όπου τα ακρότατα αυτών σηματοδοτούν σημαντική αλλαγή στην τιμή των pixel (πχ μετάβαση από μια περιοχή άσπρου χρώματος σε μία άλλη με μαύρο χρώμα και το αντίστροφο), όπως δείχνει και η εικόνα 11.



Εικόνα 11. Η συνάρτηση ∇f δίνει την κλίση (gradient) της εικόνας (Kuliah, 2010)

Στην πράξη αυτό επιτυγχάνεται φιλτράροντας²⁸ την εικόνα με ένα τελεστή (operator)²⁹ προκειμένου να υπολογιστεί η συνάρτηση της κλίσης (gradient function). Στη συνέχεια λαμβάνεται μία τιμή κατωφλίου (threshold) ώστε να εντοπιστούν τα pixel που οριοθετούν το περίγραμμα (edge pixels)³⁰.

Κατά παρόμοιο τρόπο, τα "Haar-like features" κάνουν χρήση της συνέλιξης, με αντίστοιχα φίλτρα (kernel), προκειμένου να εντοπίσουν την παρουσία ή όχι ενός

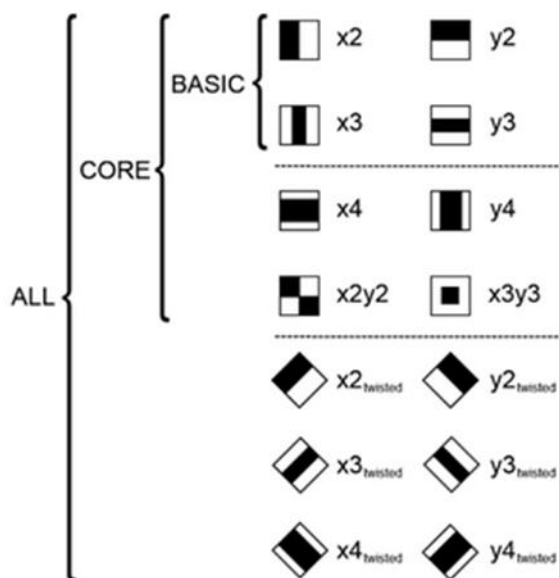
²⁸ Συνέλιξη (convolution)

²⁹ Αναφέρονται και ως φίλτρα, μάσκες, kernels.

³⁰ Η συγκεκριμένη τεχνική επεκτείνεται με τη χρήση του Canny Edge Operator όπου επιπροσθέτως εφαρμόζονται οι τεχνικές (1) του non-maximum suppression προκειμένου τα ακρότατα να "συμπιεστούν" σε μία γραμμή από pixel και (2) της hysteresis για τη σύνδεση των pixel όταν υπάρχουν κενά στις γραμμές που οριοθετούν το περίγραμμα.

συγκεκριμένου χαρακτηριστικού σε μία εικόνα³¹. Πρόκειται για τα χαρακτηριστικά/μοτίβα/σχέδια ορθογώνιου σχήματος της εικόνας 12, που σχηματίζονται με τις αντίστοιχες τιμές δεδομένων για τα pixel άσπρου και μαύρου χρώματος. Για την εκπαίδευση του αλγορίθμου δίνεται η δυνατότητα επιλογής, μέσω της παραμέτρου “διαμόρφωση(mode)” του αν θα χρησιμοποιηθούν τα πρωτότυπα (BASIC ή και CORE) ή το σύνολο (ALL) των Haar - like features. Η απεικόνιση αυτών φαίνεται στην εικόνα 12.

Τα Haar features εφαρμόζονται σε ένα σύνολο θετικών εικόνων (εικόνες που περιέχουν το αντικείμενο ενδιαφέροντος) και αρνητικών εικόνων (εικόνες που περιέχουν οτιδήποτε άλλο, εκτός από το αντικείμενο ενδιαφέροντος), προκειμένου να εξαχθούν τα χαρακτηριστικά με τα οποία θα εκπαιδευτεί ένας κατηγοριοποιητής. Κάθε χαρακτηριστικό αποτελεί μία μοναδική τιμή η οποία εξάγεται από τη διαφορά της μέσης τιμής των pixel για την μαύρη και την άσπρη περιοχή του κάθε Haar feature, εφόσον αυτό συντίθεται από ένα μαύρο και ένα άσπρο ορθογώνιο. Στην περίπτωση που το Haar feature περιλαμβάνει 3 ορθογώνια (πχ άσπρο-μαύρο-άσπρο), προστίθενται οι τιμές των pixel των δύο ακριανών ορθογωνίων (άσπρων) και αφαιρείται αυτή των κεντρικών pixel (μαύρων).



Εικόνα 12. Haar - like features (Kaehler 2016, 897)

³¹ Σύμφωνα με την αντίστοιχη ιστοσελίδα του OpenCV.org (OpenCV%20 Face Detection using Haar Cascades.html). Πρόσβαση:26/11/17

Επειδή για έναν ανιχνευτή ανάλυσης 24X24 το σύνολο των ορθογώνιων χαρακτηριστικών, για τα οποία απαιτείται να υπολογιστεί η διαφορά της μέσης τιμής των pixel, ξεπερνά τις 180.000 (Viola & Jones, 2001), γεγονός που καθιστά τον υπολογισμό τους υπολογιστικά ακριβό, χρησιμοποιείται η τεχνική των integral images. Σύμφωνα με αυτή την τεχνική, το άθροισμα των pixel κάθε ορθογώνιου μπορεί να υπολογιστεί με ένα πίνακα με τις τιμές τεσσάρων pixel.

Όμως από όλα τα Haar features που χρησιμοποιούνται, τα περισσότερα δεν επιτυγχάνουν το σκοπό τους. Στην περίπτωση της προσπάθειας ανίχνευσης ενός προσώπου, όταν ένα Haar-like feature τριών ορθογώνιων (πχ μαύρο-άσπρο-μαύρο) βρεθεί κατά τη σάρωση πάνω στο πρόσωπο, όπως φαίνεται στην εικόνα 13, τότε θα δώσει διαφορετικές τιμές για διαφορετικές περιοχές του προσώπου, ενώ θα έχει πολύ καλή απόδοση στην περιοχή των ματιών που είναι πιο σκούρα σε σχέση με το υπόλοιπο πρόσωπο ενώ στη μέση (περιοχή της μύτης) τα pixel έχουν πιο ανοιχτό χρώμα. Παρατηρώντας τις αντίστοιχες τιμές σε ανάλογες σαρώσεις με διαφορετικά πρόσωπα, παρατηρείται ότι υπάρχει "ταίριασμα" τιμών στη συγκεκριμένη περιοχή (μάτια - μύτη), για το ίδιο Haar Feature. Επίσης, το ίδιο Haar Feature δεν έχει ανάλογη απόδοση όταν σαρώσει την περιοχή στα μάγουλα ή στο μέτωπο όπου η τιμή των pixel είναι σχετικά ομοιόμορφη. Το πρόβλημα όμως που προκύπτει είναι ότι αρκετές φορές λαμβάνονται αντίστοιχες τιμές και από περιοχές της εικόνας στις οποίες δεν υπάρχει η συγκεκριμένη περιοχή του προσώπου. Οπότε, δεν είναι ασφαλές να χρησιμοποιείται μόνο ένα Haar feature προκειμένου να χαρακτηριστεί κάτι ως πρόσωπο.

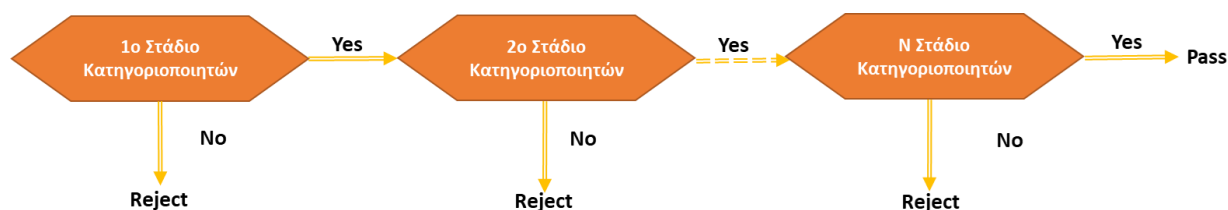


Εικόνα 13. Αναπαράσταση εφαρμογής των Haar Features σε μία εικόνα (Bhaskar, 2017)

Για την επίλυση των παραπάνω προβλημάτων που προκύπτουν από τη δοκιμή των 180.000 features σε διαφορετικές κλίμακες και προσανατολισμούς, αλλά και για τη επιτάχυνση της όλης διαδικασίας, χρησιμοποιείται η κατηγοριοποίηση που εφαρμόζει

την τεχνική της ενίσχυσης (AdaBoost), η οποία αναλύεται στην ενότητα 2.4.1, σε συνδυασμό με την κατασκευή μίας αλληλουχίας κατηγοριοποιητών (cascade of classifiers). Συγκεκριμένα:

- Αρχικά εφαρμόζεται κάθε Haar feature σε όλες τις εικόνες που συγκροτούν το δείγμα και επιλέγονται εκείνα τα features που παρουσιάζουν το μικρότερο ρυθμό σφάλματος. Αυτά αποτελούν και τους "βέλτιστους" αδύναμους κατηγοριοποιητές (weak classifiers) από τους οποίους εξάγεται ο τελικός κατηγοριοποιητής (strong classifier). Σύμφωνα με τους Viola και Jones, από τα 180.000 features κατέληξαν σε ένα αριθμό λίγο μεγαλύτερο του 6.000 ενώ τα αρχικά αποτελέσματα έδειξαν ότι ένας κατηγοριοποιητής που προήλθε από 200 features (weak classifiers) επέδειξε ποσοστό ανίχνευσης 95% (ibid).
- Για την μείωση του χρόνου υπολογισμού αντί να εφαρμόζονται και τα 6000 features στο κάθε παράθυρο που ελέγχεται, ομαδοποιούνται και εξετάζονται κατά στάδια. Στα πρώτα στάδια, με γνώμονα ότι το μεγαλύτερο μέρος μιας εικόνας δεν περιλαμβάνει το αντικείμενο ενδιαφέροντος, έχουν ενταχθεί κατηγοριοποιητές οι οποίοι αποσκοπούν να απορρίψουν την ύπαρξη του αντικειμένου ενδιαφέροντος στο μέρος της εικόνας που εξετάζουν. Όσο δεν υπάρχει απόρριψη εφαρμόζονται οι κατηγοριοποιητές των επόμενων σταδίων οι οποίοι είναι πιο πολύπλοκοι, μέχρι να επιτευχθεί το ποσοστό θετικής αναγνώρισης που τέθηκε. Η συγκεκριμένη διαδικασία παρουσιάζεται στην εικόνα 14.



Εικόνα 14. Έλεγχος των κατηγοριοποιητών κατά στάδια. Σε αυτή την ακολουθία (cascade) πρώτα εξετάζονται οι "ταχύτεροι" κατηγοριοποιητές και στη συνέχεια οι πιο "αργοί".

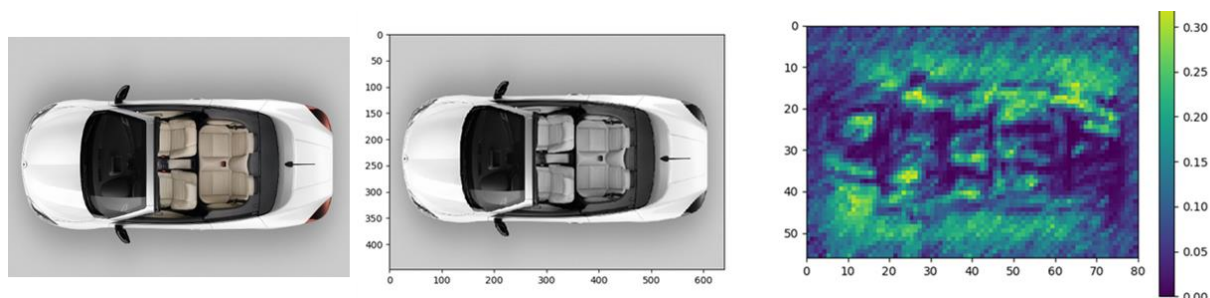
2.3.3 Ιστογράμματα της Βάθμωσης (Histogram of Oriented Gradients-HOG)

Ο HOG αποτελεί έναν "feature descriptor" ο οποίος επίσης χρησιμοποιείται για την

ανίχνευση αντικειμένων. Ως τεχνική, προσομοιάζει με τους "Edge Orientation Histograms" και "Scale Invariant Feature Transform (SIFT)" descriptors με τη διαφορά ότι ο HOG υπολογίζεται εντός ενός πυκνού πλέγματος ομοιόμορφα κατανεμημένων κελιών όπου, προκειμένου να βελτιωθεί η απόδοση, εξετάζεται η τοπικά κανονικοποιημένη αντίθεση χρησιμοποιώντας την τεχνική της αλληλοεπικάλυψης.

Στην περίπτωση του HOG, το ιστόγραμμα (histograms) της κατεύθυνσης της κλίσης (oriented gradients) της τιμής των pixel της εικόνας, είναι αυτή που χρησιμοποιείται ως χαρακτηριστικό (feature) (Dalal & Triggs, 2005). Η κλίση προκύπτει από τις παραγώγους της τιμής των pixel κατά τον x και y άξονα. Η τιμή της κλίσης, η οποία είναι υψηλή πλησίον των περιγραμμάτων (edges) και των γωνιών (corners), καθόσον στα συγκριμένα σημεία λαμβάνει χώρα απότομη αλλαγή, είναι ιδιαίτερος χρήσιμη λόγω των πληροφοριών που παρέχει (σε αντίθεση με τις "επίπεδες" περιοχές όπου η τιμή των pixel είναι σταθερή). Η διαδικασία υπολογισμού των HOG περιλαμβάνει πέντε στάδια τα οποία περιγράφονται στη συνέχεια.

Το πρώτο στάδιο αναφέρεται στην προ-επεξεργασία (pre-processing). Σε αυτό το στάδιο τα τμήματα της εικόνας που αποκόπτονται (patches), και τα οποία περιλαμβάνουν το αντικείμενο ενδιαφέροντος (εν προκειμένω το αυτοκίνητο) πρέπει να έχουν αναλογία μήκους προς ύψος (w/h) 2:1³². Στη συνέχεια, το patch επεξεργάζεται ώστε από το αρχικό μέγεθος (πχ 160X80) να μετασχηματιστεί στο 128X64 ενώ ταυτόχρονα υπόκειται και σε διόρθωση σε ότι αφορά στις τιμές gamma.



Εικόνα 15. Οπτικοποίηση της τιμής της κλίσης. Σε κάθε pixel, η κλίση έχει μέγεθος και κατεύθυνση. Για τις έγχρωμες εικόνες υπολογίζεται η κλίση και των τριών καναλιών. Η τιμή της στο κάθε pixel ισούται με το μέγιστο του μεγέθους της κλίσης των τριών

³² Αυτή είναι η αναλογία των διαστάσεων μήκουςΧπλάτος των οχημάτων. Στην έρευνα των Dalal και Triggs (2005), που αναφέρονταν στην ανίχνευση πεζών, η επιδιωκόμενη αναλογία των patches ήταν ακριβώς η αντίστροφη, 1:2.

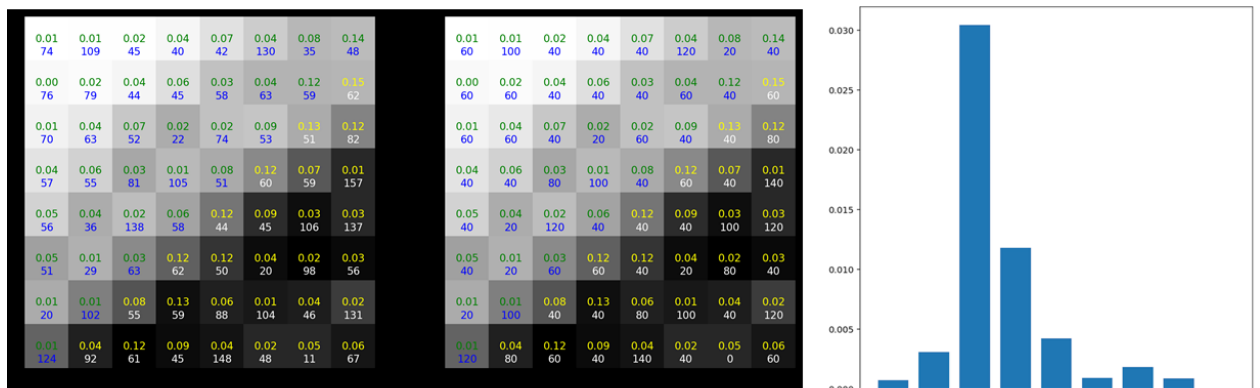
καναλιών, ενώ η κατεύθυνση θα αντιστοιχεί στη μέγιστη κλίση (Bobick, 2011-18).

Στο δεύτερο στάδιο υπολογίζεται η κλίση της τιμής των pixel (gradient) κατά την οριζόντια και κατακόρυφη διάσταση. Αυτό επιτυγχάνεται με την εφαρμογή ενός φίλτρου Kernel [-1 0 1] ή ενός τελεστή Sobel με το μέγεθος του kernel να λαμβάνει την τιμή της μονάδας. Στη συνέχεια υπολογίζεται το μέγεθος g και η κατεύθυνση θ της κλίσης όπως φαίνεται παρακάτω:

$$g = \sqrt{g_x^2 + g_y^2}$$

$$\theta = \arctan \frac{g_y}{g_x}$$

Στο τρίτο στάδιο η εικόνα (patch) χωρίζεται σε κελιά διαστάσεων 8X8 και για κάθε ένα από αυτά υπολογίζεται η τιμή της κλίσης. Το κάθε κελί περιλαμβάνει 192 τιμές από pixel ($8*8*3$) ενώ η κλίση του συντίθεται από 2 τιμές (μέγεθος g και κατεύθυνση θ), για κάθε pixel (συνολικά $8*8*2=128$ αριθμοί). Οι 128 αριθμοί δύναται να αναπαρασταθούν με τη χρήση ενός ιστογράμματος 9 διακριτών τιμών και να αποθηκευτούν με τη μορφή ενός πίνακα ή ανύσματος 9 τιμών.



Εικόνα 16. Μέγεθος (πράσινες ή κίτρινες τιμές) και κατεύθυνση σε μοίρες της κλίσης (μπλε ή άσπρες τιμές). Στη μεσαία εικόνα η κατεύθυνση έχει στρογγυλοποιηθεί σε μία από τις 9 διακριτές τιμές (0, 20, 40, ..., 160). Βάσει αυτών των τιμών συνάγεται και το ιστόγραμμα όπου φαίνεται η επικρατούσα τιμή να είναι αυτή των 40 μοιρών αποτυπώνοντας έτσι την αλλαγή στην ένταση των pixel, κατά τη διεύθυνση της διαγωνίου (Thuyen, 2017).

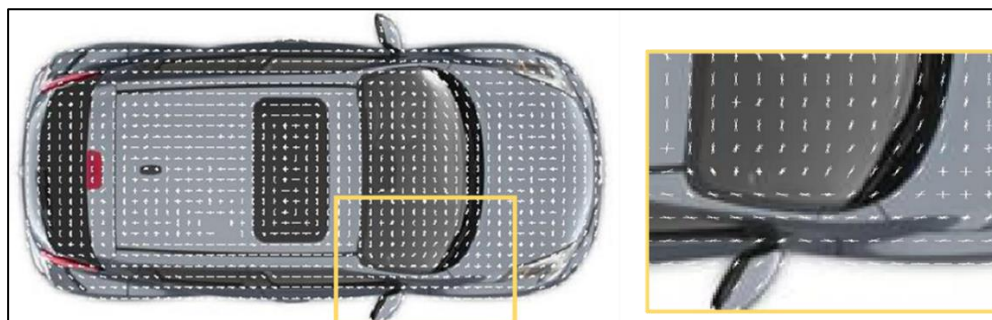
Οι διαστάσεις του κάθε κελιού επιλέχθηκαν λόγω του ότι:

- Αποτελούν υποπολλαπλάσια της αρχικής διάστασης 128X64 (σχετίζεται άμεσα με τις αντίστοιχες διαστάσεις της ερευνητικής εργασίας των Dalal και Triggs (2005).
- Το συγκεκριμένο μέγεθος επαρκεί για να καταλάβει το χώρο σημαντικών χαρακτηριστικών του αντικειμένου ενδιαφέροντος (πχ για την περίπτωση των πεζών δύναται να αποτελεί ένα σημαντικό μέρος του προσώπου).
- Είναι λιγότερο επιδεκτικές στο θόρυβο.

Στο προηγούμενο βήμα δημιουργήθηκε ένα ιστόγραμμα βάσει της κλίσης των pixel. Όμως, η κλίση εξαρτάται από τη φωτεινότητα (πχ αν αυξηθεί η φωτεινότητα της εικόνας, πολλαπλασιάζοντας την τιμή των pixel επί ένα συντελεστή μεγαλύτερο της μονάδας, τότε η κλίση θα μεταβληθεί αντίστοιχα σύμφωνα με την τιμή του συγκεκριμένου συντελεστή). Προκειμένου να αποφευχθεί αυτό το φαινόμενο και ο descriptor να μην επηρεάζεται από αλλαγές στην τιμή της φωτεινότητας, το ιστόγραμμα κανονικοποιείται. Αυτό γίνεται με την εφαρμογή της L2 νόρμας του ανύσματος (L2 - norm : $u \rightarrow \frac{u}{\sqrt{\|u\|_2^2 + \epsilon^2}}$. Αυτό λαμβάνει χώρα σε τέσσερα κελιά (16X16) αντί του ενός (8X8)

επί του οποίου υπολογίστηκε το ιστόγραμμα και λαμβάνεται έτσι το κανονικοποιημένο άνυσμα 36X1. Αυτή η τεχνική εφαρμόζεται σε όλο το patch με ένα το παράθυρο 16X16 να σαρώνει την εικόνα μετακινούμενο κάθε φορά κατά ένα διάστημα 8X8.

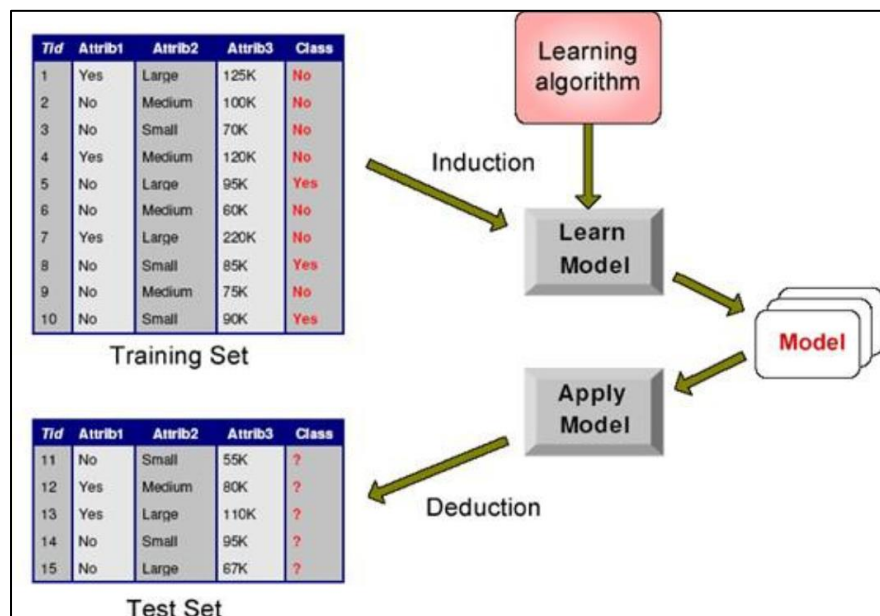
Το τελευταίο στάδιο περιλαμβάνει την δημιουργία του τελικού ανύσματος για το patch το οποίο προέρχεται από τα αντίστοιχα διαστάσεων 36X1. Για το patch ενός οχήματος που λήφθηκε για τη συγκεκριμένη μεταπτυχιακή διατριβή (128X64) το παράθυρο των 16X16 μετακινήθηκε συνολικά κατά 15 οριζόντιες θέσεις και 7 κατακόρυφες καταλαμβάνοντας έτσι συνολικά 105 διακριτές θέσεις. Επομένως η συνολική διάσταση του HOG είναι ένα άνυσμα 3780X1.



2.4 Κατηγοριοποιητές (Classifiers)

Σύμφωνα με τον Tan, «η κατηγοριοποίηση, η οποία είναι η εργασία εκχώρησης αντικειμένων σε μία από τις διάφορες προκαθορισμένες κατηγορίες, είναι ένα πρόβλημα που περιλαμβάνει πολλές και διαφορετικές εφαρμογές» (Tan, 2016). Ως κατηγοριοποίηση ορίζεται η «εργασία εκμάθησης μιας συνάρτησης - στόχου (target-function) f , η οποία απεικονίζει κάθε σύνολο χαρακτηριστικών x , σε μία από τις προκαθορισμένες ετικέτες κατηγορίας. Η συνάρτηση - στόχος είναι γνωστή και ως μοντέλο κατηγοριοποίησης (classification model)»(ibid).

«Κάθε κατηγοριοποιητής χρησιμοποιεί έναν αλγόριθμο μάθησης (learning algorithm) προκειμένου να εντοπίσει ένα μοντέλο που ταιριάζει καλύτερα στη σχέση μεταξύ του συνόλου χαρακτηριστικών και της ετικέτας κατηγορίας των δεδομένων εισόδου. Το μοντέλο που εξάγεται από τον αλγόριθμο πρέπει να ταιριάζει καλά στα δεδομένα εισόδου αλλά και να προβλέπει σωστά τις ετικέτες της κατηγορίας των εγγραφών τις οποίες δε γνωρίζει. Επομένως, ο ουσιαστικός σκοπός του αλγορίθμου είναι η δημιουργία ενός μοντέλου, με την ικανότητα γενίκευσης, που θα προβλέπει τις ετικέτες κατηγοριών τις οποίες δε γνωρίζει εκ των προτέρων» (ibid).



Εικόνα 18. Επίλυση του προβλήματος κατηγοριοποίησης (Hopkins, Garey 2004)

Στην περίπτωση του machine learning, κατηγοριοποίηση είναι το πρόβλημα της αναγνώρισης του σε ποια κατηγορία ανήκει ένα εξεταζόμενο αντικείμενο. Προκειμένου

να επιτευχθεί αυτό, η γενική μεθοδολογία που ακολουθείται είναι αυτή που φαίνεται στην εικόνα 18. Αρχικά χωρίζονται τα διαθέσιμα δεδομένα σε δύο μέρη, στο υποσύνολο της εκπαίδευσης (training set) και στο υποσύνολο του ελέγχου (test set). Το πρώτο χρησιμοποιείται για να δημιουργηθεί το μοντέλο της κατηγοριοποίησης ενώ το δεύτερο, για τον έλεγχο της ευρωστίας του κατηγοριοποιητή. Για την περίπτωση του machine learning και ειδικότερα στην διαδικασία που σχετίζεται με την προσπάθεια ανίχνευσης ενός αντικειμένου, η κατηγοριοποίηση θεωρείται ως επιβλεπόμενη (supervised) καθώς μέσω της επίβλεψης δημιουργήθηκε ένα σύνολο εικόνων, ορθής κατηγοριοποίησης, που χρησιμοποιούνται για τη φάση της εκπαίδευσης (Alpaydin, 2010).

Στην περίπτωση του HAAR και του LBP χρησιμοποιείται η τεχνική της ενίσχυσης (boosting) και συγκεκριμένα μία υλοποίηση κατηγοριοποιητή που ονομάζεται Adaptive Boost (AdaBoost). Για το HOG αν και κυριαρχεί η χρησιμοποίηση των Μηχανών Διανυσμάτων Υποστήριξης (SVM – Support Vector Machines), εντούτοις υπάρχουν και περιπτώσεις όπου υλοποιείται η κατηγοριοποίηση κάνοντας χρήση του Ada Boost.

2.4.1 Ada Boost (Gentle Ada Boost – GAB)

Ο όρος ενίσχυση (boosting) αναφέρεται με μία κατηγορία αλγορίθμων μηχανικής μάθησης (machine learning) η οποία πραγματοποιείται με επίβλεψη (supervised)³³. Βασίζεται στην ιδέα ότι ένας συνδυασμός από απλούς κατηγοριοποιητές³⁴, ο οποίος (συνδυασμός) προέρχεται από ασθενή μοντέλα (weak learners)³⁵, μέσω μια επαναληπτικής διαδικασίας, παρουσιάζει καλύτερη απόδοση συγκρινόμενος ξεχωριστά με κάθε ένα από τους απλούς κατηγοριοποιητές (Scharire, 1990). Η επαναληπτική διαδικασία στοχεύει στην προσαρμοστική κατανομή των δειγμάτων εκπαίδευσης ώστε η διαδικασία μάθησης να εστιαστεί σε εκείνα τα δείγματα η κατηγοριοποίηση των οποίων είναι δυσχερής (Tan, 2016). Ο αλγόριθμος της ενίσχυσης αναλύεται όπως φαίνεται στη συνέχεια (Ferreira & Figueiredo, 2012):

³³ Στην εκμάθηση με επίβλεψη, χρησιμοποιείται ένα σύνολο δεδομένων το οποίο διαχωρίζεται σε δύο μέρη. Σε αυτό που θα χρησιμοποιηθεί για την εκπαίδευση του αλγορίθμου (training dataset) και σε αυτό που θα χρησιμοποιηθεί για τον έλεγχο της απόδοσής του (testing dataset).

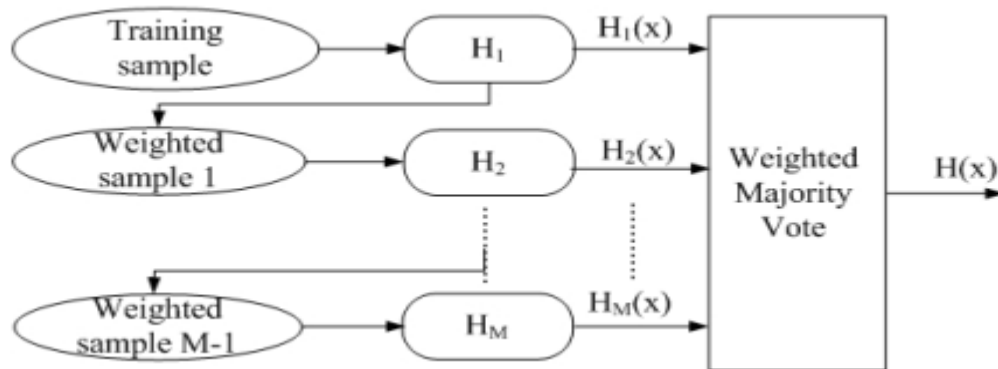
³⁴ Το σύνολο αυτών των κατηγοριοποιητών ονομάζεται "ensembles"

³⁵ Ως "Weak Learner(WL)" οι Ferreira και Figueiredo ορίζουν έναν αλγόριθμο εκμάθησης που έχει τη δυνατότητα να δημιουργήσει κατηγοριοποιητές που αποδίδουν καλύτερα από την κατηγοριοποίηση που στηρίζεται στην τυχειότητα (2012). Οπότε, ένας WL έχει περισσότερες του 50% πιθανότητες να προβεί στην επιλογή της ορθής ετικέτας για μία κατηγορία. (Freund & Scharire, 1997). Από την άλλη, ένας "Strong Learner (SL)" έχει τη δυνατότητα να αποδώσει κατηγοριοποιητές που παρουσιάζουν μικρή πιθανότητα σφάλματος.

Είσοδος : Σύνολο Δεδομένων $Z = \{z_1, z_2, \dots, z_N\}$, με $z_i = (x_i, y_i)$, όπου $x_i \in C$ και $y_i \in \{-1,+1\}$
 Έξοδος : Ένας κατηγοριοποιητής $C : C \rightarrow \{-1,+1\}$

1. Τυχαία επέλεξε, χωρίς αντικατάσταση, $L_1 < N$ δείγματα από το Z για τη δημιουργία του Z_1^*
2. Εφάρμοσε τον WL στο Z_1^* για την δημιουργία του C_1 .
3. Επέλεξε $L_2 < N$ δείγματα από το Z , με τα μισά δείγματα να είναι από εκείνα που κατηγοριοποιήθηκαν εσφαλμένα από το C_1 , δημιουργώντας έτσι το Z_2^*
4. Εφάρμοσε τον WL στο Z_2^* για την δημιουργία του C_2
5. Επέλεξε όλα τα δείγματα από το Z για τα οποία ο C_1 και ο C_2 διαφωνούν, δημιουργώντας έτσι το Z_3^*
6. Εφάρμοσε τον WL στο Z_3^* για την δημιουργία του C_3
7. Δημιούργησε τον τελικό κατηγοριοποιητή βάσει της ψήφου πλειοψηφίας: $C(x) = \text{sign} (\sum_{b=1}^3 C_b(x))$

Όπως διαπιστώνεται, το δείγμα εκπαίδευσης χωρίζεται (χωρίς αντικατάσταση) σε τρία υποσύνολα με το τρίτο να απαρτίζεται από τα στιγμιότυπα στα οποία διαφώνησαν οι $C_{1,2}$ στους δύο προηγούμενους γύρους.



Εικόνα 19. Γραφική απεικόνιση του αλγορίθμου AdaBoost (Ferreira & Figueiredo, 2012)

Μέχρι σήμερα έχουν αναπτυχθεί αρκετοί αλγόριθμοι ενίσχυσης³⁶ με τις βασικές διαφορές τους να εστιάζονται στη μέθοδο απόδοσης βαρών στα δείγματα εκπαίδευσης καθώς επίσης και στον τρόπο συνάθροισης των βασικών κατηγοριοποιητών κατά το τελευταίο βήμα του αλγορίθμου. Από τους περισσότερο δημοφιλείς αλγορίθμους είναι ο AdaBoost (Adaptive Boost) και αποτέλεσε τον πρώτο αλγόριθμο ενίσχυσης ο οποίος μπορούσε να προσαρμοστεί στους βασικούς κατηγοριοποιητές (Ferreira & Figueiredo, 2012). Για το συγκεκριμένο αλγόριθμο, η επαναληπτική διαδικασία ορίζεται από κύκλους ή γύρους (rounds) ενίσχυσης, εντός των οποίων στα στιγμιότυπα του δείγματος εκπαίδευσης αποδίδεται κάποιο βάρος το οποίο, στο τέλος κάθε γύρου,

³⁶ π.χ. AdaBoost, BrownBoost, XGBoost, LogitBoost, LPBoost, TotalBoost, MadaBoost

μπορεί να μεταβάλλεται.

Η βασική ιδέα έγκειται στη χρήση σταθμιζόμενων εκδόσεων από το ίδιο σύνολο δεδομένων εκπαίδευσης αντί της τυχαίας λήψης υποσυνόλων δειγμάτων στη θέση αυτού (όπως αναφέρθηκε στην περίπτωση της απλής ενίσχυσης). Αυτή η μεθοδολογία είχε ως αποτέλεσμα, σε αντίθεση με προηγούμενες τεχνικές ενίσχυσης, να μην απαιτείται μεγάλο σύνολο δεδομένων εκπαίδευσης (ibid).

Ο αλγόριθμος AdaBoost μπορεί να χρησιμοποιηθεί για την ανίχνευση (αναγνώριση) οχημάτων, ως μία περίπτωση δυαδικής κατηγοριοποίησης. Οι δύο κατηγορίες είναι αυτή του οχήματος και το φόντο (background) αυτού, ανεξαρτήτως του τι περιλαμβάνει το τελευταίο. Ακολουθεί τα παρακάτω βήματα:

- Βήμα 1ο : Στην αρχή, αφού αποδοθεί το ίδιο βάρος σε όλα τα δείγματα, ώστε όλα τα δείγματα να έχουν την ίδια πιθανότητα να επιλεγούν για εκπαίδευση, επιλέγεται ένα αρχικό δείγμα (από το σύνολο των δειγμάτων της εκπαίδευσης σύμφωνα με την κατανομή δειγματοληψίας).
- Βήμα 2ο : Από τον πρώτο γύρο εκπαίδευσης προκύπτει ένας κατηγοριοποιητής, ο οποίος χρησιμοποιείται προκειμένου να αποδοθούν ετικέτες στο σύνολο των δειγμάτων εκπαίδευσης. Τα δείγματα που έλαβαν λάθος ετικέτα κατηγορίας θα τους αποδοθεί αυξημένο βάρος για τον επόμενο γύρο (το αντίθετο συμβαίνει για τα δείγματα που έλαβαν την ορθή ετικέτα). Αυτό έχει ως συνέπεια οι επόμενοι ασθενείς κατηγοριοποιητές να εστιάζονται περισσότερο στα δείγματα που έχουν μεγαλύτερες πιθανότητες εσφαλμένης πρόβλεψης, τα οποία έχουν την τάση στο τέλος να κυριαρχήσουν.
- Βήμα 3ο : Επαναλαμβάνεται το δεύτερο βήμα για όσο επιτυγχάνεται μεγαλύτερη ακρίβεια και μέχρι να εξαντληθεί το όριο του αλγορίθμου εκμάθησης.
- Βήμα 4ο : Συναθροίζονται οι βασικοί κατηγοριοποιητές, που εξάγονται από κάθε επανάληψη της διαδικασίας ενίσχυσης, ώστε να δημιουργηθεί ένας ισχυρός κατηγοριοποιητής ο οποίος τελικά βελτιώνει και την προβλεπτική ικανότητα του μοντέλου.

Τα παραπάνω βήματα εκφράζονται με μαθηματικό τρόπο με τον παρακάτω αλγόριθμο (Tan, 2016)³⁷:

³⁷ Σύμφωνα με την ίδια πηγή, ο AdaBoost είναι επιρρεπής στην υπερπροσαρμογή (321).

1. $w = \{w_j = 1/N \mid j = 1, 2, 3, \dots, N\}$ {Αρχικοποίησε τα βάρη για όλα τα N δείγματα}
2. Έστω k το πλήθος των επαναλήψεων ενίσχυσης
3. Για κάθε $i=1$ έως k επανέλαβε:
4. Δημιούργησε το σύνολο εκπαίδευσης D_i κάνοντας δειγματοληψία (με αντικατάσταση) από το D σύμφωνα με το w .
5. Εκπαίδευσε ένα βασικό κατηγοριοποιητή C_i στο δείγμα D_i .
6. Εφάρμοσε το C_i σε όλα τα δείγματα του αρχικού συνόλου εκπαίδευσης D
7. $e_i = \frac{1}{N} [\sum_j w_j \delta(C_i(x_j) \neq y_i)]$ | {Υπολόγισε το σταθμισμένο σφάλμα}
8. Αν $e_i > 0.5$ τότε
9. $w = \{w_j = 1/N \mid j = 1, 2, 3, \dots, N\}$ {Επανάφερε τις αρχικές τιμές βαρών για όλα τα N δείγματα}
10. Επιστροφή στο βήμα 4
11. Τέλος Αν
12. $\alpha_i = \frac{1}{2} \ln \frac{1-e_i}{e_i}$
13. Ενημέρωσε το βάρος κάθε δείγματος σύμφωνα με την εξίσωση

$$w_i^{(j+1)} = \frac{w_i^j}{z_j} X \begin{cases} \exp^{-\alpha_j}, \text{ αν } C_j(x_i) = y_i \\ \exp^{\alpha_j}, \text{ αν } C_j(x_i) \neq y_i \end{cases}$$
14. Τέλος για κάθε
15. $C^*(x) = \underset{y}{\operatorname{argmax}} \sum_{j=1}^T a_j \delta(C_j(x) = y)$

όπου:

- $\{(x_j, y_j) \mid j = 1, 2, 3, \dots, N\}$, είναι ένα σύνολο N δειγμάτων εκπαίδευσης
- C_i , ένας βασικός κατηγοριοποιητής
- e_i , ο βαθμός σφάλματος του βασικού κατηγοριοποιητή
- α_i , η σημασία του κατηγοριοποιητή (το α_i , έχει υψηλή θετική τιμή αν ο ρυθμός σφάλματος e_i είναι κοντά στο 0 και υψηλή αρνητική τιμή, αν ο ρυθμός σφάλματος είναι κοντά στο 1. Η τιμή του α_i είναι αυτή που χρησιμοποιείται για την ενημέρωση των βαρών.
- w_i^j , το βάρος που αποδίδεται στο δείγμα (x_j, y_j) στην j επανάληψη
- Z_j , παράγοντας κανονικοποίησης που χρησιμοποιείται για να εγγυηθεί ότι $\sum_i w_i^{(j+1)} = 1$

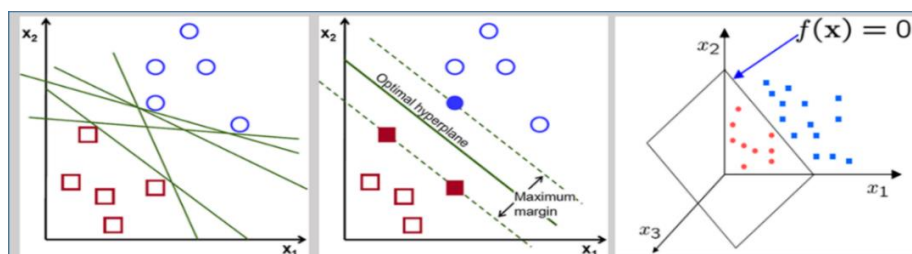
Από όταν προτάθηκε ο AdaBoost, αναπτύχθηκαν στη συνέχεια αρκετές εκδόσεις του. Από αυτές, σχετιζόμενος με τη συγκεκριμένη μεταπτυχιακή διατριβή είναι ο Gentle AdaBoost (GAB) ο οποίος αποτελεί βελτιωμένη έκδοση του Real AdaBoost³⁸ (Ferreira & Figueiredo, 2012). Στον τελευταίο, τα βήματα που ακολουθούν την έναρξη της

³⁸ Η μέθοδος της ενίσχυσης προτάθηκε το 1989, ο Ada Boost το 1996 και ο Real AdaBoost το 1999.

επαναληπτικής διαδικασίας, περιλαμβάνουν υπολογισμούς και εκτιμήσεις επί των πιθανοτήτων που υπάρχουν ώστε κάθε εκπαιδευτική ακολουθία να ανήκει σε μία κλάση, σύμφωνα με την τρέχουσα κατανομή βάρους των δειγμάτων (Friedman, Hastie, & Tibshirani, 2000). Τέλος, ο Gentle AdaBoost με τη χρήση Newton Steps παρέχει πιο αξιόπιστους και σταθερούς συνδυασμένους κατηγοριοποιητές (ensembles) (Ferreira & Figueiredo, 2012). Αυτό εξηγείται από το γεγονός ότι αποδίδει μικρότερη αξία στις ακραίες τιμές, μέσω της χρήσης παλινδρόμησης με σταθμισμένα ελάχιστα τετράγωνα, σε κάθε κύκλο της επαναληπτικής διαδικασίας (Lienhart 2003:4).

2.4.2 Μηχανή Διανυσμάτων Υποστήριξης (Support Vector Machines – SVM)

Η Μηχανή Διανυσμάτων Υποστήριξης είναι ένας αλγόριθμος μηχανικής μάθησης, με επίβλεψη, ο οποίος χρησιμοποιείται για σκοπούς κατηγοριοποίησης και παλινδρόμησης³⁹. Για τον συγκεκριμένο τύπο κατηγοριοποιητή εισάγεται η έννοια του υπερεπιπέδου μέγιστου περιθωρίου (maximal margin hyperplane). Στην εικόνα 20 παρουσιάζεται το διάγραμμα ενός δυαδικού συνόλου δεδομένων (κύκλοι και τετράγωνα). Το συγκεκριμένο δείγμα είναι δυνατόν να διαχωριστεί γραμμικά με τη χάραξη άπειρων νοητών γραμμών (υπερεπίπεδα) εκατέρωθεν των οποίων θα βρίσκονται ξεχωριστά οι κύκλοι και τα τετράγωνα.



Εικόνα 20. Διαχωρισμός δυαδικού συνόλου δεδομένων. Στο αριστερό σχήμα απεικονίζονται πέντε από τις άπειρες γραμμές οι οποίες δύνανται να διαχωρίσουν το δυαδικό δείγμα δεδομένων. Στο κεντρικό σχήμα έχει χαραχτεί το βέλτιστο υπερεπίπεδο το οποίο δημιουργεί το μέγιστο περιθώριο⁴⁰ μεταξύ των σημείων⁴¹ (OpenCV dev team 2018). Στο δεξιό σχήμα φαίνεται το αντίστοιχο υπερεπίπεδο για ένα χώρο 3-διαστάσεων με τα μπλε τετράγωνα να ικανοποιούν την εξίσωση

³⁹ Αυτή η τεχνική (SVM) προέρχεται από τη θεωρία της στατιστικής.

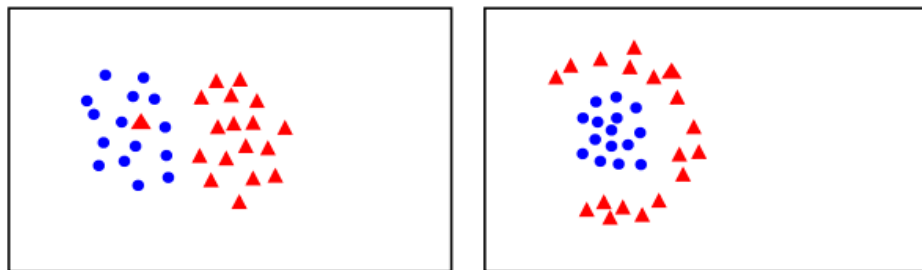
⁴⁰ Το μέγιστο περιθώριο είναι γνωστό και ως "περιθώριο κατηγοριοποιητή" (Tan, 2016)

⁴¹ Το συγκεκριμένο παράδειγμα στηρίζεται σε οπτική απεικόνιση ενός δυσδιάστατου καρτεσιανού συστήματος συντεταγμένων με τα αντίστοιχα σημεία να διαχωρίζονται από μία γραμμή. Στην πραγματικότητα έχουμε να κάνουμε με υπερεπίπεδα και διανύσματα τα οποία λειτουργούν σε ένα πολυδιάστατο χώρο.

$f(x) > 0$ και τα κόκκινα τρίγωνα την $F(x) < 0$ (Zisserman, 2011)

Παρά το γεγονός ότι τα υπερεπίπεδα στη συγκεκριμένη περίπτωση παράγουν στο σύνολό τους μηδενικό σφάλμα, εντούτοις είναι εφικτό να καθοριστεί ένα κριτήριο το οποίο θα βοηθά ώστε να επιλεγεί το βέλτιστο υπερεπίπεδο, δηλαδή εκείνο που θα λειτουργεί το ίδιο καλά, όταν εφαρμόζεται σε άγνωστα δείγματα (Friedman, Hastie, & Tibshirani, 2001). Αυτό θα συμβεί εάν ο κατηγοριοποιητής επιλέξει εκείνο το υπερεπίπεδο το οποίο θα απέχει τη μεγαλύτερη δυνατή απόσταση από το σύνολο των σημείων (υπερεπίπεδο μεγίστου περιθωρίου). Σύμφωνα με τον Tan, «τα όρια απόφασης με μεγάλα περιθώρια, τείνουν να έχουν καλύτερα σφάλματα γενίκευσης από εκείνα με μικρά περιθώρια» ενώ οι κατηγοριοποιητές που παράγουν όρια απόφασης με μικρό περιθώριο «είναι επιρρεπή στην υπερπροσαρμογή» (Tan, 2016).

Τα παραπάνω αναφέρονται σε μια απλουστευμένη μορφή γραμμικού διαχωρισμού, ενώ υπάρχουν και περιπτώσεις, όπως στην εικόνα 21, όπου ο γραμμικός διαχωρισμός δεν είναι εφικτός. Η συγκεκριμένη μεταπτυχιακή διατριβή, επιδιώκοντας την ταχύτητα στην ανίχνευση των οχημάτων, εστιάζεται στις γραμμικές μηχανές διανυσμάτων υποστήριξης καθώς παρουσιάζουν καλύτερες επιδόσεις από τις μη γραμμικές (πχ Radial Basis Function Kernel).



Εικόνα 21. Μη γραμμικά διαχωριζόμενα σύνολα (Zisserman 2011:3)

Μία γραμμική μηχανή υποστήριξης διανυσμάτων (linear SVM) αποτελεί τη λύση του παρακάτω προβλήματος: «Έστω $\{(x_i, y_i) \text{ με } i = 1, 2, \dots, n\}$ ένα σύνολο κατηγοριοποιημένων (labeled) δεδομένων με το $x \in \mathbb{R}^d$, $y_i \in \{1, -1\}$. Μία μηχανή υποστήριξης διανυσμάτων αποτελεί ένα γραμμικό κατηγοριοποιητή ο οποίος σχετίζεται με την συνάρτηση απόφασης $D(x) = \text{sign}(w^T x + b)$ ⁴² όπου $w \in \mathbb{R}^d$ και $b \in \mathbb{R}$

$$\begin{cases} \min & \frac{1}{2} \|w\|^2 \\ \text{με } & y_i(w^T x_i + b) \geq 1 \end{cases} \gg \text{(Canu, 2014)}$$

⁴² Η εξίσωση $w^T x + b = 0$ αποτελεί το όριο απόφασης (w, b : παράμετροι του μοντέλου)

Το περιθώριο ενός γραμμικού κατηγοριοποιητή σχετίζεται με το σφάλμα γενίκευσης. Το τελευταίο εξαρτάται από το σφάλμα εκπαίδευσης, το πλήθος των δειγμάτων εκπαίδευσης και την πολυπλοκότητα (ή χωρητικότητα) του μοντέλου. Σύμφωνα με τον Tan «η χωρητικότητα ενός μοντέλου είναι αντιστρόφως συσχετιζόμενη με το περιθώριό του» δηλαδή όσο μικρότερο περιθώριο διαθέτει ένα μοντέλο (οπότε η χωρητικότητά του αυξάνει), τόσο περισσότερα σύνολα εκπαίδευσης μπορεί να προσαρμόσει (Tan, 2016). Από την άλλη όμως αύξηση της χωρητικότητας προκαλεί αύξηση και του ορίου γενίκευσης (ibid). Οπότε, αυτό που επιδιώκεται με μία γραμμική μηχανή υποστήριξης διανυσμάτων είναι η μεγιστοποίηση του περιθωρίου λήψης απόφασης ώστε να ελαχιστοποιούνται τα σφάλματα γενίκευσης (ibid).

Με γνώμονα τα παραπάνω συμπεραίνεται ότι από την εκπαίδευση ενός SVM εξάγεται η εκτίμηση για τις παραμέτρους w και b , οι οποίες σχετίζονται με το όριο απόφασης. Επίσης, η μεγιστοποίηση του περιθωρίου οδηγεί σε ελαχιστοποίηση της τιμής $\frac{1}{2} \|w\|^2$ (ώστε να ισχύει $y_i(w^T x_i + b) \geq 1$).

2.5 Αξιολόγηση των Αποτελεσμάτων της Ανίχνευσης

Για τα προβλήματα που επιλύονται μέσω της μηχανικής μάθησης, όπως αυτό της ανίχνευσης οχημάτων, δύναται να χρησιμοποιηθεί ένα πλήθος διαφορετικών μοντέλων (αλγορίθμων) η απόδοση των οποίων κρίνεται βάσει ενός συνόλου δεδομένων που καλείται "validation/test dataset (δεδομένα επικύρωσης/δοκιμής)". Όπως θα αναλυθεί στη συνέχεια, τα χρησιμοποιούμενα μέτρα απόδοσης στοχεύουν στην ανάδειξη συγκριμένων χαρακτηριστικών. Ιδιαίτερα στην περίπτωση της ανίχνευσης, δεν αρκεί μόνο αυτή κάθε αυτή η ανίχνευση (δηλαδή απλά το γεγονός της ύπαρξης ενός η περισσότερων οχημάτων σε μια εικόνα) αλλά είναι εξίσου σημαντικό είναι και η εύρεση της θέσης αυτών στην εικόνα. Σε κάθε περίπτωση, οι μετρήσεις αξιολογούνται συγκρινόμενες με τα πραγματικά δεδομένα.

2.5.1 Πραγματικά Δεδομένα (Ground Truth)

Στον τομέα της μηχανικής μάθησης για την ανίχνευση αντικειμένων, ο όρος "ground truth" αναφέρεται σε ένα σύνολο εικόνων το οποίο συνοδεύεται από ένα αρχείο που περιέχει δεδομένα για τον αριθμό και τη θέση των αντικειμένων ενδιαφέροντος, στις

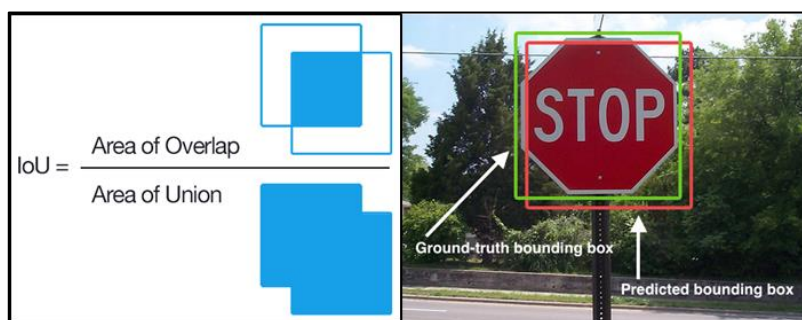
παραπάνω εικόνες⁴³. Τα δεδομένα αυτά συνήθως αφορούν στις συντεταγμένες⁴⁴ ενός ορθογώνιου (bounding box), εντός του οποίου περικλείεται το αντικείμενο.

Επί του παρόντος, κρίνεται σκόπιμο να αναφερθεί ότι η εργασία της επισήμανσης για τον καθορισμό του ground truth δεν αφορά μόνο στην διαδικασία της αξιολόγησης αλλά και σε αυτή της εκπαίδευσης. Στη δεύτερη περίπτωση, τα αντικείμενα που επισημάνθηκαν χρησιμοποιούνται προκειμένου να εξαχθούν χαρακτηριστικά σημεία (Haar, LBP, HOG feature descriptors) ώστε να εκπαιδευτεί στη συνέχεια ένας κατηγοριοποιητής⁴⁵ (Krig, 2016).

Τα συστήματα που χρησιμοποιήθηκαν στη συγκεκριμένη μεταπτυχιακή διατριβή για την επισήμανση των εικόνων είναι το Imglab, το Object Marker, και το Cascade Trainer GUI. Λεπτομέρειες για τη λειτουργία τους αναλύονται στο Παράρτημα «Α».

2.5.2 Ακρίβεια Προσδιορισμού Θέσης

Κατά τη διαδικασία αξιολόγησης ενός μοντέλου ανίχνευσης, αφού εφαρμοστεί σε ένα σύνολο εικόνων, θα εξάγει στο τέλος τις συντεταγμένες των ορθογωνίων, σύμφωνα με την εκτίμηση που κάνει για την ύπαρξη των αντικειμένων ενδιαφέροντος στα συγκεκριμένα σημεία. Αυτές οι συντεταγμένες συνήθως δεν συμπίπτουν με τις αντίστοιχες συντεταγμένες που έχουν προσδιοριστεί από τον ερευνητή (ground truth data) κατά την επισήμανση του validation/test set. Οπότε, το ερώτημα που προκύπτει είναι πόσο σωστή είναι η ανίχνευση που πραγματοποιήθηκε από το μοντέλο σε σχέση με την πραγματικότητα που καθορίστηκε από τον ερευνητή.



Εικόνα 22. Υπολογισμός IoU (Rosebrock, 2016)

⁴³ Η εργασία της καταγραφής του αριθμού και της θέσης ενός αντικείμενου πάνω σε μία εικόνα μπορεί να γίνει είτε αυτόματα είτε χειροκίνητα και ονομάζεται επισήμανση (annotation).

⁴⁴ Οι συντεταγμένες δίνονται με τη μορφή (x,y,w,h) όπου x,y οι καρτεσιακές συντεταγμένες της άνω δεξιάς γωνίας του ορθογωνίου, w το πλάτος και h το ύψος αυτού.

⁴⁵ Για παράδειγμα οι AdaBoost και SVM, για τους οποίους έγινε αναφορά σε προηγούμενες ενότητες

Το μέτρο που χρησιμοποιείται για τον καθορισμό της ακρίβειας της ανίχνευσης ενός αντικειμένου, είναι ο λόγος της τομής προς την ένωση του ορθογωνίου που εξάγεται από το μοντέλο ανίχνευσης και του αντίστοιχου που έχει επισημανθεί ως ground truth (Intersection over Union – IoU). Λόγω πλήθους παραμέτρων όπως το μέγεθος του κυλιόμενου παραθύρου που χρησιμοποιείται κατά την ανίχνευση, τη μέθοδο εξαγωγής των χαρακτηριστικών σημείων κλπ, είναι απίθανο να υπάρχει πλήρης ταύτιση μεταξύ των δύο αυτών ορθογωνίων. Όσο μεγαλύτερη είναι η επικάλυψη μεταξύ τους τόσο υψηλότερη η τιμή του IoU και επομένως πιο ακριβές το μοντέλο.

2.5.3 Μήτρα Σύγχυσης

Η μήτρα σύγχυσης (confusion matrix) είναι ένας πίνακας ο οποίος χρησιμοποιείται συχνά προκειμένου να περιγράψει την απόδοση ενός μοντέλου κατηγοριοποίησης, λαμβάνοντας υπόψη το πλήθος των εγγραφών που έχουν προβλεφθεί ορθά ή λανθασμένα (Tan, 2016). Ο πίνακας 2 παρουσιάζει τη μήτρα σύγχυσης για ένα πρόβλημα δυαδικής κατηγοριοποίησης, σαν αυτά που θα παρουσιαστούν στη συνέχεια της εργασίας

		Προβλεφθείσα Κατηγορία	
		Αρνητική	Ορθή
Πραγματική Κατηγορία	Αρνητική	a	b
	Ορθή	c	d

Πίνακας 2. Δυαδική Μήτρα Σύγχυσης (Kohavi & Provost, 1998)

Οι τιμές του πίνακα ερμηνεύονται όπως φαίνεται στη συνέχεια:

- a : Είναι ο αριθμός των σωστών προβλέψεων για τις αρνητικές περιπτώσεις (δηλ. να μην εμφανιστεί ορθογώνιο παραλληλόγραμμο που να δείχνει ανίχνευση οχήματος, σε σημείο του χώρου, όπου πραγματικά δεν υπάρχει). Στις μετρήσεις αναφέρεται ως TN (True Negative), αλλά στην περίπτωση της ανίχνευσης αντικειμένων δε χρησιμοποιείται.
- b : Είναι ο αριθμός των προβλέψεων που εσφαλμένα κατηγοριοποιούνται ως ορθές (δηλ. να ανιχνεύσει το μοντέλο την ύπαρξη οχήματος σε μια συγκεκριμένη θέση στην εικόνα, ενώ στην πραγματικότητα δεν υπάρχει). Στις μετρήσεις θα αναφέρεται ως FP (False Positive).
- c : Είναι ο αριθμός των προβλέψεων που εσφαλμένα κατηγοριοποιούνται ως

αρνητικές (δηλ. να μην ανιχνευτεί η ύπαρξη ενός οχήματος σε μία θέση, ενώ υπάρχει στην πραγματικότητα). Σε αυτή την κατηγορία θα εντάσσονται και οι περιπτώσεις όπου έχει γίνει ανίχνευση του οχήματος αλλά η σημειωθείσα τιμή του IoU είναι κάτω της τιμής κατωφλίου. Στις μετρήσεις θα αναφέρεται ως FN (False Negative).

- d : Είναι ο αριθμός των προβλέψεων που ορθώς κατηγοριοποιήθηκαν ως ορθές (δηλ. να ανιχνευτεί η ύπαρξη ενός οχήματος σε μία θέση όπου υπάρχει και στην πραγματικότητα). Στις μετρήσεις θα αναφέρεται ως TP (True Positive).

Επομένως, κατά την αξιολόγηση των μοντέλων ανίχνευσης αντικειμένων, ο συνολικός αριθμός των ορθών προβλέψεων δίνεται από την τιμή του d, ενώ ο αντίστοιχος αριθμός των εσφαλμένων προβλέψεων προκύπτει από το άθροισμα των τιμών "b+c".

2.5.4 Μέτρα Απόδοσης

Οι παραπάνω πληροφορίες που περιέχει η μήτρα σύγχυσης, είναι δυνατόν να συμπεριληφθούν στα παρακάτω μέτρα σύγκρισης της απόδοσης διαφορετικών κατηγοριοποιητών⁴⁶:

- Η Συνολική Ορθότητα (Accuracy) αποτελεί το σημείο εκκίνησης για τα μέτρα εκτίμησης απόδοσης και δίνεται από το πηλίκο των ορθών προβλέψεων προ το γενικό σύνολο των προβλέψεων.

$$\text{Accuracy} = \frac{d}{b+c+d} = \frac{\text{TP}}{\text{FP}+\text{FN}+\text{TP}}$$

- Ισοδύναμα, η απόδοση του μοντέλου μπορεί να μετρηθεί με το Ρυθμό Σφάλματος (Error Rate) που δίνεται από το πηλίκο των λανθασμένων προβλέψεων προς το συνολικό αριθμό αυτών.

$$\text{Error Rate} = \frac{b+c}{b+c+d} = \frac{\text{FP}+\text{FN}}{\text{FP}+\text{FN}+\text{TP}} = \frac{1}{\text{Accuracy}}$$

- Ακρίβεια (Precision) είναι ο αριθμός των ορθών προβλέψεων διαιρούμενος με το άθροισμα των ορθών προβλέψεων και των εσφαλμένων θετικών προβλέψεων. Καλείται και προβλεπτική τιμή ορθής κατηγοριοποίησης (Positive

⁴⁶ Στη βιβλιογραφία αποτυπώνονται και άλλες τιμές όπως : True Positive Rate, False Positive Rate, Specificity κλπ (Fawcett, 2006).

Predictive Value –PPV). Μέσω του precision δίδεται η απάντηση στο ερώτημα : «από όσα αντικείμενα ο κατηγοριοποιητής εκτίμησε ως ορθά, ποιο είναι το ποσοστό αυτών που ήταν πραγματικά σωστό;». Χαμηλή τιμή precision αποτελεί ένδειξη για μεγάλο αριθμό εσφαλμένων θετικών προβλέψεων (FP).

$$\text{Ακρίβεια (Precision)} = \frac{d}{b+d} = \frac{TP}{TP+FP}$$

- Αριθμός Εύστοχων Απαντήσεων (Recall) είναι ο αριθμός των ορθών προβλέψεων διαιρούμενος με το άθροισμα των ορθών προβλέψεων και των εσφαλμένων αρνητικών προβλέψεων. Καλείται και ευαισθησία (sensitivity) ή ρυθμός θετικών προβλέψεων (true positive rate). Μέσω του precision δίδεται η απάντηση στο ερώτημα : «από το σύνολο των αντικειμένων του δείγματος, ποιο ποσοστό αυτών εντόπισε το μοντέλο;». Χαμηλή τιμή recall αποτελεί ένδειξη για μεγάλο αριθμό εσφαλμένων αρνητικών προβλέψεων (FN).

$$\text{Recall} = \frac{d}{c+d} = \frac{TP}{TP+FN}$$

Επειδή στην πραγματικότητα οι τιμές recall και precision δρουν ανταγωνιστικά μεταξύ τους, συνήθως χρησιμοποιείται ο δείκτης F ο οποίος αποτελεί τον αρμονικό αριθμητικό μέσο των precision και recall.

$$\text{Δείκτης F (F-measure)} = \frac{2*(\text{Precision}*\text{Recall})}{\text{Precision}+\text{Recall}}$$

Κεφάλαιο

3

Ιχνηλάτηση Αντικειμένων μέσω Video Streaming

3.1 Ιχνηλάτηση

Ενώ στην περίπτωση της ανίχνευσης (detection) το αντικείμενο εντοπίζεται σε κάθε εξεταζόμενο πλαίσιο του video, ανεξαρτήτως της πληροφορίας που υπήρχε για αυτό σε προηγούμενα πλαίσια, η έννοια της ιχνηλάτησης (tracking) έχει νόημα μόνο σε εφαρμογές video και ορίζεται ως ο συνεχής εντοπισμός της θέσης του αντικειμένου ενδιαφέροντος σε διαδοχικά πλαίσια. Σκοπός της ιχνηλάτησης είναι η συσχέτιση των αντικειμένων ενδιαφέροντος (στόχος) μεταξύ συνεχόμενων και διαδοχικών πλαισίων. Πιο συγκεκριμένα, μέσω αυτής εξάγεται μία πρόβλεψη για τη θέση του αντικειμένου ενδιαφέροντος, στο επόμενο πλαίσιο, χρησιμοποιώντας αλγόριθμους προβλεπτικής δυναμικής (estimated dynamics). Στη συνέχεια ενημερώνονται τα στοιχεία της πρόβλεψης βάσει των μετρήσεων της πραγματικής κίνησης.

3.1.1 Ανάλυση της Τεχνικής της Ιχνηλάτησης

Η ιχνηλάτηση διακρίνεται από τις ακόλουθες προκλήσεις:

- Είναι δυσχερής ο υπολογισμός της οπτικής ροής (optical flow) παντού καθώς πρέπει να είναι γνωστό το είδος της συμπεριφοράς κάθε κλίσης (gradient) στην εικόνα, προκειμένου είναι εφικτή η ιχνηλάτηση συγκεκριμένων χαρακτηριστικών (features)
- Η συσχέτιση μπορεί να αποβεί ιδιαίτερος δύσκολη αποστολή, ειδικά αν ο στόχος κινείται με μεγάλη ταχύτητα, σε σχέση με το ρυθμό εναλλαγής των πλαισίων (fps)
- Το όποιο σφάλμα σε κάθε διαδοχικό πλαίσιο μπορεί να λειτουργήσει προσθετικά ώστε στο τέλος το συνολικό σφάλμα που δημιουργείται να είναι τόσο μεγάλο που να μην επιτρέψει τη συνέχιση της ιχνηλάτησης.

- Τέλος, προβλήματα δημιουργούν και οι περιπτώσεις παροδικής απόκρυψης ενός αντικειμένου.

Η διαδικασία της ιχνηλάτησης αποτελείται από δύο βασικά βήματα, τα οποία λειτουργούν μεταξύ τους ως ένας βρόγχος, την πρόβλεψη (prediction) και τη διόρθωση (correction). Με το πρώτο βήμα πραγματοποιείται μία πρόβλεψη για την μελλοντική κατάσταση του αντικειμένου, έχοντας ως δεδομένο τις προηγούμενες μετρήσεις. Λαμβάνεται η $P(\frac{X_t}{Y_0=y_0, y_1, \dots, y_{t-1}})$, όπου όλες οι μετρήσεις πριν τη χρονική στιγμή t συμβάλουν στη δημιουργία της πρόβλεψης P για το χρόνο t . Αντίστοιχα, στην περίπτωση της διόρθωσης, εξάγεται μία επικαιροποιημένη (updated) εκτίμηση για την κατάσταση του αντικειμένου βάσει της πρόβλεψης αλλά πλέον και της μέτρησης Y_t τη χρονική στιγμή t , με $P(\frac{X_t}{Y_0=y_0, y_1, \dots, y_{t-1}}, Y_t = y_t)$. (Bobick, 2011-18)

Η πρόβλεψη και η διόρθωση δίνεται αντίστοιχα από τις παρακάτω σχέσεις (ibid):

P R E D I C T I O N	$\int P(\frac{X_t}{X_{t-1}}) * P(\frac{X_{t-1}}{y_0, y_1, \dots, y_{t-1}}) * dX_{t-1}$	<p>Η σχέση της πρόβλεψης δίνει την πιθανότητα ότι το X θα βρίσκεται σε κάποιο συγκεκριμένο σημείο. Αποτελεί το άθροισμα ή το ολοκλήρωμα του να βρεθεί το X στο συγκεκριμένο σημείο (X_t) από οποιοδήποτε (X_{t-1}) στο οποίο βρισκόταν. Η πρόβλεψη έχει ολοκληρωθεί (integrated) βάσει της πιθανότητας ότι το αντικείμενο έχει βρεθεί στα σημεία ($\frac{X_{t-1}}{y_0, y_1, \dots, y_{t-1}}$) που περιγράφει η σχέση.</p>
E S T I M A T I O N	<p>Μοντέλο Παρατήρησης (Observation Model) Εκτίμηση Πρόβλεψης (Prediction Estimate)</p> $\frac{P(\frac{Y_t}{X_t}) * P(\frac{X_t}{y_0, y_1, \dots, y_{t-1}})}{P(\frac{Y_t}{y_0, y_1, \dots, y_{t-1}})}$ <p>Παράγοντας Κανονικοποίησης</p>	<p>Από τη στιγμή που πραγματοποιείται μία μέτρηση είναι δυνατόν να λάβει χώρα και η διόρθωση με τη χρήση των κανόνων Bayes. Στη σχέση δίνεται η εκτίμηση της πρόβλεψης και το μοντέλο παρατήρησης (δηλ. "Εάν το σώμα βρισκόταν σε ένα συγκεκριμένο σημείο y_t, ποια θα ήταν η πιθανότητα να ληφθεί η μέτρηση που μόλις καταγράφηκε;"). Στη συνέχεια με τη χρήση του κανόνα του Bayes αυτό μετατρέπεται σε πρόβλεψη του X_t δεδομένων των μετρήσεων, συμπεριλαμβανομένης της y_t (δηλ $P(y_0, y_1, \dots, y_{t-1}, y_t)$)</p>

Παραδείγματα αλγορίθμων που χρησιμοποιούνται για την ιχνηλάτηση είναι (Mallick, 2017):

- Dense Optical Flow: Ο συγκεκριμένος αλγόριθμος συνίσταται στον υπολογισμό του ανύσματος κίνησης κάθε pixel σε ένα πλαίσιο

- Kanade – Lucas – Tomashi (KLT): Αποτελεί έναν αλγόριθμο ιχνηλάτησης χαρακτηριστικών σημείων (features) σε διαδοχικά πλαίσια.
- Kalman Filtering : Αποτελεί έναν από τους πιο συχνά χρησιμοποιούμενους αλγόριθμους ιχνηλάτησης ο οποίος χρησιμοποιείται για να προβλεφθεί η νέα θέση ενός αντικειμένου, βάσει της πληροφορίας κίνησης που διατίθεται για αυτόν.
- Meanshift και Camshift: Οι συγκεκριμένοι αλγόριθμοι χρησιμοποιούνται για τον υπολογισμό του μέγιστου μιας συνάρτησης πυκνότητας.

Ο λόγος χρησιμοποίησης της τεχνικής της ιχνηλάτησης, αντί μόνο της ανίχνευσης, είναι κατ' αρχήν η ταχύτητα που προσφέρει η πρώτη. Οι αλγόριθμοι αυτής είναι ταχύτεροι από αυτούς της ανίχνευσης. Έτσι, στην περίπτωση της ιχνηλάτησης, αντί να δαπανάται χρόνος για έρευνα ολόκληρου του πλαισίου προκειμένου να εντοπιστεί το αντικείμενο ενδιαφέροντος, χρησιμοποιείται η προβλεπτική ικανότητα του αλγορίθμου ώστε να ερευνηθεί μια μικρή περιοχή πέριξ της πιθανής επόμενης θέσης του αντικειμένου. Συνήθως, όπως θα αναλυθεί και στο κεφάλαιο της μεθοδολογίας, αρχικά εφαρμόζεται ο αλγόριθμος ανίχνευσης και στη συνέχεια, για έναν αριθμό πλαισίων που ακολουθούν, χρησιμοποιείται ο αλγόριθμος ιχνηλάτησης. Αυτό συμβαίνει προκειμένου (1) να αντιμετωπίζονται τυχόν συσσωρευτικά σφάλματα του αλγορίθμου ιχνηλάτησης (drift) και (2) να ελεγχθεί η εικόνα για είσοδο νέων αντικειμένων ώστε να οριστεί η ιχνηλάτησή τους. Εκτός της ταχύτητας επεξεργασίας, η ιχνηλάτηση δρα εποικοδομητικά και στις περιπτώσεις που το αντικείμενο ενδιαφέροντος βρεθεί πίσω από εμπόδια που δεν επιτρέπουν την ανίχνευσή του. Σε αυτή τη περίπτωση, ο εν λόγω αλγόριθμος μπορεί να ανταπεξέλθει για ένα περιορισμένο αριθμό πλαισίων.

3.1.2 Αλγόριθμος Συσχετιστικής Ιχνηλάτησης (Correlation Tracker)

Ο εν λόγω αλγόριθμος αποτελεί ένα εργαλείο για την ιχνηλάτηση των αντικειμένων ενδιαφέροντος σε ένα video. Σύμφωνα με τον Danelljan και άλλους, που παρουσίασαν τον συγκεκριμένο αλγόριθμο, τα χαρακτηριστικά του είναι η ικανότητα επαρκούς ανταπόκρισης στις περιπτώσεις μεγάλων μεταβολών στο μέγεθος (κλίμακα) του αντικειμένου και η δυνατότητα να λειτουργεί σε συνθήκες χαμηλού ρυθμού εναλλαγής πλαισίων γεγονός το οποίο σε συνδυασμό με την ταχύτητά του επιτρέπει να λειτουργεί σε πραγματικό χρόνο (Danelljan, Häger, Khan, & Felsberg, 2014).

Ο αλγόριθμος έχει την παρακάτω μορφή (ibid):

Είσοδος:

Εικόνα I_t

Προηγούμενη θέση του αντικειμένου p_{t-1} και κλίμακα s_{t-1}

Μοντέλο μετατόπισης A_{t-1}^{trans} , B_{t-1}^{trans} και μοντέλο κλίμακας A_{t-1}^{scale} , B_{t-1}^{scale}

Έξοδος

Εκτιμώμενη θέση του αντικειμένου p_t και κλίμακα s_t

Επικαιροποιημένο μοντέλο μεταφοράς A_t^{trans} , B_t^{trans} και μοντέλο κλίμακας A_t^{scale} , B_t^{scale}

Εκτίμηση Μεταφοράς:

1. Εξήγαγε ένα δείγμα μεταφοράς z_{trans} από την εικόνα I_t στο p_{t-1} και κλίμακα s_{t-1}
2. Υπολόγισε τη συσχέτιση μεταφοράς y_{trans} χρησιμοποιώντας τις z_{trans} , A_{t-1}^{trans} και B_{t-1}^{trans} στην $y = \mathcal{F}^{-1} \left\{ \frac{\sum_{l=1}^d \overline{A^l} Z^l}{B+\lambda} \right\}$

3. Θέσε p_t τη θέση του αντικειμένου η οποία μεγιστοποιεί την y_{trans}

Εκτίμηση Κλίμακας:

1. Εξήγαγε ένα δείγμα μεταφοράς z_{scale} από την εικόνα I_t στο p_t και κλίμακα s_{t-1}
2. Υπολόγισε τη συσχέτιση μεταφοράς y_{scale} χρησιμοποιώντας τις z_{scale} , A_{t-1}^{scale} και B_{t-1}^{scale} στην $y = \mathcal{F}^{-1} \left\{ \frac{\sum_{l=1}^d \overline{A^l} Z^l}{B+\lambda} \right\}$
3. Θέσε s_t τη θέση του αντικειμένου η οποία μεγιστοποιεί την y_{scale}

Επικαιροποίηση Μοντέλου:

1. Εξήγαγε δείγματα f_{trans} και f_{scale} από την I_t στο p_t και κλίμακα s_t
2. Επικαιροποίησε το μοντέλο μεταφοράς A_t^{trans} και B_t^{trans} , χρησιμοποιώντας την
$$\begin{cases} A_t^l = (1 - \eta)A_{t-1}^l + \eta \overline{G}_t F_t^l \\ B_t = (1 - \eta)B_{t-1} + \eta \sum_{k=1}^d \overline{F}_t^k F_t^k \end{cases}$$
3. Επικαιροποίησε το μοντέλο κλίμακας A_t^{scale} και B_t^{scale} , χρησιμοποιώντας την
$$\begin{cases} A_t^l = (1 - \eta)A_{t-1}^l + \eta \overline{G}_t F_t^l \\ B_t = (1 - \eta)B_{t-1} + \eta \sum_{k=1}^d \overline{F}_t^k F_t^k \end{cases}$$

όπου:

\mathcal{F}^{-1} : ο αντίστροφος τελεστής DFT

Z: Τμήμα της εικόνας (patch) διαστάσεων MXN

η : Παράμετρος ρυθμού μάθησης

Τα A_t^l και B_t αποτελούν τον αριθμητή και παρονομαστή αντίστοιχα του φίλτρου συσχέτισης (correlation filter) H_t^l

Η παράμετρος λ ($\lambda \geq 0$) ελέγχει την επίδραση του όρου ρύθμισης

Τα F και G αποτελούν αντίστοιχα τις discrete Fourier transforms (DFTs) των patches που περιέχουν το αντικείμενο ενδιαφέροντος και της εξόδου συσχέτισης αυτών από το φίλτρο (χρησιμοποιούνται ως δείγματα εκπαίδευσης).

3.2 Video

Το video⁴⁷ είναι μια σειρά από χρονολογημένες εικόνες, κάθε μία από τις οποίες ονομάζεται πλαίσιο (frame). Ενώ στην περίπτωση της εικόνας που αναλύθηκε στην ενότητα 2.2 γίνεται λόγος για χωρική κατανομή της έντασης της φωτεινότητας, στο video προστίθεται και μία ακόμη διάσταση, η χρονική, όπου λαμβάνεται υπόψη η μεταβολή της φωτεινότητας, για κάθε pixel, σε κάθε πλαίσιο που ακολουθεί (Kokaram,). Όταν ο ρυθμός διαδοχής αυτής της αλληλουχίας από εικόνες ξεπερνά ένα συγκεκριμένο όριο, τότε ο θεατής αντιλαμβάνεται την ύπαρξη μιας συνεχούς ροής η ομαλότητα της οποίας είναι ανάλογη του αριθμού των πλαισίων που παρουσιάζονται στη μονάδα του χρόνου (frames per second – fps). Ο ρυθμός εναλλαγής των εικόνων μπορεί να κυμαίνεται στα 24 fps για τις ταινίες του κινηματογράφου και στα 25 fps (PAL, SECAM) ή 30 fps (NTSC)⁴⁸ για τις τηλεοράσεις (Adobe, 2006).

Η αρχική μορφή του video ήταν αναλογική και συντίθονταν από μία κλίμακα γκρι χρωματικής διαβάθμισης όπου στο ένα άκρο της κλίμακας βρίσκονταν το μαύρο χρώμα και στο άλλο άκρο το άσπρο (Jack, 2011). Με το πέρασμα των χρόνων και την εξέλιξη της τεχνολογίας η εικόνα έγινε έγχρωμη. Αρχικά έγινε χρήση του αναλογικού προτύπου RGB (red, green, blue) ενώ στη συνέχεια του YIQ ή YUV προκειμένου να διατηρηθεί το ίδιο εύρος ζώνης⁴⁹ με αυτό της ασπρόμαυρης εικόνας (ibid). Το επόμενο σημαντικό βήμα στην εικονοληψία σχετίζεται με την έλευση των ολοκληρωμένων ψηφιακών

⁴⁷ Σύμφωνα με το διαδικτυακό ετοιμολογικό λεξικό, η ρίζα της λέξης video είναι λατινική, προερχόμενη από το "videre" που μεταφράζεται στα ελληνικά ως "να δω (to see)", και υιοθετήθηκε το 1935 κατά αντιστοιχία της αγγλικής λέξης audio (<https://www.etymonline.com/word/video>, Πρόσβαση:16/1/18)

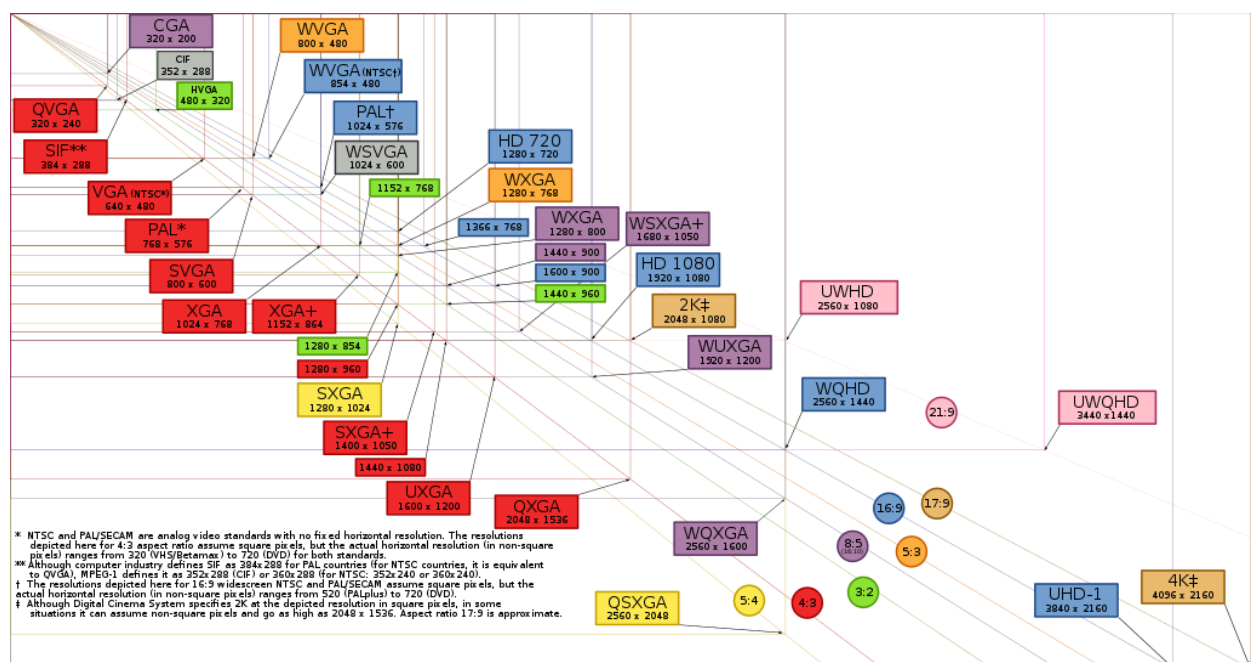
⁴⁸ Λόγω του ότι τα συγκεκριμένα πρότυπα αφορούν interlaced scanning (διεμπλεκόμενη σάρωση ή γραμμή παρά γραμμή) και όχι σε progressive (προοδευτική), ο αντίστοιχος ρυθμός εναλλαγής των δύο πεδίων (upper & lower fields) είναι 50 και 60 αντίστοιχα.

⁴⁹ Το RGB απαιτούσε τριπλάσιο εύρος ζώνης για την ίδια ασπρόμαυρη εικόνα.

κυκλωμάτων και των ηλεκτρονικών υπολογιστών γεγονός που οδήγησε στην ψηφιοποίηση του video (Apostoloroulos, Tan, & Wee, 2002).

3.2.1 Ανάλυση (Video Resolution)

Η ποιότητα του video, εκτός από τον ρυθμό εναλλαγής πλαισίων (fps) εξαρτάται και από την ανάλυση. Όσο μεγαλύτερη η ανάλυση τόσο καλύτερη η ποιότητα. Στο αναλογικό video η ανάλυση εξαρτιόνταν από τον αριθμό των γραμμών που σάρωναν τα ηλεκτρόνια για την αναπαραγωγή της εικόνας (vertical resolution). Ενώ για τα πρότυπα PAL και NTSC η κατακόρυφη ανάλυση ήταν 625 και 525 γραμμές αντίστοιχα, στην πραγματικότητα μόνο οι 576 για το PAL και οι 486 για το NTSC ήταν ορατές στο θεατή, καθώς οι υπόλοιπες χρησιμοποιούνταν για το συγχρονισμό και σήμα αμαύρωσης (blanking). Αντίστοιχα η οριζόντια ανάλυση αναφέρονταν στον αριθμό των μεταβολών μεταξύ μαύρου-άσπρου και αφορούσε στο εύρος ζώνης (bandwidth) ή στην απόκριση συχνότητας (frequency response) (Tektronix, 2000).



Εικόνα 23. Αναλύσεις video (Wikipedia.org, 2018)

Στην περίπτωση της ψηφιακής εικόνας, η μέτρηση της ανάλυσης, ιδιαίτερα της οριζόντιας είναι πιο απτή καθώς αναφέρεται σε συγκεκριμένο αριθμό από pixel. Συνήθως παρουσιάζεται με δύο αριθμούς όπου ο πρώτος σχετίζεται με τον αριθμό των οριζόντιων pixel και ο δεύτερος με τα κατακόρυφα. Οι συνήθειες αναλύσεις ψηφιακού video είναι η 480p (480 γραμμές οριζόντιας ανάλυσης και προοδευτικής σάρωσης), η

720p (High Defenition – HD), η 1080i (1080 γραμμές οριζόντιας ανάλυσης και διεμπλεκόμενης σάρωσης), η 1080p (Full HD - FHD) και προσφάτως η 2160p (Ultra HD – UHD). Το video διατήρησε και το χαρακτηριστικό του aspect ratio των εικόνων.

3.2.2 Συμπίεση Video

Με τη συμπίεση επιδιώκεται αφενός μεν η μείωση του όγκου της πληροφορίας που μεταφέρει μία ψηφιακή εικόνα, προκειμένου να είναι ευχερέστερη η διαχείρισή της (αποθήκευση, επεξεργασία, μεταφορά μεταξύ ψηφιακών μέσων απεικόνισης κλπ⁵⁰), αφετέρου δε, η διατήρηση της υψηλής ποιότητάς της (Adobe, 2004). Το ποσοστό συμπίεσης εκφράζεται ως ο λόγος μεταξύ του μεγέθους του αρχικού video προς το συμπιεσμένο. Στους τρόπους μείωσης του όγκου της πληροφορίας δύναται να συμπεριληφθεί η:

- Μείωση της ανάλυσης (ένα video ανάλυσης 960X540 περιλαμβάνει τέσσερις φορές λιγότερη πληροφορία από το ίδιο video σε ανάλυση 1920X1080)
- Μείωση του αριθμού fps (αν τα 25fps προοδευτικής σάρωσης μειωθούν σε 12fps, ο όγκος του video θα έχει μειωθεί σχεδόν κατά το ήμισυ). Εδώ θα πρέπει να αναφερθεί ότι ο ελάχιστος ρυθμός αλλαγής εικόνων προκειμένου το video να δείχνει ομαλό, είναι τα 10 fps (Adobe, 2006).

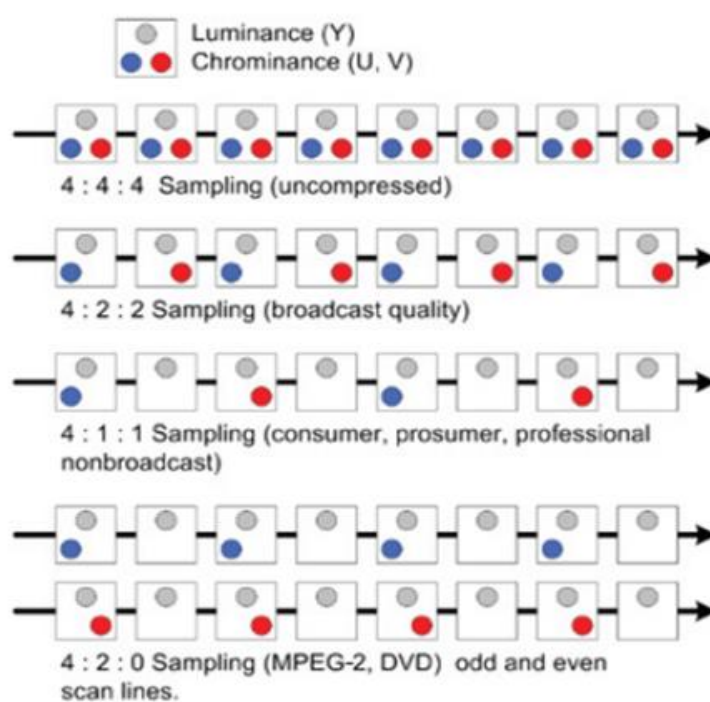
Encoding	Bit rate savings relative to			
	H.264/MPEG-4 AVC HP	H.263 CHC	MPEG-4 ASP	MPEG-2/H.262 MP
HEVC MP	35.4%*	63.7%	65.1%	70.8%
H.264/MPEG-4 AVC HP		44.5%	46.6%	55.4%
H.263 CHC			3.9%	19.7%
MPEG-4 ASP				16.2%

Εικόνα 24. Μέσος όρος κέρδους Bit Rate μεταξύ διαφορετικών προτύπων συμπίεσης (το PSNR διατηρείται σταθερό) (Ohm, J.:Sullivan G.:Tan T.:Wiegand T, 2012)

Οι παραπάνω απλές μορφές συμπίεσης δεν δύναται όμως να αποδώσουν το επιθυμητό ποιοτικό αποτέλεσμα όταν πρόκειται για οθόνη προβολής μεγάλων διαστάσεων ή για

⁵⁰ Ένα δευτερόλεπτο ασυμπίεστου έγχρωμου (8-bit) ψηφιακού video ανάλυσης 1080p περιλαμβάνει περίπου $1,25 \cdot 10^9$ bits ($1280 \cdot 1920 \cdot 3 \cdot 8 \cdot 25$)

υπηρεσίες συνδρομητικής τηλεόρασης. Η επίλυση του προβλήματος στηρίζεται στην παρατήρηση ότι η ανθρώπινη όραση δεν αντιλαμβάνεται μικρές αλλαγές οι οποίες μπορεί να λάβουν χώρα στα πλαίσια των εικόνων, όπως για παράδειγμα ανεπαίσθητες χρωματικές αλλαγές ή αλλαγές στη φωτεινότητα. Τη συγκεκριμένη αδυναμία εκμεταλλεύονται τα διάφορα πρότυπα κωδικοποίησης προκειμένου να συμπίεσουν ένα video⁵¹. Όταν το πλαίσιο ενός video συμπίεζεται ανεξαρτήτως της πληροφορίας που περιέχεται στα προηγούμενα ή στα επόμενα πλαίσια, ο τύπος της συμπίεσης είναι γνωστός ως intraframe ή χωρική (spatial) συμπίεση. Σε αντίθετη περίπτωση, ονομάζεται interframe ή χρονική (temporal) συμπίεση και στηρίζεται στο γεγονός ότι τα "γειτονικά" πλαίσια περιλαμβάνουν συναφείς πληροφορίες (Adobe, 2006).



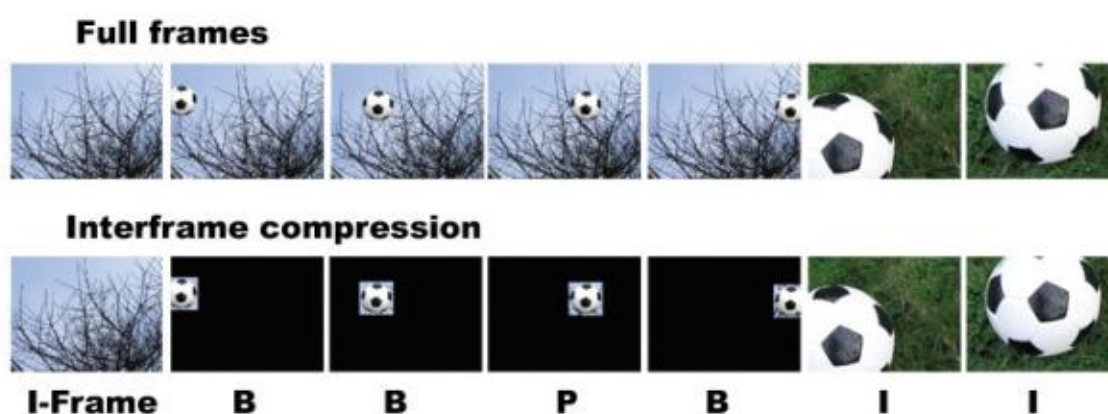
Εικόνα 25. Δειγματοληψία των στοιχείων YUV για τη συμπίεση από 4:4:4 σε 4:2:2, 4:1:1 και 4:2:0

Τα περισσότερα πρότυπα συμπίεσης (codecs) εκμεταλλεύονται την intraframe συμπίεση η οποία στηρίζεται κατά βάση στην απόρριψη μέρους των πληροφοριών για το χρώμα που περιέχονται στα πλαίσια. Ο χρωματικός κώδικας (color space) που χρησιμοποιείται για τη συμπίεση είναι ο YUV καθώς λαμβάνει υπόψη του το ανθρώπινο αισθητήριο όργανο της όρασης. Ο οφθαλμός εμφανίζεται περισσότερο ευαίσθητος στις αλλαγές της φωτεινότητας (luminance – Y) σε μια εικόνα σε σχέση με τις χρωματικές

⁵¹ Λόγω της επαγόμενης απώλειας πληροφορίας, αυτός ο τύπος συμπίεσης αναφέρεται ως lossy (απωλειακός).

διαφορές (chrominance - UV)⁵². Επομένως, μέσω κατάλληλων αλγορίθμων μειώνεται η χρωματική πληροφορία και καθίσταται έτσι εφικτή η μείωση του εύρους ζώνης του video (ibid), χωρίς να επηρεάζεται η αντιληπτή ποιότητά του.

Ένας αριθμός από codecs χρησιμοποιεί την interframe συμπίεση. Σε αυτή την περίπτωση αντί να αποθηκεύεται το σύνολο της πληροφορίας που περιέχει το πλαίσιο, αποθηκεύονται μόνο τα δεδομένα που είναι διαφορετικά. Έτσι δημιουργούνται τρεις διαφορετικοί τύποι πλαισίου: (1) τα πλαίσια τύπου I τα οποία αποτελούν και τα βασικά καθώς το μόνο είδος συμπίεσης στο οποίο υπόκεινται είναι η intraframe (2) Τα P πλαίσια, τα οποία είναι προβλεπτικά και οι πληροφορίες τους⁵³ στηρίζονται σε προηγούμενα πλαίσια (3) Τα B πλαίσια τα οποία παρεμβάλλονται (interpolated) από προηγούμενα και επόμενα I και P πλαίσια για αντιστάθμιση της εικόνας. Μία τυπική ακολουθία των παραπάνω πλαισίων μπορεί να λάβει τη μορφή «I B B P B B P B B I» (ibid)



Εικόνα 26. Απεικόνιση της χρονικής συμπίεσης. Επειδή μόνο η μπάλα έχει αλλάξει θέση στις διαδοχικές εικόνες, χρησιμοποιώντας ένα διάνυσμα κίνησης (motion vector) περιγράφεται: (1) η μετατόπιση θέσης του block το οποίο κωδικοποιείται/ αποκωδικοποιείται στο τρέχον πλαίσιο και (2) η θέση του block με την καλύτερη αντιστοίχιση στο πλαίσιο αναφοράς (Adobe,2006)

Το πρότυπο συμπίεσης που χρησιμοποιείται στη συγκεκριμένη μεταπτυχιακή διατριβή είναι το H.264 (ή MPEG-4 Part 10 Advanced Video Coding) το οποίο επιτρέπει την

⁵² Το μεταδιδόμενο σήμα της τηλεόρασης χρησιμοποιεί το ήμισυ της chrominance πληροφορίας σε σχέση με τη luminance.

⁵³ Μπορεί να περιλαμβάνουν λιγότερο και από το 1/10 της πληροφορίας των I πλαισίων

κωδικοποίηση των αντικειμένων του video χρησιμοποιώντας αλγόριθμους χωρικής και χρονικής συμπίεσης καθώς και αντιστάθμισης της κίνησης (προσφέρει ακριβή περιγραφή των μετατοπίσεων για τα αντικείμενα που κινούνται). Η εξοικονόμηση σε bit rate μπορεί να φτάσει και το 50% συγκρινόμενη με το MPEG-2 part 2.

3.2.3 Video Streaming

Η ψηφιοποίηση της εικόνας, σε συνδυασμό με την διάδοση του internet καθώς επίσης και των προσωπικών υπολογιστών που καθιστούν εύκολη την επεξεργασία του video, οδήγησε στην απαίτηση για διακίνησή του video. Έτσι, με τον όρο Video streaming περιγράφεται «η μετάδοση συμπιεσμένου⁵⁴ συνήθως video, μέσω ενός ιδιωτικού ή δημοσίου δικτύου (internet, LAN, δορυφορικό), το οποίο μετά τη λήψη του από μία συσκευή υπολογιστικών δυνατοτήτων (τηλεόραση, έξυπνο κινητό, Η/Υ) αποσυμπιέζεται και αναπαράγεται» (Benes,).

Πλέον το video streaming αποτελεί στοιχείο της καθημερινότητας και η χρήση του έχει ουσιαστικά εκτοξευτεί, ως αποτέλεσμα της ευρείας διάδοσης των έξυπνων κινητών τηλεφώνων, της αύξησης του εύρους ζώνης, της συνεχούς αναβάθμισης των δικτύων κινητής τηλεφωνίας (3G, 4G) και των βελτιώσεων στον τομέα της συμπίεσης του video (Riiser, 2013).

Σύμφωνα με τον Do και άλλους, το video streaming λαμβάνει χώρα μέσω δύο βασικών μορφών, της ζωντανής μετάδοσης (live streaming) και της αποστολής/λήψης video κατά απαίτηση (Video on Demand - VoD) του χρήστη (Do, Hua, & Tantaoui, 2004). Μία ουσιαστική διαφορά μεταξύ αυτών των κατηγοριών είναι ότι κυρίαρχο στοιχείο για την αξιολόγηση της ποιότητας της παρεχόμενης υπηρεσίας στην πρώτη περίπτωση είναι το χρονικό διάστημα που μεσολαβεί (υστέρηση) μεταξύ της χρονικής στιγμής που λαμβάνει χώρα ένα πραγματικό γεγονός και του χρόνου που αναπαράγεται στο δέκτη του χρήστη (ibid). Από την άλλη, στην περίπτωση του VoD, ο χρόνος αναπαραγωγής του video δεν ενδιαφέρει και συνήθως οι μετρήσεις ποιότητας υπηρεσίας στρέφονται σε δείκτες όπως η παρεχόμενη ευκρίνεια, ο χρόνος που μεσολαβεί από την αίτηση για λήψη του video μέχρι την αναπαραγωγή του (καθυστέρηση/buffer time), η απώλεια πακέτων κλπ (Staelens et al., 2010).

⁵⁴ Η απαίτηση για συμπίεση του video πριν την αποστολή του οφείλεται στα όρια που θέτει το εύρος ζώνης που διατίθεται στα δίκτυα

Η απαίτηση για live streaming δεν αφορά μόνο τα μέσα κοινωνικής δικτύωσης ή την ειδησεογραφία αλλά και τον τομέα της επιτήρησης (Gualdi, Prati, & Cucchiara, 2008), αντικείμενο που άπτεται της συγκεκριμένης μεταπτυχιακής διατριβής. Εν προκειμένω, τα ζητούμενα είναι:

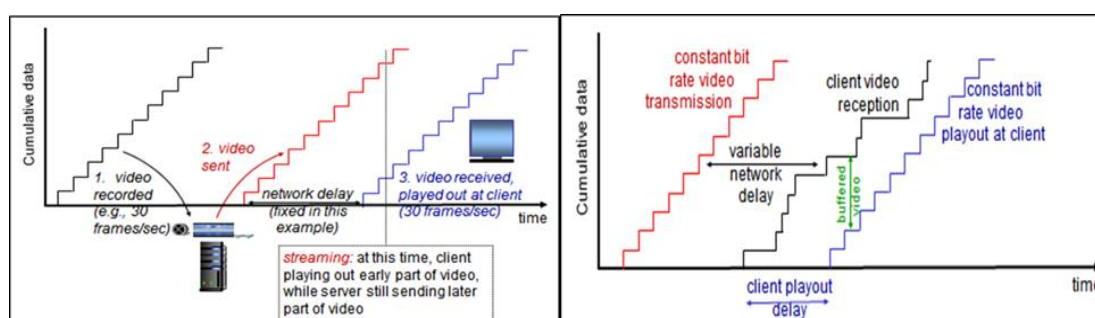
- Η μικρή καθυστέρηση (near real time streaming) κατά τη μεταφορά⁵⁵ του video, προκειμένου η λαμβανόμενη πληροφορία να είναι αξιοποιήσιμη. Λόγω του ότι δεν πρόκειται για interactive εφαρμογή τα χρονικά περιθώρια ολοκλήρωσης της παραπάνω διαδικασίας δεν είναι το ίδιο στενά με αυτά των interactive εφαρμογών για τα οποία ο αποδεκτός χρόνος δε θα πρέπει να ξεπερνά τα 150ms, (Apostolopoulos et al., 2002).
- Η εύρεση ενός σημείου ισορροπίας μεταξύ της ταχύτητας του χρησιμοποιούμενου δικτύου (εν προκειμένω το δίκτυο 3G/4G της κινητής τηλεφωνίας) και της ποιότητας της εικόνας στο δέκτη, ώστε να είναι δυνατή η επεξεργασία του video με τις τεχνικές του computer vision.

Η βασική ιδέα πίσω από το video streaming στηρίζεται στην κατάτμηση του video σε μέρη, την αποστολή αυτών και τη δυνατότητα του δέκτη να λάβει αυτά τα τμήματα, να τα τοποθετεί στη σωστή χρονολογική σειρά και να τα αναπαράγει (χωρίς να απαιτείται η αναμονή για λήψη του συνολικού αρχείου). Κάθε πλαίσιο του video χωρίζεται σε αριθμημένα πακέτα τα οποία φέρουν ξεχωριστή σφραγίδα έτσι ώστε να γνωρίζει το codec αναπαραγωγής του video σε ποια θέση πρέπει να τοποθετήσει τα πακέτα. Το video streaming μέσω του διαδικτύου περιλαμβάνει δυσκολίες καθόσον το τελευταίο παρέχει υπηρεσίες best effort. Τα στοιχεία εκείνα τα οποία μεταβάλλονται δυναμικά και επηρεάζουν τη ροή των δεδομένων είναι το εύρος ζώνης, η διακύμανση της καθυστέρησης (delay jitter) και ο ρυθμός απώλειας πακέτων (loss rate) (ibid).

Αναλυτικότερα, το εύρος ζώνης μεταβάλλεται χρονικά με αποτέλεσμα όταν ο πομπός μεταδίδει με μεγαλύτερο ρυθμό, να υπάρχει συμφόρηση και να χάνονται πακέτα ενώ στην αντίθετη περίπτωση να υπάρχουν καθυστερήσεις στο δέκτη. Επίσης, η διακύμανση της χρονικής διαφοράς που μεσολαβεί μεταξύ εκπομπής και λήψης δύναται να οδηγήσει σε προβλήματα κατά την αναδόμηση του video στην πλευρά του δέκτη. Τέλος, οι

⁵⁵ Το τμήμα της μεταφοράς περιλαμβάνει την λήψη του video, την κωδικοποίησή του προκειμένου να μειωθεί το μέγεθός του, την εκπομπή του μέσω του δικτύου κινητής τηλεφωνίας, την λήψη του από το σταθμό εδάφους, την αποκωδικοποίηση και επεξεργασία του.

απώλειες εξαρτώνται και από το είδος του δικτύου που χρησιμοποιείται. Στην περίπτωση των ενσύρματων δικτύων, αυτά είναι περισσότερο επιρρεπή στην απώλεια εξολοκλήρου ενός πακέτου ενώ αντίστοιχα τα ασύρματα δίκτυα εισάγουν σφάλματα δυαδικού ψηφίου (bit error) και σφάλματα ριπής (burst errors). Για την αντιμετώπιση των παραπάνω, επιστρατεύονται μέθοδοι όπως εμπροσθόδοτη διόρθωση σφάλματος (Forward Error Correction), η επανεκπομπή, η απόκρυψη σφάλματος, η εισαγωγή play-out buffer, η χρησιμοποίηση UDP πρωτοκόλλου έναντι του TCP για αποφυγή επιπλέον καθυστερήσεων σε χρονικά ευαίσθητα δεδομένα, κλπ



Εικόνα 27. Απεικόνιση του video streaming. Στην αριστερή εικόνα παρουσιάζεται μία ιδεατή κατάσταση όπου υπάρχει συνεχής αναπαραγωγή του video στο δέκτη στο ρυθμό που παράγεται το video από την πλευρά του πομπού. Στην πραγματικότητα όμως η κατάσταση που λαμβάνει χώρα είναι αυτή της δεξιάς εικόνας όπου υπάρχουν μεταβλητές καθυστερήσεις στο δίκτυο (delay jitter). Αντιμετωπίζεται με την εισαγωγή playout buffer στην πλευρά του δέκτη (Kurose, J.F : Ross, K.W., 2012).

3.3 Μεταφορά Δεδομένων Μέσω του Δικτύου Κινητής Τηλεφωνίας

Αν και η πρώτη επίδειξη χρήσης κινητών τηλεφώνων χειρός πραγματοποιήθηκε το 1973 από τον John Mitchell και τον Martin Cooper (Brophy, 2012), οι συσκευές εισήχθησαν στην αγορά των ΗΠΑ μετά από δέκα έτη (1984) και δεν χρησιμοποιήθηκαν ευρέως μέχρι και τα μέσα της δεκαετίας του 1990 (Balzano, Garay, & Manning, 1995). Από τότε και μέχρι το 2000, οι συνδρομητές της κινητής τηλεφωνίας έφτασαν παγκοσμίως τα 500 εκατ. (Maier, Blakemore, & Koivisto, 2000) ενώ σήμερα ξεπερνούν τα 7,5 δισεκατ. (Statista, 2018), αποδεικνύοντας ότι πρόκειται για έναν από τους πιο ταχέως αναπτυσσόμενους τομείς.

3.3.1 Συστήματα Κινητής Τηλεφωνίας

Σύμφωνα με την Διεθνή Κοινοπραξία Ηλεκτρολόγων Μηχανικών (International Engineering Consortium): «Τα συστήματα κινητής τηλεφωνίας χρησιμοποιούν ένα μεγάλο αριθμό ασύρματων πομπών χαμηλής ισχύος οι οποίοι δημιουργούν, σε χωρικό επίπεδο, ένα κυψελοειδές δίκτυο⁵⁶. Η δυνατότητα μεταβολής της ισχύος εκπομπής επιτρέπει στις κυψέλες να προσδιορίζουν το μέγεθός τους σύμφωνα με την πυκνότητα των χρηστών και τις απαιτήσεις που υπάρχουν για επικοινωνία εντός μια δεδομένης περιοχής. Καθώς οι χρήστες μετακινούνται από την περιοχή κάλυψης μίας κυψέλης σε μία άλλη, πραγματοποιείται και αντίστοιχη μεταβίβαση της κλήσης (handoff) μεταξύ των δύο κυψελών. Η κυψελοειδής μορφή του δικτύου δίνει τη δυνατότητα επαναχρησιμοποίησης των διαθέσιμων συχνοτήτων (καναλιών), από κυψέλες που απέχουν ορισμένη απόσταση μεταξύ τους»⁵⁷.

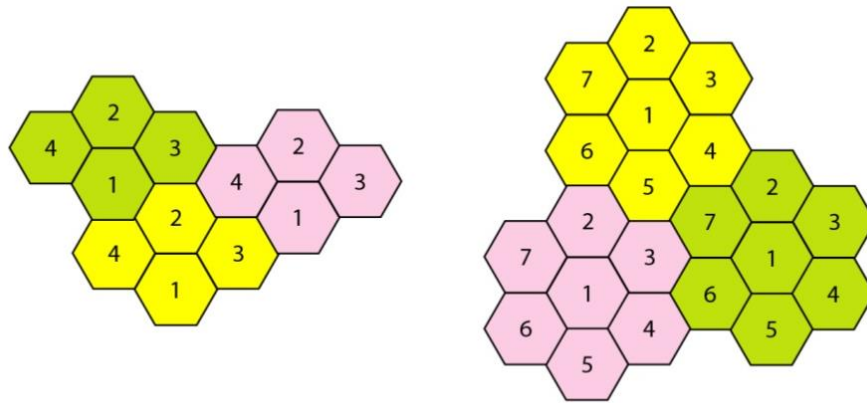
Αν και η κινητή επικοινωνία υπάρχει για περισσότερα από 100 χρόνια, εν τούτοις, μόλις τα τελευταία έτη έχει ωριμάσει σαν τεχνολογία σε τέτοιο σημείο, ώστε να καθίσταται εφικτή αλλά και πρακτική η επικοινωνία σε κάθε σημείο της γης. Σύμφωνα με τους Simon & Alejandro, οι βασικοί σκοποί των συστημάτων κινητής τηλεφωνίας είναι οι παρακάτω (Simon & Alejandro, 2007):

- Κάλυψη (Coverage) : Τα κυψελοειδή συστήματα θα πρέπει να επιδιώκουν την επίτευξη της μεγαλύτερης δυνατής χωρικής κάλυψης, λαμβάνοντας υπόψη τις περιοχές στις οποίες κινούνται οι χρήστες. Αυτό αναφέρεται όχι μόνο σε εξωτερικούς αλλά και σε εσωτερικούς χώρους.
- Ευκαμψία (mobility): Θα πρέπει να είναι δυνατή η χρήση μίας συσκευής ώστε να διανέμει κατάλληλα τους πόρους που διαθέτει με σκοπό την υποστήριξη της διαλειτουργικότητας μεταξύ διαφορετικών προτύπων.
- Χωρητικότητα (Capacity): Αναφέρεται στη δυνατότητα υποστήριξης μεγάλου αριθμού χρηστών με ελαχιστοποίηση των αποκρυστάκων (dropped) η αποκλεισμένων (blocked) κλήσεων. Σε αυτό συνδράμει και η βέλτιστη εκμετάλλευση διαχείρισης συχνοτήτων.
- Παροχή υψηλής ποιότητας υπηρεσιών οι οποίες συνίστανται σε υπηρεσίες φωνής και δεδομένων.
- Επίτευξη των παραπάνω σε κόστος προσιτό για το χρήστη.

⁵⁶ Η κυψέλη αποτελεί τη βασική υπηρεσία

⁵⁷ <https://www.coursehero.com/file/22688480/E-Cellular/#/> (Πρόσβαση:23/1/18)

Σύμφωνα με τους Lempiäinen & Manninen, οι παράγοντες που λαμβάνονται υπόψη κατά το σχεδιασμό ενός δικτύου κινητής τηλεφωνίας, πλέον της προσπάθειας επίτευξης των παραπάνω στόχων, είναι επίσης το κόστος ανάπτυξης του δικτύου, η τοποθεσία εγκατάστασης των μονάδων / σταθμών του δικτύου και οι προοπτικές που υπάρχουν για την περαιτέρω ανάπτυξη του δικτύου (Lempiäinen & Manninen, 2002).



Εικόνα 28. Εφαρμογή της τεχνικής επαναχρησιμοποίησης συχνοτήτων σε κυψελοειδή δίκτυα επικοινωνιών⁵⁸ (στο αριστερό σχήμα $N=4$ και στο δεξιό $N=7$)

Η μεταφορά δεδομένων μέσω του δικτύου κινητής τηλεφωνίας, επηρεάζεται από ένα πλήθος παραγόντων οι σημαντικότεροι από τους οποίους συνίστανται στα παρακάτω:

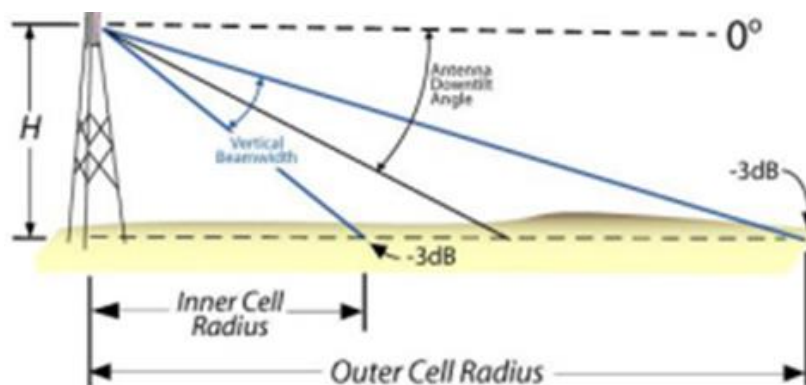
- Στην τεχνολογική γενιά των συστημάτων που προσφέρει ο πάροχος (δίκτυο 2G, 3G και 4G). Αποτελεί το σημαντικότερο παράγοντα και θα γίνει εκτενής αναφορά στις υποενότητες που ακολουθούν.
- Οι μεταβολές στο bandwidth στη διάρκεια του χρόνου είναι ιδιαίτερα σημαντικές στα δίκτυα των παρόχων κινητής τηλεφωνίας, λόγω των φαινομένων που σχετίζονται με την διάδοση των Ηλεκτρομαγνητικών κυμάτων (LOS/Non - LOS, direct signal/ Scatter/ reflection/ defraction)
- Το βασικό πρόβλημα των δικτύων κινητής τηλεφωνίας είναι το φαινόμενο της διάλειψης⁵⁹ (William Stallings, 2007). Με τον συγκεκριμένο όρο περιγράφεται η επίδραση των απότομων μεταβολών στο μέσο διάδοσης των ραδιοκυμάτων. Συγκεκριμένα η διάλειψη μεταβάλλεται με το χρόνο, τη θέση του δέκτη ή την τιμή της συχνότητας (ibid). Μπορεί να οφείλεται στην διάδοση πολλαπλών

⁵⁸ <https://www.slide share.net/ajal4u/frequency-re-use-nb> (Πρόσβαση: 12/1/18)

⁵⁹ Fading

διαδρομών (fast ή multipath fading), στο φαινόμενο της σκίασης που προκαλείται από εμπόδια στην κατεύθυνση διάδοσης (shadowing ή slow fading) ή στην απώλεια διάδοσης (Simon & Alejandro, 2007). Ακόμη και εάν η ισχύς του λαμβανόμενου σήματος βρίσκεται σε αναμενόμενα επίπεδα, τα διαφορετικά φαινόμενα διάδοσης ραδιοκυμάτων μπορεί να επηρεάσουν το λαμβανόμενο σήμα (William Stallings, 2007).

- Η ασύρματη μετάδοση video υπόκειται στο φαινόμενο της διακοπής λόγω αλλαγής κυψέλης. Στη συγκεκριμένη κατηγορία εντάσσεται και η δυνατότητα του παρόχου για ομαλό handover.



Εικόνα 29. Καθ' ύψος κάλυψη ενός Base Station

- Δυσχέρειες πιθανώς να προκύπτουν και από την διάρκεια ζωής της μπαταρίας των ασύρματων συσκευών που εμπλέκονται στη μετάδοση
- Η καθ' ύψος κατευθυντικότητα του λοβού των κεραιών των σταθμών βάσης της κινητής τηλεφωνίας. Οι κεραιές παρέχουν ομοιόμορφη αζιμουθιακή κάλυψη ενώ παρουσιάζουν κατευθυντικότητα ως προς το κατακόρυφο επίπεδο. Η επιδίωξη είναι ο κεντρικός λοβός της δέσμης ακτινοβολίας (HPBW) της κεραίας, ως προς το κατακόρυφο επίπεδο, να μην κατευθύνεται παράλληλα με το έδαφος αλλά να διαθέτει μία κλίση προς αυτό, καθόσον εκεί βρίσκονται οι χρήστες της κινητής τηλεφωνίας. Αυτή η τεχνική αυξάνει την ισχύ του σήματος στο προβαλλόμενο εδαφικό ίχνος (footprint), ενώ ταυτόχρονα ορίζει και τα εξωτερικά όρια επικοινωνιακής κάλυψης της συγκεκριμένης κυψέλης, όπως φαίνεται στην εικόνα 29.
- Στο φαινόμενο Doppler το οποίο οδηγεί σε μετατόπιση της συχνότητας (αύξηση ή μείωση), όταν υπάρχει σχετική ταχύτητα μεταξύ της κινητής συσκευής και του σταθμού βάσης (Vali, Berber, & Nguang, 2009). Το φαινόμενο της

πολυδιοδευσης (multipath) σε συνδυασμό με τη μετατόπιση Doppler, έχει ως αποτέλεσμα τα κύματα να καταλήγουν στο δέκτη από πολλαπλές κατευθύνσεις όπου κάθε μία αποδίδει διαφορετική μετατόπιση βάσει της γεωμετρίας της στο χώρο σε σχέση με την κινητή συσκευή. Το φαινόμενο αυτό ονομάζεται διάχυση Doppler (Doppler Spread) και το επαγόμενο εύρος φάσματος Doppler Bandwidth. Στο πεδίο του χρόνου, η μετατόπιση Doppler οδηγεί στο χρόνο συνοχής T_c οποίος είναι αντιστρόφως ανάλογο του Doppler Spread f_m και ορίζεται ως το χρονικό διάστημα εντός του οποίου το κανάλι μπορεί να θεωρηθεί ως σταθερό, με $T_c = \frac{9}{16\pi f_m}$ (Aragon-Zavala, 2008). Στην περίπτωση των UAV και των Ε/Π, που εξετάζονται στη συγκεκριμένη μεταπτυχιακή διατριβή, η ταχύτητα πλεύσης κυμαίνεται μεταξύ των τιμών 0 έως 130 Knots (240km/h ή 67 m/s) και επομένως η μέγιστη μετατόπιση Doppler δεν ξεπερνά τα $\pm 450\text{Hz}$ ⁶⁰.

Video Resolution	Μέγεθος Πλαισίου		Μέγεθος Video (25fps) ⁽⁴⁾	
	MP	RGB 8-bit (MByte) ⁽¹⁾	RGB 3X8bit ⁽²⁾	H-264 ⁽³⁾
HD 1920X1080	2.1	6.22	156 MB	7 MB
HD 1280X720	0.9	2.76	69,1 MB	3,11 MB
960X540	0.5	1.56	32,4 MB	1,46 MB
480X270	0,13	0.39	9,72 MB	0,44 MB

ΠΑΡΑΤΗΡΗΣΕΙΣ
(1) Uncompressed
(2) Uncompressed BMP/TIFF
(3) BluRay H.264 (1080p25 = 56 Mbps)
(4) Το μέγεθος του video υπολογίστηκε σε χρονική διάρκεια 1sec οπότε οι τιμές του πίνακα αναφέρονται ταυτόχρονα και στο Bitrate (MB/s)

Πίνακας 3. Μέγεθος πλαισίου και video σε διαφορετικές αναλύσεις αυτού

Τέλος, προκειμένου να υπάρχει μία τάξη σύγκρισης μεγεθών μεταξύ της ταχύτητας των ασύρματων δικτύων κινητής τηλεφωνίας, που θα αναλυθούν στη συνέχεια, και του απαιτούμενου εύρους ζώνης για τη μεταφορά video (upload), παρατίθενται στον πίνακα 3 οι αντίστοιχες τιμές για διαφορετικά resolution.

3.3.2 Συστήματα Δεύτερης Γενιάς (2G)

Αναφορικά με την ανάπτυξη των δικτύων κινητής τηλεφωνίας αυτά εξελίχθηκαν από

⁶⁰ Οι μακροκυψέλες σε σχέση με τις κυψέλες μικρότερου μεγέθους (picocells, femtocells), υποστηρίζουν μεγαλύτερες ταχύτητες κίνησης των χρηστών με τις πρώτες να εξυπηρετούν έως 350Km/h και τις δεύτερες έως 30Km/h (Rysavy Research, 2013).

τα πρώτης γενιάς (1G) την δεκαετία του 1980, τα οποία ήταν αναλογικά, σε αυτά της δεύτερης γενιάς (2G) στις αρχές της δεκαετίας του 1990. Τα τελευταία παρείχαν καλύτερη ποιότητα σήματος, υψηλότερους ρυθμούς μεταφοράς δεδομένων και μεγαλύτερη χωρητικότητα (William Stallings, 2007). Η κυριότερες διαφορές τους εστιάζονται στο γεγονός ότι τα συστήματα της δεύτερης γενιάς ήταν πλέον ψηφιακά, παρείχαν δυνατότητα κρυπτασφάλισης, μετάδοσης μηνυμάτων καθώς επίσης και δυνατότητα εντοπισμού και διόρθωσης σφαλμάτων (ibid). Βασίστηκαν σε συστήματα πολυπλεξίας με κάθε κανάλι να διαμοιράζεται δυναμικά σε ένα αριθμό χρηστών μέσω των τεχνικών Time Division Multiple Access (TDMA) ή Frequency Division Multiple Access (FDMA) (ibid).

Παραδείγματα της συγκεκριμένης τεχνολογίας είναι τα συστήματα Global System for Mobile Communications (GSM), IS - 95 CDMS και IS - 136 TDMS⁶¹. Από αυτά το πλέον διαδεδομένο είναι το GSM το οποίο λειτουργούσε σε δύο μπάντες συχνοτήτων (850/900 MHz, 1800/1900 MHz) και διέθετε εύρος ζώνης 200kHz (Ghosh, Zhang, Andrews, & Muhamed, 2010). Πρακτικά η συγκεκριμένη τεχνολογία προσέδιδε δυνατότητα μεταφοράς δεδομένων της τάξης των 9,6Kbps (circuit-switched data).

Στα μέσα της δεκαετίας του 1990, το Ευρωπαϊκό Ινστιτούτο Προτύπων Τηλεπικοινωνιών (European Telecommunications Standards Institute -ETSI), παρουσίασε το GSM Packet Radio Systems (GPRS), το οποίο αποτέλεσε σημαντική αναβάθμιση του GSM σε ότι αφορά στους ρυθμούς μετάδοσης δεδομένων. Ενώ ο μέγιστος ονομαστικός ρυθμός μετάδοσης άγγιζε τα 160kbps ουσιαστικά οι χρήστες είχαν πραγματικό ρυθμό της τάξης των 20 έως 40kbps (ibid). Μία ακόμα αναβάθμιση αποτέλεσε το Enhanced Data Rate for GSM Evolution (EDGE) στο οποίο οι τυπικές ταχύτητες για τους χρήστες κυμαίνονταν μεταξύ 80 και 120kbps (οι ονομαστικές ταχύτητες έφταναν τα 236 kbps για download, ενώ το upload μπορούσε να φτάσει στα 118,4 kbps) (De Vriendt, Lainé, Lerouge, & Xu, 2002).

3.3.3 Συστήματα Τρίτης Γενιάς (3G)

Τα συστήματα 2G στηρίζονταν σε κυκλώματα μεταγωγής (circuit-switched) τα οποία παρείχαν δυνατότητες μικρού ρυθμού μεταφοράς δεδομένων και περιορισμένη

⁶¹ Το IS-95 αναπτύχθηκε στη Β. Αμερική και μέρη της Ασίας ενώ το IS-136 αν και αναπτύχθηκε αρχικά επίσης στη Β. Αμερική, στη συνέχεια αντικαταστάθηκε από το GSM.

χωρητικότητα. Οι προδιαγραφές που τέθηκαν από τη Διεθνή Ένωση Τηλεπικοινωνιών (International Telecommunications Union -ITU), προέβλεπαν τη δυνατότητα υποστήριξης ρυθμού 2Mbps σε σταθερό περιβάλλον εργασίας και εντός κτιρίων, 384kbps για χρήση από πεζούς σε αστικό περιβάλλον και 144kbps σε ανοιχτές περιοχές οδικού περιβάλλοντος (ibid). Οι εν λόγω ρυθμοί θα προσέδιδαν καλύτερη ποιότητα υπηρεσιών αλλά και δυνατότητα υποστήριξης εφαρμογών πολυμέσων, συμπεριλαμβανομένου του streaming (video call, audio-video streaming, video on demand, mobile TV), καθώς επίσης και μεταβλητών ρυθμών δεδομένων ανά χρήστη (William Stallings, 2007).

Το πρότυπο Universal Mobile Telephone Service (UMTS) αναπτύχθηκε από τον ETSI ως 3G σύστημα βασιζόμενο στην εξέλιξη του GSM⁶². Η βασική διαφοροποίηση αφορούσε στην σχεδίαση του W-CDMA (Wide-band CDMA) το οποίο περιλάμβανε δύο διαμορφώσεις λειτουργίας Frequency Division Duplex (FDD) και Time Division Duplex (TDD) (ETSI 2004, 15) ενώ υποστήριζε τους ρυθμούς μεταφοράς δεδομένων που αναφέρθηκαν στην προηγούμενη παράγραφο.

Η επόμενη αναβάθμιση του UMTS-WCDMA αναφέρεται ως High-Speed Packet Access, (HSPA) και αποτελεί συνδυασμό του High-Speed Downlink Packet Access (HSDPA)⁶³ και του High-Speed Uplink Packet Access (HSUPA)⁶⁴. Το HSDPA προήλθε από την απαίτηση για υψηλότερη διεκπεραιωτικότητα (throughput) κατά το download. Η θεωρητική ταχύτητα άγγιζε τα 14.4Mbps ενώ η πρακτική κυμαίνονταν από 0,5 έως 2Mbps (Ghosh et al., 2010). Αντίστοιχα το HSUPA επιτύγχανε θεωρητικό upload 5,8Mbps ενώ στην πράξη οι χρήστες απολάμβαναν ρυθμούς της τάξης των 0,5 έως 1Mbps (ibid)

Το 2007 παρουσιάστηκε το HSPA+⁶⁵ το οποίο μέσω εισαγωγής της διαμόρφωσης

⁶² Λόγω του ότι η επιτροπή 3GPP (3rd Generation Partnership Project) ολοκλήρωσε και εξέδωσε το πρότυπο για τη λειτουργία του UMTS το 1999, το τελευταίο φέρει συνήθως την ονομασία Release 99.

⁶³ Παρουσιάστηκε το 2002 ως Release 5. Μεταξύ Release 99 και Release 5 μεσολάβησε η Release 4 το σημαντικότερο στοιχείο της οποίας ήταν η εφαρμογή chip rate του LCR στα 1,28Mcps. Το πλεονέκτημά του εγγυται στο γεγονός ότι υποστηρίζεται από μονές (unpaired) μπάντες συχνοτήτων των 1,6MHz γεγονός που το καθιστά πιο ευπροσάρμοστο από το FDD και το 3.84Mcps TDD τα οποία απαιτούν ελάχιστο εύρο ζώνης 5MHz (ETSI Mobile Competence Centre, 8 - 10 December, 2004).

⁶⁴ Παρουσιάστηκε το 2002 ως Release 6.

⁶⁵ Release 7

64QAM καθώς της χρήσης δύο κεραιών στους σταθμούς βάσης αλλά και στα τερματικά κινητά τηλέφωνα⁶⁶, επιτύγχανε θεωρητική ταχύτητα της τάξης των 28Mbps (download) και 11 Mbps (upload). Στη συγκεκριμένη έκδοση δεν ήταν δυνατή η ταυτόχρονη χρήση της συγκεκριμένης διαμόρφωσης αλλά και του MIMO. Αυτό επετεύχθη με την 8η έκδοση οπότε η θεωρητική ταχύτητα άγγιξε τα 42Mbps (ibid).

3.3.4 LTE και Συστήματα Τέταρτης Γενιάς (4G)

Πολλές φορές το Long Term Evolution (LTE) αναφέρεται ως 4G σύστημα, γεγονός το οποίο τεχνικά δεν ισχύει βάσει των προτύπων που έθεσε για το τελευταίο η ITU. Η μόνη δικαιολογία για αυτό το χαρακτηρισμό είναι η μεγάλη απόσταση που χωρίζει το LTE από τα υπόλοιπα συστήματα της 3 και 3+ γενιάς⁶⁷ (ibid). Οι στόχοι που τέθηκαν για το LTE⁶⁸ είναι 300 Mbps downlink και 71Mbps uplink (Rysavy Research, 2013).

Ο ορισμός της ITU για τα 4G συστήματα φέρει την ονομασία IMT-Advanced ενώ τα εν λόγω συστήματα θα είναι βασισμένα στο πρωτόκολλο IP (IPv6). Το 2009 συστάθηκε από την 3GPP μία επιτροπή που θα μελετούσε τις απαιτήσεις του LTE-Advanced⁶⁹ το οποίο στη συνέχεια προτάθηκε προς την ITU ως το IMT – Advanced πρότυπο (Ghosh 2010, 39-40). Σύμφωνα με αυτό τα 4G συστήματα θα έχουν την δυνατότητα:

- Παροχής ρυθμών δεδομένων 100 Mbps, ενώ ο χρήστης κινείται σε υψηλές ταχύτητες σε σχέση με το σταθμό βάσης (πχ τρένα), και 1 Gbps downlink ή 0,5 Gbps uplink, όταν η θέση του είναι σχετικά σταθερή (ITU-R, 2008).
- Διακίνησης video υψηλής ανάλυσης, τηλεδιάσκεψης μέσω βιντεοκλήσης και γενικά εφαρμογές που απαιτούν μεγαλύτερες ταχύτητες μεταφοράς δεδομένων (Hossain, Rasti, Tabassum, & Abdelnasser, 2014).

3.3.5 Το Internet μέσω Κινητής Τηλεφωνίας στην Ελλάδα

Στην Ελλάδα το mobile internet καθίσταται διαθέσιμο από 3 παρόχους, την Cosmote, τη Vodafone Greece και τη Wind Hellas. Αναλόγως της περιοχής, το internet δύναται να υποστηρίζεται μέσω δικτύου 3G ή 4G. Στην εικόνα 30 φαίνεται η κάλυψη του δικτύου 3G και 4G, των εταιριών Cosmote και Vodafone για την περιοχή της Θεσσαλίας (η αντίστοιχη ιστοσελίδα της Wind δε διέθετε χάρτη περιοχής αλλά παρείχε χάρτη

⁶⁶ MIMO - Multiple Input Multiple Output

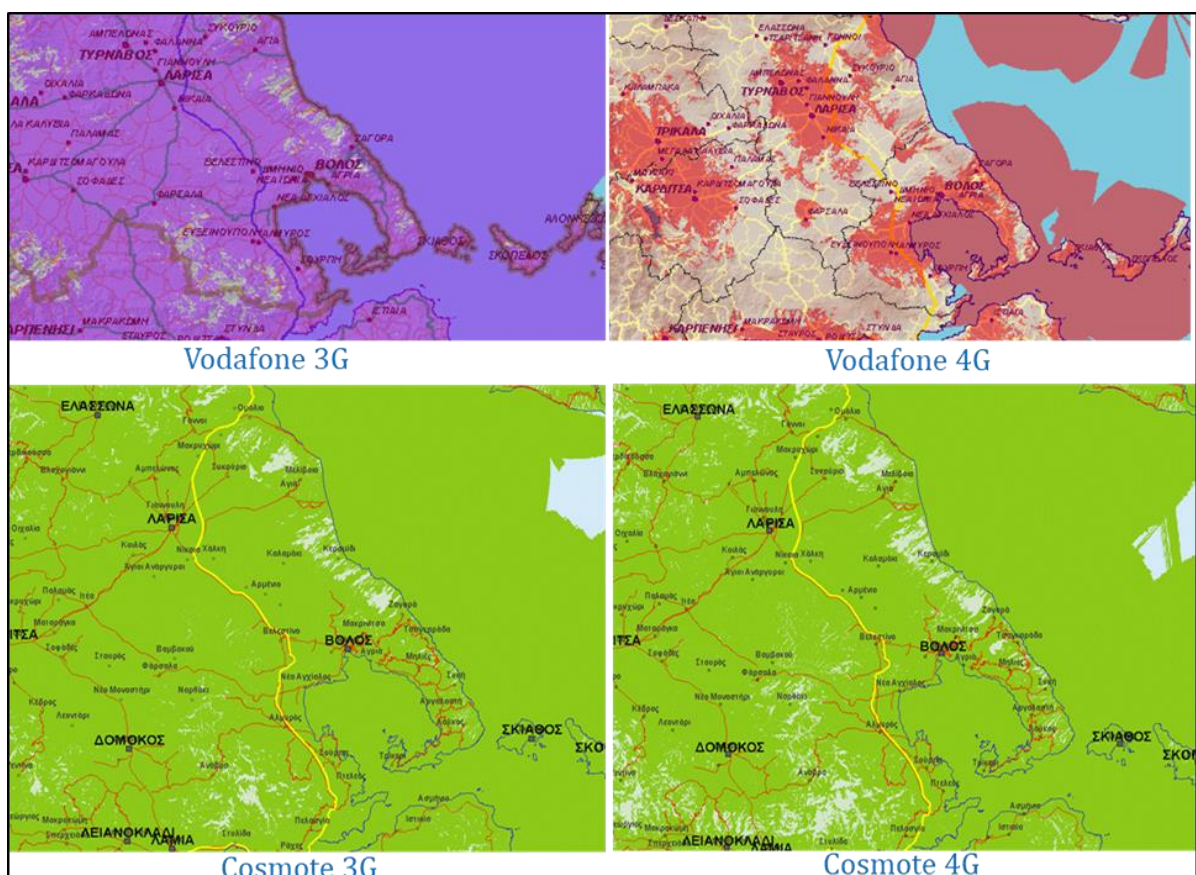
⁶⁷ Αυτός είναι και ο λόγος που συμπεριελήφθη στη συγκεκριμένη υποενότητα

⁶⁸ Release 8 LTE

⁶⁹ Release 10

κάλυψης για το σύνολο του Ελλαδικού χώρου).

Επίσης, λόγω της διεξαγωγής των δοκιμών για video streaming χρησιμοποιώντας το δίκτυο της Cosmote, το επίσημο site της εταιρίας⁷⁰ σημειώνει αναφορικά με τις επιτυγχανόμενες ταχύτητες: «Το δίκτυο COSMOTE 4G τεχνολογίας LTE (Long Term Evolution) προσφέρει [...] περισσότερο από 4 φορές υψηλότερη μέση ταχύτητα από το δίκτυο 3G. Η μέση ταχύτητα πρόσβασης μέσω του δικτύου 3G της COSMOTE ξεπερνάει τα 9 Mbps. Τα στοιχεία αυτά πιστοποιούνται από την τελευταία έρευνα της ανεξάρτητης εταιρείας P3, η οποία διενεργήθηκε το Μάρτιο του 2015».



Εικόνα 30. Χάρτες κάλυψης παρόχων κινητής τηλεφωνίας της περιοχής της Θεσσαλίας⁷¹ (Cosmote, 2018; Vodafone, 2018)

Μελετώντας τη στόχευση των εταιριών κινητής τηλεφωνίας (κάλυψη, χωρητικότητα, εύρος ζώνης) υπό το πρίσμα της δυνατότητας υλοποίησης αυτών των στόχων σε

⁷⁰ <https://www.cosmote.gr/cs/cosmote/gr/4gplus.html> (Πρόσβαση: 28/1/18)

⁷¹ Σύμφωνα με το <http://mobilemaps.freesite.host/el/>, (Πρόσβαση:28/1/18) η Vodafone δε διαχωρίζει το EDGE από το 3G. Όπως αναφέρθηκε στην υποενότητα 2.10.2 το EDGE αποτελεί την τελευταία αναβάθμιση του συστήματος GSM (2G), οπότε δεν θα πρέπει να θεωρείται ως 3G σύστημα.

πρακτικό επίπεδο⁷², αλλά και των ιδιαίτερων προκλήσεων που τίθενται αναφορικά με την καθ' ύψος κάλυψη που δύναται να παρέχει το υφιστάμενο κυψελοειδές δίκτυο, καθιστούν ενδιαφέρουσα την έρευνα της συγκεκριμένης μεταπτυχιακής διατριβής στο τμήμα που αφορά στο live streaming, από μία εναέρια πλατφόρμα προς ένα επίγειο σταθμό, κάνοντας χρήση του δικτύου κινητής τηλεφωνίας.

⁷² Η πρακτική υλοποίηση εξαρτάται από τα φαινόμενα που επηρεάζουν την ασύρματη διάδοση (διάλειαση, κύριοι - πλευρικοί λοβοί εκπομπής κεραιών) καθώς επίσης και οικονομικούς παράγοντες οι οποίοι συνίστανται στις ισοροπίες που πρέπει να τηρηθούν μεταξύ μιας επένδυσης και του αναμενόμενου κέρδους από αυτή. Για μια εταιρίας κινητής τηλεφωνίας η επένδυση μπορεί να αφορά στην αναβάθμιση ενός δικτύου από 3G σε 4G, στην τοποθέτηση περισσότερων κεραιών για κάλυψη μεγαλύτερης περιοχής ή εξυπηρέτηση περισσότερων χρηστών κλπ.

Κεφάλαιο 4 Βάσεις Δεδομένων

4.1 Διαδικτυακές Βάσεις Δεδομένων Οχημάτων

Κατά την προσέγγιση του θέματος και αφού μελετήθηκε η σχετική βιβλιογραφία, διαπιστώθηκε η απαίτηση για τη χρησιμοποίηση μίας βάσης δεδομένων με θετικές και αρνητικές εικόνες οι οποίες θα ανταποκρίνονταν στη στόχευση της εργασίας. Από έρευνα που πραγματοποιήθηκε στο διαδίκτυο για βάσεις δεδομένων με αντικείμενο αυτό της ανίχνευσης, διαπιστώθηκε ότι αυτές εντάσσονται σε τρεις κύριες κατηγορίες επιδιώκοντας: (1) ανίχνευση προσώπου, (2) ανίχνευση πεζών και (3) ανίχνευση οχημάτων. Εστιάζοντας στην τελευταία κατηγορία, που αποτελεί και ένα από τα ερευνητικά αντικείμενα της συγκεκριμένης μεταπτυχιακής διατριβής, ακολουθεί συνοπτική παρουσίαση των αντίστοιχων διαδικτυακών βάσεων που εντοπίστηκαν.:

4.1.1 Βάση Δεδομένων του California Institute of Technology

Περιλαμβάνει 526 εικόνες οχημάτων, σε μορφή jpeg. Οι φωτογραφίες απεικονίζουν το πίσω μέρος των οχημάτων. Η ανάλυσή τους είναι 320X240 pixel και το δείγμα λήφθηκε στους αυκινητοδρόμους της νότιας Καλιφόρνια (Oliveira & Santos, 2008).



Εικόνα 31. Βάση δεδομένων του California Institute of Technology

4.1.2 Βάση Δεδομένων CBCL StreetScenes Challenge Framework

Η εν λόγω βάση δεδομένων αποτελεί μία συλλογή εικόνων (3460 θετικών και 87 αρνητικών), επισημάνσεων (annotations), λογισμικού (κώδικα) και μέτρων απόδοσης προκειμένου να χρησιμοποιηθούν στην έρευνα της μηχανικής εκμάθησης. Οι εικόνες

λήφθηκαν στην περιοχή της Βοστώνης και κάθε εικόνα έχει επισημανθεί με ορθογώνια παραλληλόγραμμα που περικλείουν το αντικείμενο ενδιαφέροντος (οχήματα, πεζοί, ποδήλατα, δέντρα, ουρανός, δρόμος και πεζοδρόμιο) (Stan, 2007).



Εικόνα 32. Βάση δεδομένων CBCL StreetScenes Challenge Framework

4.1.3 UIUC Image Database

Η βάση δεδομένων UIUC περιλαμβάνει φωτογραφίες οχημάτων, με τις λήψεις να έχουν πραγματοποιηθεί από το πλευρό αυτών. Οι εικόνες που χρησιμοποιούνται για την εκπαίδευση του αλγορίθμου αποτελούνται από 550 θετικές φωτογραφίες και 500 αρνητικές, ανάλυσης 100X40pixels, σε png μορφή. Επίσης περιλαμβάνει και ένα δείγμα 170 φωτογραφιών για τη χρησιμοποίησή του ως test set (Agarwal, Awan, & Roth, 2004).



Εικόνα 33. UIUC Image Database

4.1.4 LISA – Q Front FOV Database

Η συγκεκριμένη βάση δεδομένων περιλαμβάνει 3 video. Το δύο από αυτά αποτελούνται 1600 και 300 πλαίσια αντίστοιχα και λήφθηκαν απογευματινές ώρες με ηλιοφάνεια. Το τρίτο video περιέχει 300 πλαίσια που κινηματογραφήθηκαν πρωινές ώρες κάτω από συνθήκες νέφωσης (Sivaraman & Trivedi, 2010).



Εικόνα 34. LISA – Q Front FOV Database

4.1.5 Βάση Δεδομένων του University of Central Florida (UCF)

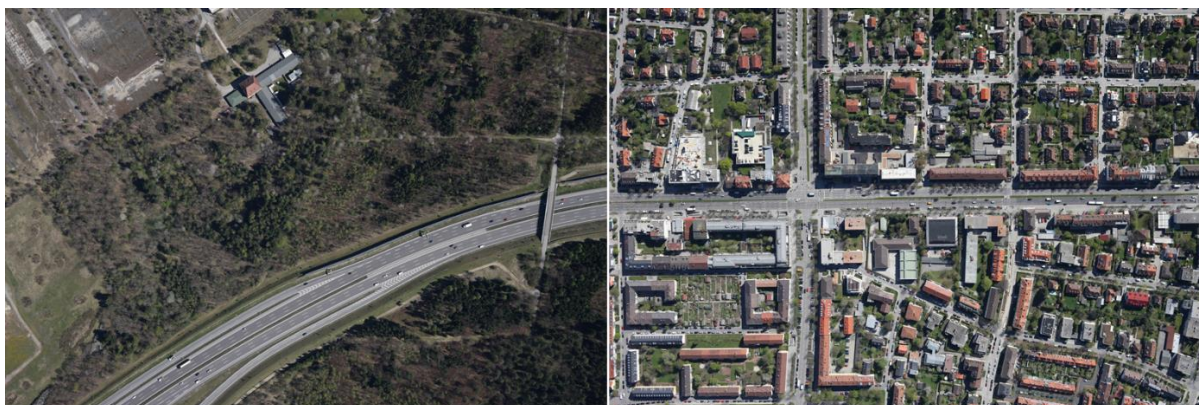
Η συγκεκριμένη βάση δεδομένων περιλαμβάνει video τα οποία λήφθηκαν με τη βοήθεια ενός αερο-οχήματος τύπου zeppelin (ελέγχονταν ασύρματα από το έδαφος). Η συλλογή διαθέτει ποικιλία λήψεων από διαφορετικά ύψη και προοπτική, με βιντεοκάμερα υψηλής ανάλυσης. Αν και η συγκεκριμένη βάση ταιριάζει με το αντικείμενο της μεταπτυχιακής διατριβής, εντούτοις δε χρησιμοποιήθηκε λόγω του περιορισμένου αριθμού οχημάτων αλλά και της κακής ποιότητας της λήψης σε ότι αφορά στη σταθεροποίηση της κάμερας (Soomro, Zamir, & Shah, 2012).



Εικόνα 35. Βάση δεδομένων του University of Central Florida (UCF). Οι φωτογραφίες αποτελούν στιγμιότυπα από τα video.

4.1.6 DLR Munich Vehicle Dataset

Η συγκεκριμένη βάση συγκεντρώθηκε με λήψεις φωτογραφιών, σε μορφή jpeg πάνω από την πόλη του Μονάχου. Χρησιμοποιήθηκε κάμερα Canon EOS 1D MarkIII ανάλυσης 5616X3744 pixel και εστιακού βάθους 50mm. Οι εικόνες λήφθηκαν από ύψος 1000μ (K. Liu & Mattyus, 2015).



Εικόνα 36. DLR Munich Vehicle dataset

4.1.7 Dutch UAS Dataset 001

Η Dutch UAS Dataset 001 περιλαμβάνει πλαίσια από video με ρινόκερους, ζέβρες και

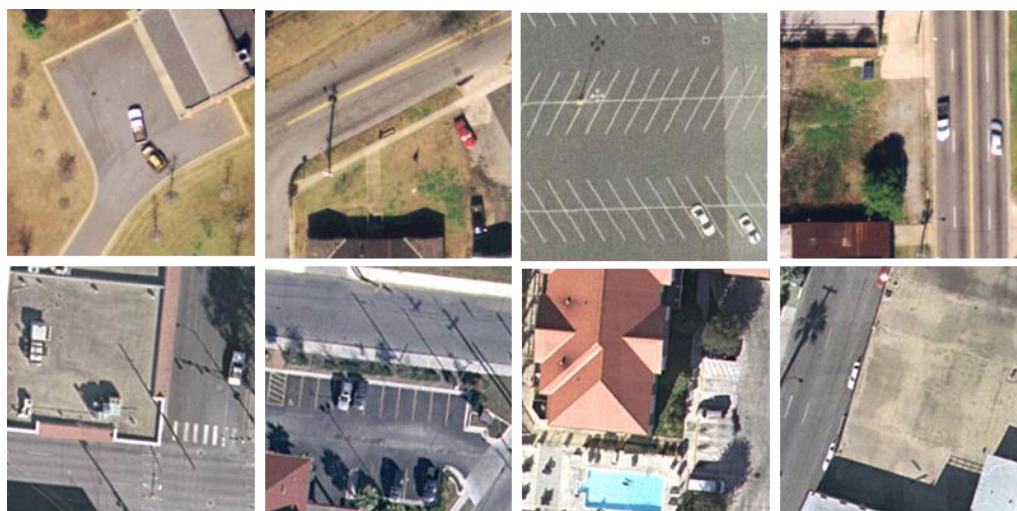
οχήματα τα οποία βρίσκονται σε καταφύγια προστασίας άγριων ζώων στην Αφρική. Οι λήψεις πραγματοποιήθηκαν από UAV σταθερών πτερύγων με κάμερα υψηλής ευκρίνειας. Λόγω του μεγάλου ύψους πτήσης, παρά την ευκρίνεια της κάμερας δε διακρίνονται καθαρά τα αντικείμενα ενδιαφέροντος τα οποία έχουν επισημανθεί χωρίς όμως να γίνεται αναφορά στο είδος τους (Visser et al., 2014).



Εικόνα 37. Dutch UAS Dataset 001

4.1.8 Overhead Imagery Research Data Set (OIRDS)

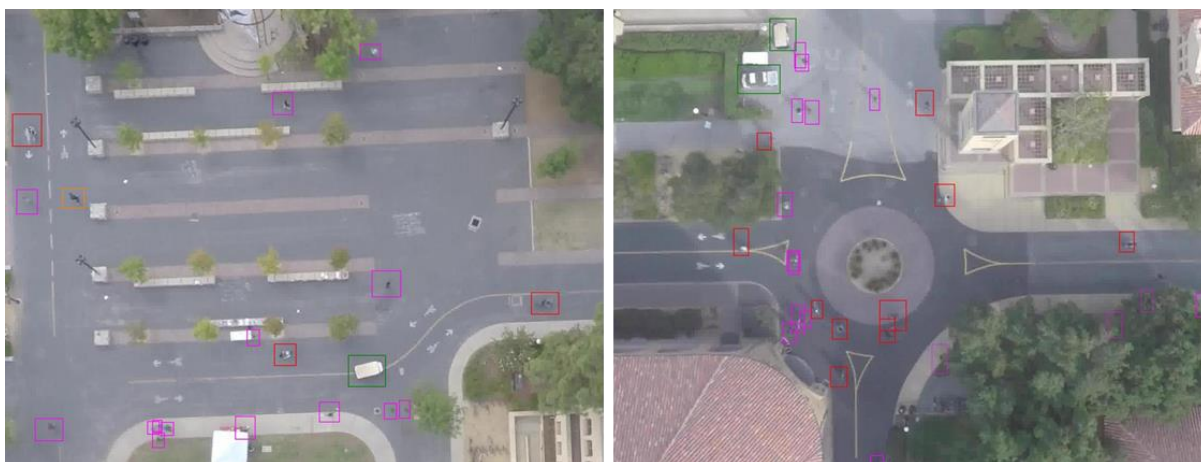
Η Overhead Imagery Research Data Set είναι μία συλλογή από εναέριες φωτογραφίες στις οποίες έχει υλοποιηθεί η εργασία της επισημάνσης. Αντικείμενο των λήψεων ήταν τα επιβατικά οχήματα (Tanner et al., 2009). Η βιβλιοθήκη περιέχει 900 εικόνες με 1800 επισημασμένα οχήματα. Αυτές προέρχονται από το United States Geological Survey (USGS) και το πρόγραμμα VIVID (Video Verification of Identity) της DARPA. Η ανάλυση των φωτογραφιών κυμαίνεται από 256X256 έως 640X480, με επικρατούσα τιμή την 257X257 pixel.



Εικόνα 38. Overhead Imagery Research Data Set (OIRDS)

4.1.9 Stanford Drone Dataset

Η βάση δεδομένων του Stanford αποτελείται από εναέριες λήψεις που πραγματοποιήθηκαν με drone στο χώρο του πανεπιστημίου. Σε αυτές τις εικόνες και τα video έχει πραγματοποιηθεί η εργασία της επισήμανσης για πεζούς, ποδηλάτες, οχήματα, λεωφορεία και οχηματίδια του golf.



Εικόνα 39. Stanford Drone Dataset. Τα οχήματα έχουν επισημανθεί με ορθογώνιο παραλληλόγραμμο πράσινου χρώματος (http://cvgl.stanford.edu/projects/uav_data)

4.1.10 INRIA Aerial Image Labeling

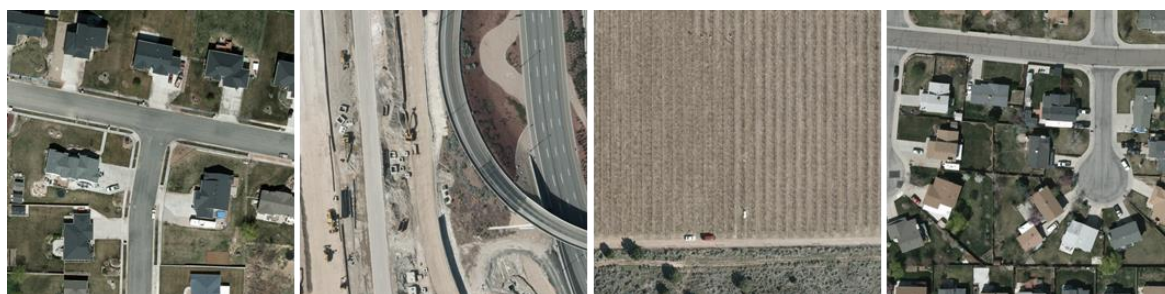
Η συγκεκριμένη βάση δεδομένων αποτελείται εναέριες ορθοαναγόμενες εικόνες (orthorectified), με χωρική ανάλυση 0,3m. Από το γεγονός ότι η εργασία της επισήμανσης για τις υπόψη εικόνες πραγματοποιήθηκε για κτίρια, συνάγεται και ο σκοπός δημιουργίας της (Maggiori, Tarabalka, Charpiat, & Alliez, 2017).



Εικόνα 40. INRIA Aerial Image Labeling.

4.1.11 Vehicle Detection in Aerial Imagery (VEDAI)

Η συγκεκριμένη βάση περιλαμβάνει εικόνες από εναέρια λήψη και σχεδιάστηκε με σκοπό τη χρησιμοποίησή της σε εφαρμογές μηχανικής όρασης. Δημιουργήθηκε για την ανίχνευση οχημάτων, το μέγεθος των οποίων καλύπτει πολύ μικρό αριθμό από pixel (Razakarivony & Jurie, 2016). Οι 1210 εικόνες περιλαμβάνουν ποικιλία περιοχών (δάση, αστικό περιβάλλον, αγροτικές περιοχές κλπ), διατίθενται σε δύο διαφορετικές αναλύσεις 512x512 και 1024x1024, είναι σε διαμόρφωση jpeg, ενώ τα οχήματα βρίσκονται σε τυχαίες διευθύνσεις.



Εικόνα 41. VEDAI Database

4.2 Μία Νέα Βάση Δεδομένων

Από τις παραπάνω δεκατρείς βάσεις δεδομένων, γίνεται άμεσα κατανοητό ότι θα πρέπει να αποκλειστούν οι τέσσερις πρώτες⁷³ καθόσον δε διαθέτουν εναέριες λήψεις. Επίσης, η έλλειψη σταθεροποίησης στην εικόνα, δεν επιτρέπει τη χρήση της βάσης δεδομένων του University of Central Florida (UCF). Ως πηγές, αυτές που προσεγγίζουν περισσότερο το σκοπό της συγκεκριμένης διατριβής είναι οι υπόλοιπες έξι πλην όμως λόγω του μεγάλου γωνιακού εύρους που καλύπτει ο εναέριος φακός αλλά και του ύψους πτήσης από το οποίο έγιναν οι λήψεις, το μέγεθος των οχημάτων καταλαμβάνει μικρό αριθμό από pixel.

Θα μπορούσε να θεωρηθεί ότι οι συγκεκριμένες βάσεις δεδομένων⁷⁴ στοχεύουν στην ανάπτυξη αλγορίθμων οι οποίοι χρησιμοποιούνται στο πρώτο στάδιο της έρευνας για ανίχνευση οχημάτων. Συνήθως, το αρχικό αυτό στάδιο αποσκοπεί στην κάλυψη μιας μεγάλης σε έκταση περιοχής, προκειμένου να υπάρξει μία αρχική ένδειξη για τις περιοχές στις οποίες εντοπίζονται οχήματα. Στη συνέχεια, και βάσει αυτής της αρχικής

⁷³ Του California Institute of Technology, η CBCL StreetScenes Challenge Framework, η UIUC Image Database και η LISA – Q Front FOV Database

⁷⁴ Των υποενοτήτων 4.1.7 έως 4.1.13

πληροφορίας ο φακός δύναται να εστιάσει σε συγκεκριμένες περιοχές ενδιαφέροντος, ώστε να ξεκινήσει η δεύτερη φάση της έρευνας που σχετίζεται με την πρόσκτηση (acquisition) των οχημάτων. Επειδή η μεταπτυχιακή διατριβή εντάσσεται στη δεύτερη φάση της έρευνας, δηλαδή στην πρόσκτηση και ιχνηλάτηση των οχημάτων εντός μια περιοχής περιορισμένου εύρους, απαιτήθηκε η δημιουργία μίας νέας βάσης δεδομένων

4.2.1 Μέσα που Χρησιμοποιήθηκαν για τις Εναέριες Λήψεις

Για τις εναέριες λήψεις χρησιμοποιήθηκε το UAV Phantom 3 Pro της DJI. Το συγκεκριμένο αερόχημα διαθέτει βεληνεκές έως 4800μ (απόσταση από το χειριστή που το ελέγχει). Η ναυτιλία υποβοηθείται από σύστημα GPS/INS, ενώ ο σταθμός εδάφους, μέσω του οποίου ελέγχεται, διαθέτει δύο κεραίες. Η κάμερά του έχει δυνατότητα καταγραφής video ανάλυσης έως 4K. Λεπτομέρειες για τα χαρακτηριστικά και τις δυνατότητες του Phantom 3 Pro, αλλά και της κάμερας που διαθέτει, φαίνονται στους πίνακες 3 και 4 αντίστοιχα⁷⁵.

Δυνατότητες Αερο-οχήματος UAV Phantom 3 Pro	
Βάρος	1280g
Μέγιστη Ταχύτητα Ανόδου	5 m/s
Μέγιστη Ταχύτητα Καθόδου	3 m/s
Μέγιστη Ταχύτητα Πτήσης	16 m/s ή 31 Knots
Διάρκεια Πτήσης	23min
Ακρίβεια Στοιχείων Αιώρησης	V:±0,5m , H: :±1,5m
Συχνότητα Λειτουργίας R/C	2,400 – 2,483 GHz
Ισχύς Εκπομπής (EIRP)	FCC/CE/MIC: 20/16/16 dBm

Πίνακας 4. Προδιαγραφές Phantom 3 Pro



Εικόνα 42. DJI Phantom 3 Pro

⁷⁵ <https://www.dji.com/phantom-3-pro/info> (Πρόσβαση: 7/3/18)

Δυνατότητες Κάμερας UAV Phantom 3 Pro	
Αισθητήρας	1/2.3" CMOS
Ανάλυση	12,4 Megapixel
Φακός	FOV 94° 20 mm (35 mm format equivalent) f/2.8 focus at ∞
Μέγεθος Εικόνας	4000X3000
Διαμορφώσεις Video	UHD: 4096x2160p 24/25, 3840x2160p 24/25/30 FHD: 1920x1080p 24/25/30/48/50/60 HD: 1280x720p 24/25/30/48/50/60
Συμπίεση Video	MP4, MOV (MPEG-4 AVC/H.264)

Πίνακας 5. Προδιαγραφές Κάμερας

4.2.2 Σχεδίαση των Προδιαγραφών της Νέας Βάσης Δεδομένων

Η βιντεοσκόπηση για τη δημιουργία της βάσης δεδομένων, η οποία στη συνέχεια θα ονομάζεται Vertical Recorded Vehicles with UAS Data Base (VERVwUAS DB), πραγματοποιήθηκε στην περιοχή του ν. Μαγνησίας. Οι λήψεις, όπως φαίνεται και στο χάρτη της εικόνας 43, έλαβαν χώρα εντός του αστικού ιστού (περιοχές 2 και 4) καθώς επίσης και στα περίχωρα της πόλης του Βόλου (περιοχές 1 και 3). Πιο συγκεκριμένα, το δείγμα εκπαίδευσης (training set) λήφθηκε από τις περιοχές 1 και 2 ενώ το δείγμα ελέγχου (test set) περιλαμβάνει εικόνες και video και από τις τέσσερις περιοχές. Οι συγκεκριμένες περιοχές επιλέχτηκαν προκειμένου να συμπεριληφθεί στο δείγμα όσο το δυνατόν μεγαλύτερη ποικιλία αναφορικά με τον περιβάλλοντα χώρο. Έτσι, η VERVwUAS DB περιλαμβάνει απεικονίσεις από αστικό, ημιαστικό και αγροτικό περιβάλλον.

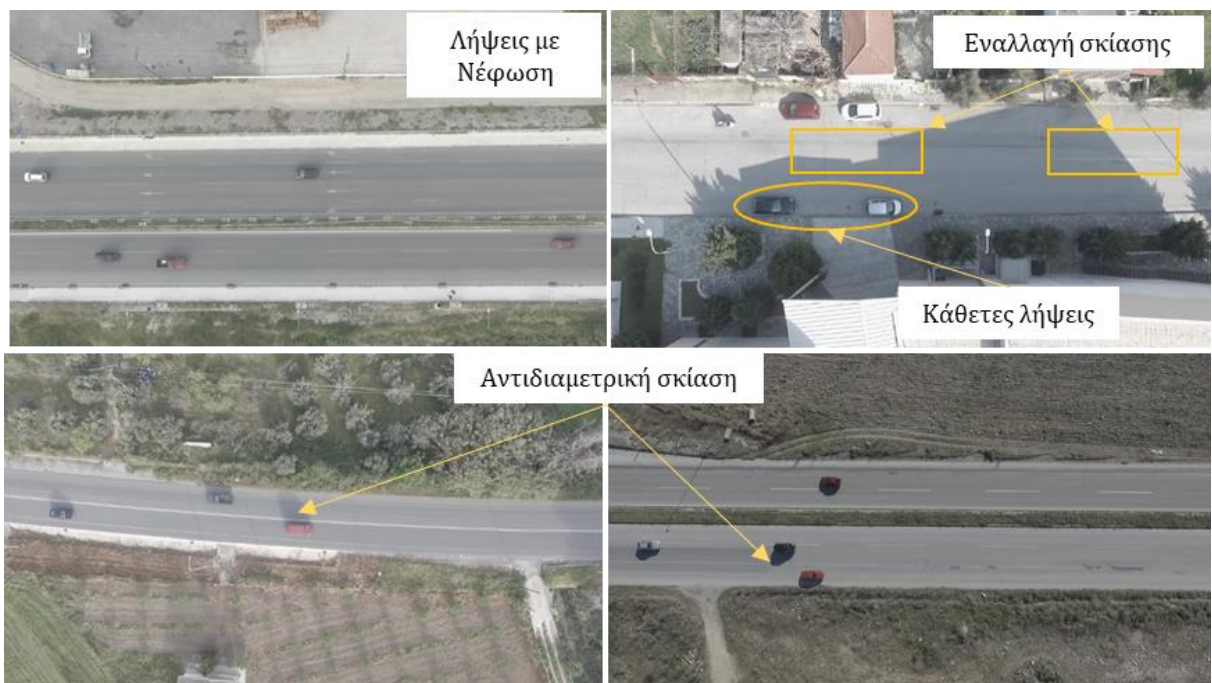


Εικόνα 43. Περιοχές βιντεοσκόπησης για τη δημιουργία της VERVwUAS DB



Εικόνα 44. Στιγμιότυπα από τα video της VERVwUAS DB

Επίσης, επιδιώχθηκε η συλλογή αρνητικών δειγμάτων (background) από διαφορετικούς τύπους εδαφών που συνδυάζουν ποικιλία ως προς την περιοχή (αστική, αγροτική), το χρωματισμό, την υφή και τις συνθήκες φωτισμού. Οι λήψεις που έγιναν στο φυσικό περιβάλλον περιλαμβάνουν δασώδεις, θαμνώδεις, λιβαδικές και οργωμένες/καλλιεργημένες εκτάσεις.



Εικόνα 45. Στιγμιότυπα από τα video της VERVwUAS DB

Εκτός του περιβάλλοντος χώρου, ένας ακόμη παράγοντας που λήφθηκε υπόψη ήταν η σκίαση. Παρατηρώντας τις εικόνες που προέρχονται από τα αντίστοιχα video, φαίνεται ότι υπάρχουν λήψεις στις οποίες η δημιουργούμενη σκιά βρίσκεται, αναλόγως της θέσης του ηλίου κάθε φορά, εκατέρωθεν του διαμήκη άξονα του οχημάτων. Επίσης υπάρχουν και λήψεις που πραγματοποιήθηκαν κάτω από νέφωση ή και κατά το λυκόφως με αποτέλεσμα την απουσία σκιάς.

Η ανάλυση του video που επιλέχτηκε κατά την καταγραφή ήταν 1920X1080. Ο λόγος επιλογής της συγκεκριμένης ποιότητας ευκρίνειας ήταν διττός. Από τη μία, όντας εικόνα υψηλής ευκρίνειας (HD – High Defenition)⁷⁶, διαθέτει ικανοποιητικό αριθμό pixel για την απεικόνιση των οχημάτων⁷⁷, ενώ από την άλλη παρέχει αρκετές επιλογές υποβιβασμού της ανάλυσης, είτε επειδή το εύρος ζώνης του παρόχου κινητής τηλεφωνίας μπορεί να μην επαρκεί είτε επειδή ο χρόνος σάρωσης της εικόνας από τον αλγόριθμο ανίχνευσης υπερβαίνει τον μέγιστο αποδεκτό, για την διατήρηση της ιχνηλάτησης σε σχεδόν πραγματικό χρόνο (near real time).



Εικόνα 46. Στιγμιότυπα από τα video της VERVwUAS DB

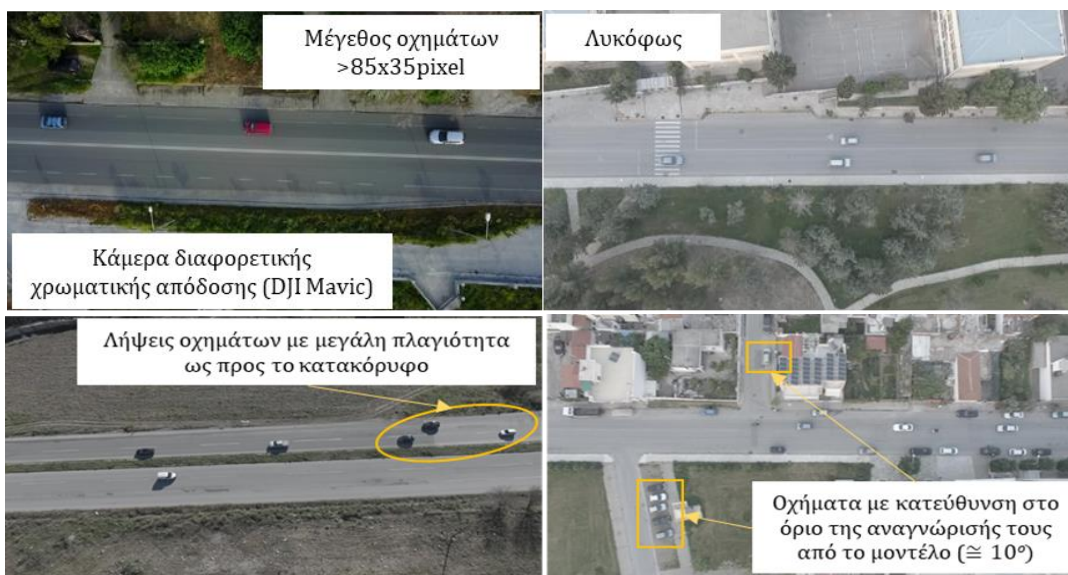
⁷⁶ Η συγκεκριμένη ανάλυση αποτελεί το ενδιάμεσο format μεταξύ των ευραίως χρησιμοποιούμενων 720p, 1080p και 4K.

⁷⁷ Η συγκεκριμένη παράμετρος σχετίζεται και με την απόσταση λήψης η οποία αναλύεται στην επόμενη παράγραφο.

Το ύψος πτήσης καθορίστηκε στα 70m. Οι λόγοι που οδήγησαν στην επιλογή του συγκεκριμένου ύψους σχετίζονται με θέματα καθετότητας της λήψης αλλά και του μεγέθους των οχημάτων. Συγκεκριμένα η περιοχή που καλύπτει το κάθε πλαίσιο έχει διαστάσεις 125μΧ80μ, γεγονός που πρακτικά συνεπάγεται ότι το οπτικό πεδίο (FOV) της κάμερας, κατά το διαμήκη άξονα, είναι περίπου 24 φορές μεγαλύτερο από τη μήκος ενός μέσου οχήματος. Ταυτόχρονα, η παρεχόμενη ευκρίνεια των οχημάτων (85Χ35 pixel για εικόνα διαστάσεων 1920Χ1080), επιτρέπει στο χρήστη και την οπτική αναγνώριση του τύπου των οχημάτων. Επίσης, σε ότι αφορά στην προοπτική θέασης των αυτών⁷⁸ αυτή θα ήταν ανάλογη με το αν το drone βρισκόταν σε ύψος 1000μ με το γωνιακό εύρος λήψης της κάμερας (Angle of View - AoV) να είναι περίπου 8,5° .

Για τους σκοπούς της φάσης δοκιμής του μοντέλου, έχουν πραγματοποιηθεί και λήψεις με τα παρακάτω χαρακτηριστικά :

- Προέρχονται από κάμερα με εμφανή διαφορά στην απόδοση των χρωμάτων σε σχέση με αυτή της κάμερας του DJI Phantom 3 Pro. Οι συγκεκριμένες λήψεις πραγματοποιήθηκαν στις περιοχές 3 και 4 με τη χρησιμοποίηση του UAS Mavic Pro της DJI, σε χαμηλότερα ύψη, ώστε οι διαστάσεις των οχημάτων να καταλαμβάνουν μεγαλύτερο χώρο στην εικόνα.



Εικόνα 47. Στιγμιότυπα από τα video της VERVwUAS DB

⁷⁸ Στα 70μ που ίπτονταν το UAS, το AoV ισούνταν με την τιμή που δίνεται από τις προδιαγραφές της κάμερας, δηλαδή 94°, ενώ το εστιακό μήκος του φακού (lens focal length) ήταν 20mm. Η αντίστοιχη εστιακή τιμή για ύψος πτήσης στα 1000μ, προκειμένου να καλύπτεται η ίδια εδαφική έκταση, ισούται με 300mm.

- Η κατεύθυνση των οχημάτων σε σχέση με τον οριζόντιο άξονα της εικόνας να είναι στο όριο των 10° .
- Η γωνία λήψης δεν είναι κατακόρυφη (vertical) αλλά διαθέτει κλίση η οποία φτάνει μέχρι και τις 25° προκειμένου να δοκιμαστούν τα όρια αναγνώρισης ως προς την κατακόρυφη διάσταση.

Το γεγονός ότι η διεύθυνση του διαμήκη άξονα των οχημάτων ταυτίζεται με τη οριζόντια διάσταση της λήψης ($\pm 10^\circ$), αποτελεί ταυτόχρονα και ένα μειονέκτημα της VERV UAS-DB. Για τους λόγους που θα αναλυθούν ενδελεχώς στη συνέχεια του κειμένου, ο κινητήριος μοχλός για την επιλογή αυτή είναι το μειονέκτημα των μοντέλων μηχανικής όρασης που θα αξιολογηθούν, να επηρεάζεται η ανίχνευση από την κατεύθυνση των αντικειμένων ενδιαφέροντος⁷⁹. Επίσης αναφορικά με το είδος των οχημάτων, λόγω των χρονικών περιορισμών που τίθονταν για την ολοκλήρωση της μεταπτυχιακής διατριβής, η βάση δεδομένων περιλαμβάνει κυρίως ιδιωτικής χρήσης επιβατηγά οχήματα και σε ένα πολύ μικρό ποσοστό λοιπά οχήματα (πχ λεωφορεία, φορτηγά οχήματα κλπ).

⁷⁹ Όπως αναφέρεται στη σχετική βιβλιογραφία δεν είναι "orientation invariant".

Βάσεις Δεδομένων	Θέση Φακού		Εικόνες	Video	An	Περιβάλλον Λήψης			Χρήση Διαφορετικών Φακών	Κίνηση Κάμερας στη Λήψη	Παρ/σεις
	Επ	Εν				ΑΣΤ	Η/ΑΣΤ	ΑΓΡ			
California Institute of Technology			526 εικόνες (320x240 p)	-----						-----	Αυτοκινητόδρομοι στην Καλιφόρνια
CBCL StreetScenes Challenge Framework			3460 θετικές 87 αρνητικές (1280x960p)	-----						-----	Οδοί στη Βοστώνης
UIUC Image Database			550 θετικές 500 αρνητικές (100x40p) 170 ως test set	-----						-----	Περιλαμβάνει και πρόγραμμα αξιολόγησης
LISA - Q Front FOV Database			-----	3 video (13,5MB) (704x480p)							
University of Central Florida (UCF)			-----	4 video (2,01 GB) (720x480p)							
DLR Munich Vehicle Dataset			10 ως train set 10 ως test set (5616x3744p)	-----						-----	
Dutch UAS Dataset 001			577 εικόνες (3000x220p)	-----						-----	Στέπες στην Αφρική
Overhead Imagery Research Data Set (OIRDS)			900 εικόνες (256x256 έως 640x480p)	-----						-----	Προέρχονται από United States Geological Survey και το πρόγραμμα Video Verification of Identity
Stanford Drone Dataset			-----	69 GB (3840x2160p)							Εντός του campus του πανεπιστημίου
INRIA Aerial Image Labeling			180 εικόνες (5000x5000p)	-----						-----	
Vehicle Detection in Aerial Imagery (VEDAI)			1210 εικόνες σε δύο διαστάσεις (512x512p και 1024x1024p)	-----						-----	Προέρχεται από δορυφορικές εικόνες του Utah AGRC
Vertical Recorded Vehicles with UAS Data Base (VERVwUAS DB)			1400 cropped εικόνες οχημάτων 4876 αρνητικές	46 video (22,2 GB) (1960x1080p)							
Υπόμνημα Επ: Επίγεια λήψη Εν:Εναέρια λήψη An : Annotation											
			ΑΣΤ : Αστικό Η/ΑΣΤ : Ημισιαστικό ΑΓΡ : Αγροτικό								

Πίνακας 6. Βασικά Χαρακτηριστικά των Dataset οχημάτων

Κεφάλαιο

5

Μεθοδολογία

5.1 Εκπαίδευση Κατηγοριοποιητών

Αφού καθορίστηκαν οι προδιαγραφές της VERVwUAS DB, σύμφωνα με όσα προαναφέρθηκαν και συλλέχθηκε ο απαιτούμενος αριθμός από video, ξεκίνησε η επεξεργασία τους προκειμένου να υλοποιηθεί η διαδικασία της εκπαίδευσης και ελέγχου. Αναλυτικά τα βήματα που ακολουθήθηκαν παρουσιάζονται στις παρακάτω υποενότητες.

Κατά την ανάλυση θα παρουσιάζονται επιμέρους αποτελέσματα - συμπεράσματα τα οποία όμως εστιάζονται αποκλειστικά στην διαδικασία εκπαίδευσης και στο πώς αυτά καθόρισαν την περαιτέρω στρατηγική της μεθοδολογίας. Τα αποτελέσματα που αφορούν στην απόδοση των κατηγοριοποιητών μηχανικής μάθησης που μελετώνται (HAAR/GAB, LBP/GAB, HOG/GAB και HOG/SVM), καθώς επίσης και οι δυνατότητες των δικτύων 3G/4G να υποστηρίξουν video streaming, που αποτελούν και τα ερευνητικά ερωτήματα της μεταπτυχιακής διατριβής, θα παρουσιαστούν ξεχωριστά στο 6ο κεφάλαιο.

5.1.1 Δημιουργία Δειγμάτων Εκπαίδευσης και Ελέγχου

Τα video που καταγράφηκαν χωρίστηκαν σε δύο μέρη. Το μεγαλύτερο τμήμα χρησιμοποιήθηκε για τη διαδικασία της εκπαίδευσης (train set) ενώ το άλλο για τη δημιουργία του δείγματος ελέγχου (test set). Επίσης, η διαδικασία εκπαίδευσης των HAAR/GAB, LBP/GAB και HOG/GAB, όπως θα αναλυθεί και στη συνέχεια, παρουσίαζε διαφορές σε σχέση με την αντίστοιχη του HOG/SVM.

Το δείγμα εκπαίδευσης για τα μοντέλα HAAR/GAB, LBP/GAB και HOG/GAB αποτελείται από δύο διακριτούς φακέλους, αυτόν με τα θετικά δείγματα και τον αντίστοιχο με τα αρνητικά. Σύμφωνα με την υφιστάμενη βιβλιογραφία ο αριθμός των δεύτερων θα πρέπει να είναι τουλάχιστον διπλάσιος από τον αντίστοιχο των πρώτων. Ειδικότερα:

- Στον πρώτο φάκελο (p:positive) περιλαμβάνονται αποκλειστικά και μόνο εικόνες οχημάτων οι οποίες προήλθαν από περικοπή (crop) οχημάτων από τα αντίστοιχα πλαίσια των video. Για την εργασία της περικοπής χρησιμοποιήθηκε το πρόγραμμα Cascade Trainger GUI στο οποίο εισάγεται η διεύθυνση (path) του video και στη συνέχεια, με τη χρήση δρομέα που διαθέτει, καθίσταται εφικτή η μετάβαση μεταξύ των πλαισίων. Επίσης, το πρόγραμμα δίνει τη δυνατότητα ορισμού της αναλογίας των διαστάσεων του ορθογωνίου που περικόπτεται. Μετά από έρευνα στο διαδίκτυο για την αναλογία μήκους – πλάτους των επιβατηγών οχημάτων, βρέθηκε ότι μια μέση τιμή είναι αυτή του 2,42:1. Βάσει της ανάλυσης του video (1920x1080) και του ύψους πτήσης του drone, οι διαστάσεις των εικόνων που περικόπηκαν κυμαίνονται μεταξύ 62x26, για μικρά οχήματα και 158x66 για φορτηγά οχήματα, με τη μέση τιμή να είναι 85x35pixel). Δείγμα από οχήματα που έχουν περικοπεί από τα πλαίσια των video, φαίνονται στην εικόνα 48.



Εικόνα 48. Θετικά δείγματα οχημάτων 4 διαφορετικών μεγεθών με τη διαδικασία της περικοπής

- Στον δεύτερο φάκελο (n:negative) περιλαμβάνονται φωτογραφίες οι οποίες περιέχουν οτιδήποτε άλλο εκτός από οχήματα. Επειδή ένα από τα σημαντικότερα θέματα κατά την εκπαίδευση των cascades είναι αυτό της αναλογίας των διαστάσεων των εικόνων, και για τα αρνητικά δείγματα τηρήθηκε η ίδια αναλογία (2,42:1) πλην όμως δεν υπήρξε δέσμευση ως προς το μέγεθος της εικόνας που θα τηρούνταν ως δείγμα. Και σε αυτή την περίπτωση το χρησιμοποιούμενο λογισμικό ήταν το Cascade Trainger GUI. Οι εικόνες που

περιέχονται στο συγκεκριμένο φάκελο προέρχονται από τις περιοχές λήψης 1 και 2^{80,81}.



Εικόνα 49. Δείγματα αρνητικών εικόνων

Ενώ για τα μοντέλα HAAR/GAB, LBP/GAB και HOG/GAB πραγματοποιήθηκε περικοπή θετικών εικόνων και διαχωρισμός τους από τις αρνητικές εικόνες, στην περίπτωση του HOG/SVM έλαβε χώρα επισήμανση (annotation) των οχημάτων πάνω στις αντίστοιχες εικόνες, χρησιμοποιώντας το εργαλείο imglab. Στη συνέχεια, οι συντεταγμένες των ορθογωνίων (bounding boxes : x,y,w,h) αποθηκεύτηκαν σε ένα αρχείο xml. Ο αλγόριθμος, κατά τη φάση της εκπαίδευσης, θεωρεί ως θετικό δείγμα τα annotated τμήματα της εικόνας ενώ το υπόλοιπο αυτής αποτελεί το αρνητικό δείγμα (εικόνα 50).



Εικόνα 50. Παράδειγμα annotated εικόνας από το imglab

Θα πρέπει να σημειωθεί ότι αυτή καθ' αυτή η διαδικασία της περικοπής ή του annotation εισάγει ευθύς εξ' αρχής κάποιο σφάλμα καθόσον αντικείμενα, όπως τα οχήματα, τα οποία ουσιαστικά διαθέτουν ακανόνιστο σχήμα, περιγράφονται ουσιαστικά από ένα

⁸⁰ Βλέπε υποενότητα 4.2.2/παράγραφο 1η

⁸¹ Οι συγκεκριμένες αρνητικές εικόνες περικόπηκαν εξ' αρχής για την εκπαίδευση του κατηγοριοποιητή και στη συνέχεια προστέθηκαν και άλλες ως αποτέλεσμα της διαδικασίας hard negative mining.

ορθογώνιο.

Το δείγμα ελέγχου, όπως ήδη αναφέρθηκε, προέρχεται και από τις τέσσερις περιοχές στις οποίες πραγματοποιήθηκαν λήψεις. Η συγκεκριμένη απόφαση στηρίζεται στην πρόθεση μελέτης του πόσο γενικευμένη μπορεί να είναι η κατηγοριοποίηση σε σχέση με το δείγμα εικόνων που χρησιμοποιήθηκε για την εκπαίδευσή του. Το δείγμα ελέγχου δημιουργήθηκε από 54 φωτογραφίες οι οποίες αποτελούν πλαίσια των video που καταγράφηκαν στις περιοχές 1 έως 4⁸². Σε αυτές τις εικόνες περιέχονται συνολικά 252 οχήματα. Το συγκεκριμένο δείγμα υποβλήθηκε στη συνέχεια στη διαδικασία του annotation με τη βοήθεια του εργαλείου ObjectMarker. Τα αποτελέσματα καταχωρήθηκαν αυτόματα σε ένα αρχείο τύπου txt και στη συνέχεια μεταφέρθηκαν σε φύλλο excel.

5.1.2 Δοκιμαστική Φάση Εκπαίδευσης των Αλγορίθμων

Η διαδικασία εκπαίδευσης του αλγορίθμου πραγματοποιήθηκε σε δύο στάδια. Το πρώτο στάδιο αποτέλεσε ουσιαστικά τη φάση της δοκιμής. Η φάση της δοκιμής πραγματοποιήθηκε με ένα μικρό set δειγμάτων τα οποία προέρχονταν από την 1η περιοχή, προκειμένου να κατανοηθεί ο τρόπος με τον οποίο επηρεάζουν οι παράμετροι της εκπαίδευσης το τελικό προϊόν. Λεπτομέρειες του δείγματος εκπαίδευσης που χρησιμοποιήθηκε σε αυτή τη φάση φαίνονται στον πίνακα 6.

FEATURE EXTRACTION	LEARNING ALGORITHM	ΑΡΙΘΜΟΣ ΘΕΤΙΚΩΝ ΔΕΙΓΜΑΤΩΝ	ΑΡΙΘΜΟΣ ΑΡΝΗΤΙΚΩΝ ΔΕΙΓΜΑΤΩΝ	ΑΡΙΘΜΟΣ ANNOTATED ΕΙΚΟΝΩΝ	ΣΥΝΟΛΟ ANNOTATED ΟΧΗΜΑΤΩΝ
HAAR	Gentle Ada Boost (GAP)	348	922	-----	-----
HOG	-//-	-//-	-//-	-----	-----
LBP	-//-	-//-	859 ⁸³	-----	-----
HOG	SVM	-----	-----	12	40

Πίνακας 7. Στοιχεία του δείγματος εκπαίδευσης στη φάση της δοκιμής

Για την εξαγωγή των χαρακτηριστικών σημείων και την εκπαίδευση του κατηγοριοποιητή, χρησιμοποιήθηκαν οι παράμετροι του πίνακα 7. Ο λόγος που τοποθετήθηκε ποσοστό λήψης θετικών εικόνων οφείλεται στο γεγονός ότι σε περιπτώσεις που χρησιμοποιούσαν το 100% των θετικών δειγμάτων τερματίζονταν η εκπαίδευση σε κάποιο ενδιάμεσο στάδιο εμφανίζοντας το μήνυμα «OpenCV Error: Bad argument (Can not get new positive sample. The most possible reason is insufficient

⁸² Για το δείγμα ελέγχου δεν χρησιμοποιήθηκαν video που αποδόθηκαν για τη διαδικασία της εκπαίδευσης των περιοχών 1 και 2.

⁸³ Από τις 922 εικόνες τελικά διατηρήθηκαν 859.

count of samples in given vec-file.)». Μετά από σύντομη έρευνα στο διαδίκτυο για την αντιμετώπιση του συγκεκριμένου προβλήματος, προτεινόταν είτε η μείωση της τιμής του Min Hit Rate είτε η λήψη ενός ποσοστού από το σύνολο των θετικών δειγμάτων. Όπως φαίνεται στον πίνακα, προτιμήθηκε η δεύτερη επιλογή.

CASCADE	ΣΤΑΔΙΑ	M	Π	ΠΟΣΟΣΤΟ ΘΕΤΙΚΩΝ ΕΙΚΟΝΩΝ	ΑΡΙΘΜΟΣ ΑΡΝΗΤΙΚΩΝ ΕΙΚΟΝΩΝ	HAAR MODE	MIN HIT RATE	MAX FAR	WTR	MWC
HAAR_GAP	20	58	24	90% (313)	922	BASIC	0.995	0,5	0,95	100
HOG_GAP	20	-/-	-/-	-/-	-/-	---	-/-	-/-	-/-	-/-
HAAR_GAP	20	-/-	-/-	75% (261)	-/-	---	-/-	0,3	-/-	-/-
LBP_GAP	15	-/-	-/-	100%(348)	859	---	-/-	0,5	-/-	-/-
LBP_GAP	20	-/-	-/-	90% (313)	-/-	---	-/-	-/-	-/-	-/-
LBP_GAP	25	-/-	-/-	75%(261)	-/-	---	-/-	-/-	-/-	-/-
M : Μήκος (W-Width)		WTR : Weight Trim Rate			58:24 = 2,416					
Π : Πλάτος (H – Height)		MWC : Max Weak Count			Με κόκκινο περίγραμμα οι επιπλέον δοκιμές του ίδιου cascade αλλάζοντας παραμέτρους					
FAR : False Alarm Rate		GAP : Gentle Ada Boost								

Πίνακας 8. Παράμετροι εκπαίδευσης των Cascades κατά τη φάση των δοκιμών

Η εφαρμογή των παραπάνω παραμέτρων για την εκπαίδευση των κατηγοριοποιητών ακολουθίας δημιούργησε για κάθε μία περίπτωση το αντίστοιχο cascade.xml αρχείο. Στον πίνακα 8 παρουσιάζονται τα αποτελέσματα και οι παρατηρήσεις επί της διαδικασίας της εκπαίδευσης.

Για την εκπαίδευση σε HOG_SVM δεν ακολουθήθηκε η χρήση των βιβλιοθηκών του OpenCV, αλλά αυτών της Dlib. Σύμφωνα με την οικεία ιστοσελίδα⁸⁴, η «Dlib αποτελεί μία σύγχρονη εργαλειοθήκη γραμμένη σε C++ η οποία περιλαμβάνει αλγόριθμους μηχανικής μάθησης». Η συγκεκριμένη βιβλιοθήκη παρέχει το εργαλείο imglab με το οποίο πραγματοποιήθηκε και η εργασία του annotation σε 12 εικόνες της περιοχής 1, επισημαίνοντας τη θέση 40 συνολικά οχημάτων που εμπεριέχονταν σε αυτές. Με την ολοκλήρωση του annotation δημιουργήθηκε το mydataset.xml αρχείο στο οποίο τοποθετήθηκαν οι συντεταγμένες (x, y, width, height) από τα οχήματα. Τροφοδοτώντας το πρόγραμμα εκπαίδευσης με το συγκεκριμένο αρχείο δημιουργήθηκε το detector.svm αρχείο το οποίο χρησιμοποιείται στο πρόγραμμα ανίχνευσης. Ο συνολικός χρόνος εκπαίδευσης διήρκησε 5m37s⁸⁵.

⁸⁴ <http://dlib.net/>

⁸⁵ Η εκπαίδευση του HOG/SVM έλαβε χώρα σε Laptop Lenovo T510 Intel® Core (TM) i5 CPU M520 @ 2.4 GHz, μνήμη RAM 4 GB, 64-bit, με λειτουργικό Windows 7 Pro και onboard κάρτα γραφικών Intel(R) HD Graphics

CASCADE	ΣΤΑΔΙΑ ΠΟΥ ΥΛΟΠΟΙΗΘΗΚΑΝ	ΔΙΑΡΚΕΙΑ ΕΚΠ/ΣΗΣ ⁸⁶	ΠΑΡ/ΣΕΙΣ
HAAR_GAP_20stages	17	3h29m53s	α. Ο συγκεκριμένος χρόνος καταγράφηκε με την ολοκλήρωση του 17ου σταδίου β. Τερματίστηκε στην έναρξη του 18ου σταδίου με το μήνυμα «Required leaf false alarm rate achieved. Branch training terminated.» ⁸⁷
HOG_GAP_20stages	15	2h9m20s	α. Ο συγκεκριμένος χρόνος καταγράφηκε με την ολοκλήρωση του 15ου σταδίου β. Τερματίστηκε στην έναρξη του 16ου σταδίου με το μήνυμα «Required leaf false alarm rate achieved. Branch training terminated.»
HAAR_GAP_20stages (MaxFAR : 0,3)	11	>10 ημέρες	Διακόπηκε λόγω πτώσης της τάσης του ηλεκτρικού ρεύματος. Σε κάθε περίπτωση ο χρόνος που απαιτήθηκε για την εκπαίδευση των 11 σταδίων θα καθιστούσε απαγορευτική τη συνέχισή της, δεδομένου και της εκθετικής αύξησης του χρόνου εκπαίδευσης για κάθε επόμενο στάδιο.
LBP_GAP_15stages	15	5m17s	Ολοκληρώθηκε επιτυχώς
LBP_GAP_20stages	20	29m28s	-//-
LBP_GAP_25stages	25	37h32m0s	-//-

Πίνακας 9. Στοιχεία εκπαίδευσης των Cascade αλγορίθμων

Με την ολοκλήρωση της δοκιμαστικής φάσης της εκπαίδευσης, στοιχεία της οποίας παρουσιάστηκαν στον παραπάνω πίνακα, και της εφαρμογής της σε ένα μέρος του test set διαπιστώθηκε⁸⁸:

- Ότι η εφαρμογή της διαδικασίας της ανίχνευσης ήταν πιο γρήγορη με τη χρήση του LBP_GAP, ενώ ταυτόχρονα δε φαίνονταν να υστερεί σημαντικά, ως προς την απόδοση, από τον HAAR_GAP. Το συγκεκριμένο συμπέρασμα υπό το πρίσμα της απαίτησης για λειτουργία της ανίχνευσης σε πραγματικό χρόνο είναι ιδιαίτερος σημαντικό και θα πρέπει να συνεξεταστεί ποσοτικά πλέον στη φάση της κανονικής εκπαίδευσης.

⁸⁶ Η εκπαίδευση των HAAR/GAB, LBP/GAB και HOG/GAB πραγματοποιήθηκε σε desktop PC που διαθέτει επεξεργαστή Intel® Core (TM) i5-4460 CPU @ 3.2 GHz, μνήμη RAM 4 GB, 64-bit, το λειτουργικό του είναι Windows 10 Pro και φέρει κάρτα γραφικών NVIDIA GeForce GTX 750.

⁸⁷ Το συγκεκριμένο μήνυμα δεν αποτελεί σφάλμα. Σημαίνει ότι βάσει των παραμέτρων που δόθηκαν για την εκπαίδευση, αλλά και τον υφιστάμενο αριθμό των δειγμάτων, έχει επιτευχθεί η απαιτούμενη ακρίβεια. Δηλαδή, ορίζοντας $\text{minHitRate} = 0.995$, το μοντέλο δεν θα επιτρέψει περισσότερα από 5% θετικών δειγμάτων να κατηγοριοποιηθούν εσφαλμένα ενώ με $\text{maxFalseAlarmRate} = 0.5$ επιδιώκεται κάθε στάδιο να επιτύχει ορθή κατηγοριοποίηση των αρνητικών η οποία θα είναι λίγο καλύτερη από την τυχαιότητα (συμβολίζεται με το 0,5). Επομένως, προκειμένου η εκπαίδευση να ολοκληρωθεί με τον ζητούμενο αριθμό σταδίων θα πρέπει είτε να τεθούν παράμετροι οι οποίες είναι πιο δύσκολο να επιτευχθούν είτε να αυξηθεί ο αριθμός των δειγμάτων.

⁸⁸ Λόγω της δοκιμαστικής φύσης της συγκεκριμένης φάσης, οι διαπιστώσεις ήταν περισσότερο ποιοτικής φύσης και όχι ποσοτικής με ακριβείς καταγραφές.

- Η εμφανής διαφορά μεταξύ LBP_GAP_20stages και LBP_GAP_25stages αναφορικά με τον αριθμό των False Positives. Επομένως, για τον αλγόριθμο που θα προκριθεί από την εκπαίδευση 20 σταδίων, θα πρέπει να εξεταστεί η περεταίρω εκπαίδευσή του (πιθανώς σε βάρος υπερπροσαρμογής του μοντέλου στο δείγμα εκπαίδευσης).
- Η εκθετική αύξηση του χρόνου εκπαίδευσης σε κάθε στάδιο. Για παράδειγμα, κατά την εκπαίδευση του LBP_GAP_25stages, ενώ για σύνολο των 24 σταδίων απαιτήθηκε χρόνος περίπου 12h30m, μόνο για το 25ο στάδιο απαιτήθηκαν 25h. Εν προκειμένω, η εκπαίδευση του LBP_GAP ήταν ταχύτερη από τα υπόλοιπα μοντέλα κατηγοριοποιητών ακολουθίας (αναφορές στο διαδίκτυο κάνουν λόγο ότι απαιτεί 3 φορές λιγότερο χρόνο σε σχέση με το Haar). Η συγκεκριμένη παρατήρηση έχει μεγάλη αξία καθόσον ο χρόνος, αποτελεί περιορισμένο πόρο. Ένας τρόπος εξοικονόμησης χρόνου θα μπορούσε να είναι η εκμετάλλευση περισσότερων υπολογιστικών πόρων.
- Ότι παρά τον πολύ μικρό αριθμό δειγμάτων του HOG_SVM απέδωσε ικανοποιητικά σε όλους τους εξεταζόμενους δείκτες (TP, FP) και καλύτερα από τα υπόλοιπα μοντέλα, πλην όμως ήταν αρκετά αργός στη σάρωση της εικόνας.
- Ότι τα 15 στάδια εκπαίδευσης του HOG_GAP δεν επιτρέπουν την εξαγωγή συμπερασμάτων ως προς την απόδοσή τους και θα πρέπει να επανεξεταστεί κατά την φάση της κανονικής εκπαίδευσης των αλγορίθμων.
- Η απαίτηση για λήψη μέρους των θετικών δειγμάτων κατά την εκπαίδευση των κατηγοριοποιητών ακολουθίας.
- Ότι όσο μειώνονταν το video resolution, επί του οποίου δοκιμαζόταν οι αλγόριθμοι, μειώνονταν όχι μόνο ο χρόνος επεξεργασίας της εικόνας (λόγω της μικρότερης έκτασης που θα πρέπει να καλύψουν τα παράθυρα σάρωσης της εικόνας) αλλά και ο αριθμός των True Positive Detection.
- Ότι για την ολοκλήρωση της εκπαίδευσης με αποφυγή λήψης του μηνύματος «Required leaf false alarm rate achieved», εκτός της αύξησης των δειγμάτων, και εφόσον απαιτηθεί, θα πρέπει να μειωθεί και η τιμή του maxFAR. Ταυτόχρονα θα πρέπει να αποφευχθεί η αύξηση του MinHitRate άνω του 0,995, προκειμένου να μην υπάρξει υπερπροσαρμογή στο μοντέλο εκπαίδευσης.

5.1.3 Εκπαίδευση των Κατηγοριοποιητών

Για τη φάση της εκπαίδευσης, δημιουργήθηκε αρχικά στην VERVwUAS DB ένα δείγμα 4400 αρνητικών εικόνων (859 από την αγροτική περιοχή και 3541 από τον αστικό ιστό) και 1400 θετικών εικόνων (348 εικόνες από την αγροτική περιοχή και οι υπόλοιπες 1052 από την αστική περιοχή). Αντίστοιχα, το δείγμα εκπαίδευσης για το μοντέλο HOG/SVM δημιουργήθηκε από 184 εικόνες οι οποίες προέρχονται από πλαίσια των video από τις περιοχές 1 και 2.

Μετά την ολοκλήρωση της φάσης των δοκιμών και αφού πραγματοποιήθηκαν οι πτήσεις και στην 2η περιοχή, ολοκληρώθηκε η διαδικασία συμπλήρωσης του training set. Τα συμπεράσματα της δοκιμαστικής φάσης ήταν αυτά που καθόρισαν και τη στρατηγική διεξαγωγής της κανονικής εκπαιδευτικής διαδικασίας.

Αρχικά αποφασίστηκε να μη διεξαχθεί η εκπαίδευση στους προσωπικούς Η/Υ (desktop και laptop), αλλά σε διαδικτυακό server. Η συγκεκριμένη απόφαση είχε διττή στόχευση. Από τη μία, λόγω του ότι η εκπαίδευση είναι χρονοβόρα, θα αποφεύγονταν περιπτώσεις διακοπής της λόγω απρόοπτων καταστάσεων (πχ όπως συνέβη στην φάση δοκιμών με τη διακοπή του ηλεκτρικού ρεύματος). Από την άλλη, θα δίνονταν η δυνατότητα δημιουργίας μίας διαμόρφωσης η οποία θα παρείχε περισσότερους υπολογιστικούς πόρους και επομένως θα καθιστούσε ταχύτερη την όλη διαδικασία. Με γνώμονα τα παραπάνω, χρησιμοποιήθηκε η πλατφόρμα που προσφέρει η Google (Google Cloud Platform), η οποία περιέχει και υπηρεσίες μηχανικής μάθησης.

Αφού ενοικιάστηκε ένας server που λειτουργούσε σε Windows 10 με 8 CPUs στα 52 GB⁸⁹ και εγκαταστάθηκαν τα απαιτούμενα προγράμματα για την εργασία της εκπαίδευσης, πραγματοποιήθηκε μία δοκιμή⁹⁰ για να πιστοποιηθεί η λειτουργικότητα του συστήματος. Η ουσιαστική διαφορά των ρυθμίσεων σε σχέση με τις προηγούμενες δοκιμαστικές εκπαιδεύσεις, ήταν ο ορισμός της τιμής των "Pre-calculated Values Buffer Size" και "Pre-calculated Indices Buffer Size" σε 32768 Mb αντί των 2048 Mb, που χρησιμοποιήθηκαν στον προσωπικό Η/Υ. Αν και η δοκιμή ήταν επιτυχής, το γεγονός ότι διήρκεσε 10m29s προξένησε προβληματισμό καθόσον για το ίδιο training data set και τις ίδιες παραμέτρους, πλην των τιμών του Buffer Size, ο οικιακός υπολογιστής ήταν

⁸⁹ Η ονομασία που δόθηκε αυτόματα από τη Google στο δημιουργούμενο server (instance 1) ήταν «lucid – parsec-197322»

⁹⁰ Η εκπαίδευση έγινε σε LBP_GAP cascades λόγω του ότι η εκπαίδευσή τους διαρκεί λιγότερο από τα HAAR_GAP και HOG_GAP

κατά 17s ταχύτερος.

Μετά από μελέτη του προβλήματος, εκτιμήθηκε ότι η αιτία πίσω από αυτό ήταν η αδυναμία διεξαγωγής παράλληλης επεξεργασίας. Η αναζήτηση τρόπων, μέσω του διαδικτύου, για την επίλυση της κατάστασης οδήγησε σε αναφορά στην ιστοσελίδα του OpenCV.org⁹¹ που επεσήμαινε ότι: « Η εφαρμογή `opencv_traincascade` δύναται να εκμεταλλευτεί τα TBB⁹² για multi-threading λειτουργία. Προκειμένου να χρησιμοποιηθεί σε διαμόρφωση πολλαπλών πυρήνων (multicore mode), το OpenCV θα πρέπει να δημιουργηθεί με TBB».

Ανεξαρτήτως από την ύπαρξη επισημάνσεων στο αντίστοιχο forum του OpenCV.org⁹³ περί μικρής μείωσης του χρόνου εκπαίδευσης, παρά την παραλληλοποίηση της επεξεργασίας με TBB⁹⁴, επιχειρήθηκε να γίνει `compile` του OpenCV με TBB. Ακόμα και αν επιβεβαιώνονταν ο παραπάνω ισχυρισμός, θα μπορούσε τουλάχιστον να γίνει μία σύγκριση μεταξύ των χρόνων εκπαίδευσης με ή χωρίς παράλληλη επεξεργασία, να εξαχθούν και να καταγραφούν τα ανάλογα συμπεράσματα. Δυστυχώς, όμως δεν κατέστη εφικτό να γίνει `compile` του OpenCV με TBB. Εν προκειμένω, κρίνεται σκόπιμο να αναφερθεί ότι οι μόνες αναφορές στο διαδίκτυο που έκαναν λόγο για επιτυχές `compilation`, αφορούσαν λειτουργικό Unix (Ubuntu). Επιπλέον, προκειμένου να αντισταθμιστεί αυτή η αδυναμία, ενοικιάστηκε και δεύτερος `server`⁹⁵ ώστε να πραγματοποιείται ταυτόχρονα η εκπαίδευση σε LBP_GAB και HAAR_GAP. Έτσι ξεκίνησε ουσιαστικά η Φάση Ια της εκπαίδευσης με τις χρησιμοποιούμενες

⁹¹ https://docs.opencv.org/3.0-beta/doc/user_guide/ug_traincascade.html (Πρόσβαση:21/3/18)

⁹² Σύμφωνα με την ιστοσελίδα των Intel® Threading Building Blocks (<https://www.threadingbuildingblocks.org/>)(Πρόσβαση:21/3/18), «αυτά διευκολύνουν τη συγγραφή παράλληλων C++ προγραμμάτων τα οποία εκμεταλλεύονται στο έπακρο την πολυπύρηνη απόδοση».

⁹³ <http://answers.opencv.org/question/63/enable-multithreading-with-tbb-during-cascade-training/>(Πρόσβαση:20/3/18)

⁹⁴ Σε απάντηση που επισυνάπτεται από τη Maria Dimashova στις 4 Ιουλίου 2012, αναφέρεται ότι : «Ακόμα και με χρήση των TBB δεν θα παρατηρηθεί επαρκής ανάθεση φόρτου εργασίας, από το OpenCV, σε πυρήνες CPU κατά την εφαρμογή της εκπαίδευσης `traincascade`. Σχεδόν ένας πυρήνας θα λειτουργεί για το σύνολο του χρόνου εκπαίδευσης. Αυτό συμβαίνει επειδή μόνο ένα μικρό τμήμα του κώδικα εκπαίδευσης παραλληλοποιείται από τα TBB. Αυτός αφορά στην εύρεση του βέλτιστου διαχωρισμού των κόμβων του δέντρου απόφασης (tree node) και στον προ-υπολογισμό μέρους των τιμών των χαρακτηριστικών, πριν την εκπαίδευση του νέου σταδίου. Όμως, για μεγάλο μέρος του χρόνου, η εκπαίδευση `traincascade` αναζητά τα αρνητικά δείγματα τα οποία αναγνωρίστηκαν ως θετικά, από τα προηγούμενα στάδια, προκειμένου να εκπαιδεύσει το επόμενο. Η συγκεκριμένη διαδικασία δεν παραλληλοποιείται».

⁹⁵ Και ο δεύτερος `server` («dogwood-cinema-197917») δημιουργήθηκε ώστε να διαθέτει τα ίδια χαρακτηριστικά με τον πρώτο.

παράμετρους να φαίνονται στον πίνακα 9.

CASCADE	ΣΤΑΔΙΑ	M	Π	ΠΟΣΟΣΤΟ ΘΕΤΙΚΩΝ ΕΙΚΟΝΩΝ	ΑΡΙΘΜΟΣ ΑΡΝΗΤΙΚΩΝ ΕΙΚΟΝΩΝ	HAAR MODE	MIN HIT RATE	MAX FAR	WTR	MWC
HAAR_GAP	20	58	24	80 - 90% (1190-1260)	4400	BASIC	0.995	0,5	0,95	100
LBP_GAP	20	-/-	-/-	-/-	-/-	---	-/-	-/-	-/-	-/-

Πίνακας 10. Αρχικές παράμετροι εκπαίδευσης των LBP_GAB και HAAR_GAP

Με βάση την πρότερη εμπειρία από τη δοκιμαστική φάση, το αρχικά χρησιμοποιούμενο ποσοστό θετικών εικόνων ήταν 90%. Τόσο για το συγκεκριμένο ποσοστό όμως όσο και για τα αντίστοιχα που ακολούθησαν με φθίνουσα τιμή μέχρι το 80% είχαν ως αποτέλεσμα τη διακοπή της εκπαίδευσης με την ένδειξη : «OpenCV Error: Bad argument (Can not get new positive sample. The most possible reason is insufficient count of samples in given vec-file.)». Αλλάζοντας στρατηγική αποφασίστηκε η χρησιμοποίηση μικρής σχετικά τιμής για το χρησιμοποιούμενο ποσοστό των θετικών/αρνητικών εικόνων και σε περίπτωση λήψης μηνύματος : «Required leaf false alarm rate achieved. Branch training terminated», η σταδιακή αύξηση της τιμής μέχρι εκείνο το σημείο που θα περατώνονταν η εκπαίδευση έχοντας ολοκληρώσει το σύνολο των καθορισμένων σταδίων. Αναλυτικά στοιχεία για τις χρησιμοποιούμενες παραμέτρους και τα αποτελέσματα της εκπαίδευσης φαίνονται στους πίνακες 10, 11 και 12 για τα μοντέλα LBP_GAB, HAAR_GAP και HOG_GAP αντίστοιχα (έχουν τηρηθεί τα log files και οι χρόνοι διάρκειας εκπαίδευσης για όσες από αυτές πέτυχαν αριθμό υλοποιηθέντων σταδίων μεγαλύτερο του 18 (δηλ. 18 από 20) για τα LBP/HOG_GAB ή του 17 (δηλ. 18 από 20) για το Haar.

A/A	ΠΟΣΟΣΤΟ ΘΕΤΙΚΩΝ ΕΙΚΟΝΩΝ	ΑΡΙΘΜΟΣ ΑΡΝΗΤΙ- ΚΩΝ ΕΙΚΟΝΩΝ	MIN HIT RATE	MAX FAR	ΣΤΑΔΙΑ ΠΟΥ ΥΛΟΠΟΙ Η- ΘΗΚΑΝ	ΔΙΑΡΚΕΙΑ ΕΚΠ/ΣΗΣ	ΠΑΡ/ΣΕΙΣ
1	60%	3000	0.995	0,5	18	5h58m4s	Τερματίστηκε στην έναρξη του 19ου σταδίου με το μήνυμα «Required leaf false alarm rate achieved. Branch training terminated».
2	-/-	3500	-/-	-/-	15	6h23m33s	<ul style="list-style-type: none"> • Τερματίστηκε στην έναρξη του 16ου σταδίου με το μήνυμα «Required leaf false alarm rate achieved. Branch training terminated». • Συνάγεται το συμπέρασμα ότι διατηρώντας σταθερό τον αριθμό των θετικών δειγμάτων και αυξάνοντας τον αριθμό των αρνητικών, μειώνεται ο αριθμός των απαιτούμενων

A/A	ΠΟΣΟΣΤΟ ΘΕΤΙΚΩΝ ΕΙΚΟΝΩΝ	ΑΡΙΘΜΟΣ ΑΡΝΗΤΙΚΩΝ ΕΙΚΟΝΩΝ	MIN HIT RATE	MAX FAR	ΣΤΑΔΙΑ ΠΟΥ ΥΛΟΠΟΙΗΘΗΚΑΝ	ΔΙΑΡΚΕΙΑ ΕΚΠ/ΣΗΣ	ΠΑΡ/ΣΕΙΣ
							σταδίων.
3	70%	3500	-/-	-/-	18	5h 47m 5s	<ul style="list-style-type: none"> • Τερματίστηκε στην έναρξη του 19ου σταδίου με το μήνυμα «Required leaf false alarm rate achieved. Branch training terminated». • Σε σχέση με το 1ο setting) εκπαίδευσης (A/A 1) φαίνεται ότι αύξηση τόσο των θετικών όσο και αρνητικών δειγμάτων οδηγεί στον ίδιο αριθμό ολοκληρωθέντων σταδίων.
4	72%	-/-	-/-	-/-	10	Δεν καταγράφηκε	Τερματίστηκε με το μήνυμα « OpenCV Error: Bad argument (Can not get new positive sample ... vec-file) ⁹⁶ ».
5	-//-	3000	-/-	-/-	13	Δεν καταγράφηκε	Τερματίστηκε με το μήνυμα « OpenCV Error: Bad argument (Can not get new positive sample ... vec-file)».
6	71%	3500	-/-	-/-	14	Δεν καταγράφηκε	<ul style="list-style-type: none"> • Τερματίστηκε στην έναρξη του 15ου σταδίου με το μήνυμα «Required leaf false alarm rate achieved. Branch training terminated». • Συγκρίνοντας τις τιμές του συγκεκριμένου setting σε σχέση με αυτές του 3ου, 4ου και 5ου αλλά και τα αποτελέσματα που έδωσαν αυτές οι τέσσερις εκπαιδεύσεις, διαφάνηκε ότι θα πρέπει να αναζητηθούν και άλλες εναλλακτικές πέραν της τροποποίησης του αριθμού των ληφθέντων δειγμάτων.
7	70%	3500	0,996	-/-	18	5h14m55s	<ul style="list-style-type: none"> • Τερματίστηκε στην έναρξη του 19ου σταδίου με το μήνυμα «Required leaf false alarm rate achieved. Branch training terminated». • Παρά το ζήτημα υπερπροσαρμογής στο δείγμα εκπαίδευσης, η τιμή του Min Hit Rate αυξήθηκε προκειμένου να μελετηθεί η συμπεριφορά της εκπαίδευσης. • Σε σχέση με το 3ο setting εκπαίδευσης που είχε τον ίδιο αριθμό θετικών-αρνητικών δειγμάτων, το δημιουργούμενο από την εκπαίδευση cascade.xml αρχείο είχε μέγεθος 31 KB αντί 26 KB του A/A 3.
8	70%	3500	0,995	0,49	16	Δεν καταγράφηκε	• Τερματίστηκε στην έναρξη του 17ου σταδίου με το μήνυμα «Required leaf false

⁹⁶ Η αιτία του συγκεκριμένου μηνύματος είναι η ακόλουθη: Κατά την εκπαιδευτική διαδικασία μπορεί να απορριφθούν θετικά δείγματα που κατηγοριοποιήθηκαν εσφαλμένα σε ένα συγκεκριμένο στάδιο και να μην επιτραπεί η περαιτέρω χρήση τους. Σε αυτή την περίπτωση το σύστημα θα ανατρέξει στα θετικά δείγματα που αφέθηκαν εκτός (βάσει τις τιμές του "numPOS" σε σχέση με τον πραγματικό αριθμό των θετικών δειγμάτων), προκειμένου να τα εισάγει στη διαδικασία της εκπαίδευσης. Για παράδειγμα, αν οριστεί αυτή η τιμή σε 650 τότε στο log file φαίνεται ότι στο 1ο στάδιο εκπαίδευσης ο λόγος POS count:consumed=650:650. Σε επόμενα στάδια αυτός ο λόγος να γίνει 650:663. Επομένως αν το περιθώριο μεταξύ "numPOS" και συνολικών θετικών δειγμάτων είναι μικρό, πιθανώς θα ληφθεί το συγκεκριμένο μήνυμα.

A/A	ΠΟΣΟΣΤΟ ΘΕΤΙΚΩΝ ΕΙΚΟΝΩΝ	ΑΡΙΘΜΟΣ ΑΡΝΗΤΙΚΩΝ ΕΙΚΟΝΩΝ	MIN HIT RATE	MAX FAR	ΣΤΑΔΙΑ ΠΟΥ ΥΛΟΠΟΙΗΘΗΚΑΝ	ΔΙΑΡΚΕΙΑ ΕΚΠ/ΣΗΣ	ΠΑΡ/ΣΕΙΣ
						κε	alarm rate achieved. Branch training terminated».
9	-/-	-/-	-/-	0,48	17	Δεν καταγράφη κε	<ul style="list-style-type: none"> • Στοιχείο που χρήζει επισήμανσης είναι ότι παρά τα ίδια settings με το A/A 3 η μείωση της τιμής του Max FAR, οδήγησε σε ενωρίτερο τερματισμό της εκπαίδευσης, αντί το αντίθετο. • Τερματίστηκε στην έναρξη του 18ου σταδίου με το μήνυμα «Required leaf false alarm rate achieved. Branch training terminated». • Σε σχέση με την προηγούμενη εκπαίδευση φαίνεται ότι μείωση της τιμής του Max FAR, οδηγεί σε αργότερο τερματισμό της εκπαίδευσης.
10	-/-	3200	-/-	0,48	17	Δεν καταγράφη κε	<ul style="list-style-type: none"> • Τερματίστηκε στην έναρξη του 18ου σταδίου με το μήνυμα «Required leaf false alarm rate achieved. Branch training terminated». • Η μείωση του αριθμού των αρνητικών δειγμάτων, δεν συνοδεύτηκε από αντίστοιχη αύξηση του αριθμού σταδίων που ολοκληρώθηκαν.
11	-/-	3500	-/-	0,465	18	33h41m19s	Τερματίστηκε στην έναρξη του 19ου σταδίου με το μήνυμα «Required leaf false alarm rate achieved. Branch training terminated».
12	-/-	-/-	-/-	0,445	18	39h5m11s	Τερματίστηκε στην έναρξη του 19ου σταδίου με το μήνυμα «Required leaf false alarm rate achieved. Branch training terminated».
13	-/-	3000	-/-	-/-	16	Δεν καταγράφη κε	<ul style="list-style-type: none"> • Τερματίστηκε στην έναρξη του 16ου σταδίου με το μήνυμα «Required leaf false alarm rate achieved. Branch training terminated». • Παρά το γεγονός ότι από τις προηγούμενες εκπαιδεύσεις φαίνεται ότι η μείωση των αρνητικών δειγμάτων επιφέρει αύξηση των υλοποιηθέντων σταδίων εκπαίδευσης, αυτό δεν επιβεβαιώθηκε στη συγκεκριμένη περίπτωση.
14	-/-	4000	-/-	-/-	17	Δεν καταγράφη κε	Τερματίστηκε στην έναρξη του 17ου σταδίου με το μήνυμα «Required leaf false alarm rate achieved. Branch training terminated».
Παράμετροι Εκπαίδευσης							
Values Buffer Size : 32768 Mb Indices Buffer Size : 32768 Mb Acceptance Ratio Break Value : -1 Number of Threads : 5				Number of Stages : 20 Sample Width : 58 Sample Height : 24 Boost Type : GAP		Weight Trim Rate : 0,95 Maximal Depth Weak Tree : 1 Maximal Weak Trees : 100	

Πίνακας 11. Στοιχεία εκπαίδευσης του LBP_GAB (Φάση IIα)

A/A	ΠΟΣΟΣΤΟ ΘΕΤΙΚΩΝ ΕΙΚΟΝΩΝ	ΑΡΙΘΜΟΣ ΑΡΝΗΤΙΚΩΝ ΕΙΚΟΝΩΝ	MIN HIT RATE	MAX FAR	ΣΤΑΔΙΑ ΠΟΥ ΥΛΟΠΟΙΗΘΗΚΑΝ	ΔΙΑΡΚΕΙΑ ΕΚΠ/ΣΗΣ	ΠΑΡ/ΣΕΙΣ
1	50%	2000	0,98	0,5	15	2h 7m 56s	<ul style="list-style-type: none"> • Τερματίστηκε στην έναρξη του 16ου σταδίου με το μήνυμα «Required leaf false alarm rate achieved. Branch training terminated». • Η εκπαίδευση του δείγματος ξεκίνησε χρησιμοποιώντας μικρότερες τιμές θετικών – αρνητικών δειγμάτων, σε σχέση με τις αντίστοιχες του LBP, σύμφωνα με τα συμπεράσματα της δοκιμαστικής φάσης εκπαίδευσης. • Παρά το ζήτημα υπερπροσαρμογής στο δείγμα εκπαίδευσης, η τιμή του Min Hit Rate αυξήθηκε προκειμένου να μελετηθεί η συμπεριφορά της εκπαίδευσης. • Ο χρόνος εκπαίδευσης που καταγράφηκε ήταν με την ολοκλήρωση του 15ου σταδίου.
2	-/-	2500	-/-	-/-	17	6h 44m 36s	Τερματίστηκε στην έναρξη του 18ου σταδίου με το μήνυμα «Required leaf false alarm rate achieved. Branch training terminated».
3	-/-	3000	-/-	-/-	17	10h 32m 9s	Τερματίστηκε στην έναρξη του 18ου σταδίου με το μήνυμα «Required leaf false alarm rate achieved. Branch training terminated».
4	60%	3000	0,99 5	-/-	19	8h 23m 44s	Τερματίστηκε στην έναρξη του 20ου σταδίου με το μήνυμα «Required leaf false alarm rate achieved. Branch training terminated».
5	65%	-/-	-/-	-/-	18	5h 23m 20s	Τερματίστηκε στην έναρξη του 19ου σταδίου με το μήνυμα «Required leaf false alarm rate achieved. Branch training terminated».
6	68%	-/-	-/-	-/-	16	Δεν καταγράφηκε	Τερματίστηκε στην έναρξη του 17ου σταδίου με το μήνυμα «Required leaf false alarm rate achieved. Branch training terminated».
7	-/-	3500	-/-	-/-	19	8h 53m 44s	Τερματίστηκε στην έναρξη του 20ου σταδίου με το μήνυμα «Required leaf false alarm rate achieved. Branch training terminated».
8	70%	-/-	-/-	-/-	17	Δεν καταγράφηκε	Τερματίστηκε με το μήνυμα « OpenCV Error: Bad argument (Can not get new positive sample ... vec-file)».
9	68%	-/-	-/-	0,49	20	35h 16m 8s	Ολοκληρώθηκε η εκπαίδευση
Παράμετροι Εκπαίδευσης							
Values Buffer Size : 32768 Mb Indices Buffer Size : 32768 Mb Acceptance Ratio Break Value : -1 Number of Threads : 5				Number of Stages : 20 Sample Width : 58 Sample Height : 24 Boost Type : GAP		Weight Trim Rate : 0,95 Maximal Depth Weak Tree : 1 Maximal Weak Trees : 100 HAAR Mode : BASIC	

Πίνακας 12. Στοιχεία εκπαίδευσης του HAAR_GAB (Φάση ΙΙα)

A/A	ΠΟΣΟΣΤΟ ΘΕΤΙΚΩΝ ΕΙΚΟΝΩΝ	ΑΡΙΘΜΟΣ ΑΡΝΗΤΙΚΩΝ ΕΙΚΟΝΩΝ	MIN HIT RATE	MAX FAR	ΣΤΑΔΙΑ ΠΟΥ ΥΛΟΠΟΙΗΘΗΚΑΝ	ΠΑΡ/ΣΕΙΣ	
1	68%	3500	0,995	0,5	15	Τερματίστηκε εσκεμμένα μετά από 26 περίπου ώρες λόγω του ότι βάσει της εκθετικής αύξησης του χρόνου διάρκειας της εκπαίδευσης σε κάθε επόμενο στάδιο, διαφαίνονταν ότι για την ολοκλήρωσή της θα απαιτούνταν $\cong 2$ μήνες.	
2	60%	3000	-/-	-/-	19	45h43m23s	Τερματίστηκε στην έναρξη του 20ου σταδίου με το μήνυμα «Required leaf false alarm rate achieved. Branch training terminated».
3	-/-	3000	-/-	0,49	16	Τερματίστηκε εσκεμμένα μετά από 23 περίπου ώρες λόγω του ότι βάσει της εκθετικής αύξησης του χρόνου διάρκειας της εκπαίδευσης σε κάθε επόμενο στάδιο, διαφαίνονταν ότι για την ολοκλήρωσή της θα απαιτούνταν $\cong 1$ μήνας.	
4	62%	3000	-/-	0,5	19	81h23m44s	Τερματίστηκε στην έναρξη του 20ου σταδίου με το μήνυμα «Required leaf false alarm rate achieved. Branch training terminated».
Παράμετροι Εκπαίδευσης							
Values Buffer Size : 32768 Mb Indices Buffer Size : 32768 Mb Acceptance Ratio Break Value : -1 Number of Threads : 5				Number of Stages : 20 Sample Width : 58 Sample Height : 24 Boost Type : GAP		Weight Trim Rate : 0,95 Maximal Depth Weak Tree : 1 Maximal Weak Trees : 100	

Πίνακας 13. Στοιχεία εκπαίδευσης του HOG_GAB (Φάση Ια)

Η αδυναμία ολοκλήρωσης της εκπαίδευσης σε LBP και HOG είχε ως αποτέλεσμα την επανεξέταση της όλης διαδικασίας. Έχοντας ως σημείο εκκίνησης το forum του OpenCV.org και τις βιβλιογραφικές αναφορές στις οποίες παρέπεμπαν, αντλήθηκαν ή και έγιναν περισσότερο κατανοητά τα παρακάτω:

- Η παράμετρος "numNeg" δεν αφορά αυτές καθ' αυτές τις εικόνες του αρνητικού δείγματος αλλά αναφέρεται σε "negative windows" τα οποία αντλούνται από τις αρνητικές εικόνες. Αποτελεί τον αριθμό των αρνητικών δειγμάτων που χρησιμοποιούνται σε κάθε στάδιο εκπαίδευσης.
- Για περιπτώσεις όπου το φόντο (background) παρουσιάζει ποικιλομορφία, όπως στη συγκεκριμένη περίπτωση, μία καλή αναλογία μεταξύ θετικών και αρνητικών δειγμάτων, για την εκκίνηση της εκπαίδευσης, ισούται με 0,5.
- Κάθε αρνητικό παράθυρο που κατηγοριοποιείται σωστά σε ένα στάδιο δεν χρησιμοποιείται ξανά σε επόμενα στάδια αυτής. Οπότε είναι σημαντικό να υπάρχει όσο το δυνατόν μεγαλύτερη "μοναδικότητα" στα παράθυρα που αντλούνται για την εκπαίδευση. Για το σκοπό, χρησιμοποιήθηκαν και επιπλέον εικόνες από το διαδίκτυο οι οποίες θα ενίσχυαν τη συγκεκριμένη απαίτηση.

- Οι διαστάσεις width (πλάτος) και height (ύψος) καθορίζουν το μικρότερο μέγεθος αντικειμένου που θα δύναται να ανιχνεύσει ο αλγόριθμος. Αυτή είναι και η αιτία για την οποία υπήρξε ανεπιτυχής η προσπάθεια ανίχνευσης οχημάτων σε video ανάλυσης 960x540 (το μέσο μέγεθος των οχημάτων σε αυτή την περίπτωση ήταν 43x18, δηλαδή 62% μικρότερο από το παράθυρο 58x24 με την οποία εκπαιδεύτηκε ο αλγόριθμος).

Λαμβάνοντας υπόψη τα παραπάνω, επαναπροσεγγίστηκε η τακτική δημιουργίας του training data set. Πιο συγκεκριμένα επειδή από την περιοχή 1 είχαν ληφθεί 347 δείγματα και προκειμένου να είναι ισοδύναμο αριθμητικά το εν λόγω set με αυτό της περιοχής 2, λήφθηκαν 450⁹⁷ δείγματα από την τελευταία περιοχή, σχηματίζοντας έτσι ένα set 797 θετικών εικόνων. Αναφορικά με το αρνητικό set εικόνων, λήφθηκαν αυτές με τις μεγαλύτερες διαστάσεις ενώ προστέθηκαν και επιπλέον εικόνες από το διαδίκτυο. Συνολικά το αρνητικό δείγμα σχηματίστηκε από 1621 εικόνες (476 από το διαδίκτυο, 455 από την αστική περιοχή και 690 από την αγροτική περιοχή).

Ο πρώτος αλγόριθμος ανίχνευσης, ο οποίος εκπαιδεύτηκε με αυτό το νέο δείγμα εικόνων, ήταν ο LBP_GAP, λόγω του ότι απαιτεί το μικρότερο χρόνο. Εκπαιδεύτηκε σε 20 στάδια και το εξαχθέν αρχείο ανίχνευσης (cascade.xml) εφαρμόστηκε δοκιμαστικά σε δύο video που προέρχονταν από την περιοχή 1 και 2 αντίστοιχα. Ενώ τα αποτελέσματα ήταν ικανοποιητικά για την περιοχή 2 (αστικός ιστός), δε συνέβαινε το ίδιο για την περιοχή 1. Αναζητώντας τα αίτια πίσω από αυτή την αστοχία, αποκαλύφθηκε ότι ενώ οι θετικές εικόνες στο αρχείο που τηρούνταν ήταν καταχωρημένες με αύξουσα ονομαστική και χρονολογική σειρά (δηλαδή πρώτες αυτές από την περιοχή 1 που λήφθηκαν τον Ιανουάριο και στη συνέχεια οι αντίστοιχες της περιοχής 2), το pos.txt file που δημιουργήθηκε, από το οποίο και προέρχεται το pos-samples.vec, τις καταχωρούσε ξεκινώντας από τις πιο πρόσφατες. Αυτό είχε ως αποτέλεσμα κατά την εκτέλεση του opencv_haartraining να λαμβάνονται πρώτες για εκπαίδευση οι φωτογραφίες της περιοχής 2. Το συγκεκριμένο γεγονός, σε συνδυασμό με τον ορισμό 638 θετικών δειγμάτων στην τιμή της παραμέτρου numPos, κατέληξε στη λήψη για εκπαίδευση του συνόλου των 450 εικόνων της περιοχής 2 και μόνο 188

⁹⁷ Δόθηκε λίγο μεγαλύτερη βαρύτητα στα δείγματα που προέρχονται από τον αστικό ιστό λόγω της μεγαλύτερης πολυπλοκότητας της περιοχής.

από την περιοχή 1 (638-450 =188)⁹⁸. Το ζήτημα επιλύθηκε χρησιμοποιώντας το λογισμικό "BulkFileChanger⁹⁹", με το οποίο πραγματοποιήθηκε αλλαγή του attribute "created", προκειμένου το σύστημα να θεωρεί πιο πρόσφατες τις εικόνες από την περιοχή 1. Η επανεκπαίδευση του δείγματος και τα αποτελέσματα ανίχνευσης στα παραπάνω video, επιβεβαίωσαν την υπόθεση περί της προέλευσης της αρχικής αστοχίας. Αντίστοιχα, σε ότι αφορά στην εκπαίδευση του HOG_SVM, από το training set των 184 εικόνων, επισημάνθηκαν οι συντεταγμένες 852 οχημάτων που περιλαμβάνονταν εντός αυτών¹⁰⁰.

Έτσι, εφόσον επιλύθηκαν τα ζητήματα των training sets, ξεκίνησε η τελική διαδικασία της εκπαίδευσης των μοντέλων που καθορίστηκαν από τα ερευνητικά ερωτήματα (ουσιαστικά πρόκειται για τη φάση ΙΙβ:επαναπροσέγγιση της εκπαίδευσης). Το σύνολο των εκπαιδεύσεων, με τις χρησιμοποιούμενες παραμέτρους για τα HAAR/GAB, LBP/GAB, HOG/GAB και HOG/SVM, φαίνονται στους πίνακες 14, 14 και 15, αντίστοιχα.

CAS-CADE	ΣΤΑΔΙΑ	M	Π	ΠΟΣΟΣΤΟ ΘΕΤΙΚΩΝ ΕΙΚΟΝΩΝ	ΑΡΙΘΜΟΣ ΑΡΝΗΤΙΚΩΝ Ν ΕΙΚΟΝΩΝ	HAAR MODE	MAX FAR	ΔΙΑΡΚΕΙΑ ΕΚΠ/ΣΗΣ
HAAR	20	29	12	85%(677)	1600	BASIC	0.5	24m 7s
-/-	-/-	-/-	-/-	-/-	-/-	CORE	-/-	38m 44s
-/-	-/-	-/-	-/-	-/-	-/-	ALL	-/-	43m 49s
-/-	25	-/-	-/-	-/-	-/-	ALL	-/-	18h3m27s
LBP	-/-	-/-	-/-	-/-	-/-	-----	-/-	6m 32s
-/-	25	-/-	-/-	-/-	-/-	-----	-/-	3h 21m 4s
-/-	-/-	-/-	-/-	82%(653)	1621	-----	0.45	4d 20h 26m 53s
HOG	20	58	24 ¹⁰¹	85%(677)	1600	-----	0,5	1d 1h 28m 4s

Παράμετροι Εκπαίδευσης

Values Buffer Size : 32768 Mb	Boost Type : GAP	Maximal Depth Weak Tree : 1
Indices Buffer Size : 32768 Mb	min Hit Rate:0,995	Maximal Weak Trees : 100
Acceptance Ratio Break Value : -1	Weight Trim Rate : 0,95	
Number of Threads : 5		

Πίνακας 14. Χρησιμοποιούμενες παράμετροι για την εκπαίδευση των HAAR/GAB, LBP/GAB και HOG/GAB

⁹⁸ Το συγκεκριμένο πρόβλημα (αδυναμία ανίχνευσης οχημάτων στην περιοχή 1), ήταν πιο έντονο στα xml αρχεία που προήλθαν από την εκπαίδευση με 1400 θετικές εικόνες. Σε αυτή τη περίπτωση, λαμβάνοντας numPos = 75% (1050), οι εικόνες της περιοχής 1 έμεναν εκτός εκπαιδευτικής διαδικασίας.

⁹⁹ BulkFileChanger v1.51, Copyright (c) 2010 - 2017 Nir Sofer, Web site: <http://www.nirsoft.net>. Το συγκεκριμένο πρόγραμμα δίνει τη δυνατότητα τροποποίησης ορισμένων χαρακτηριστικών των αρχείων (πχ ημερομηνία δημιουργίας/προσβασης/τροποποίησης κλπ)

¹⁰⁰ Η δημιουργία του train set παρουσιάζεται αναλυτικά στην εικόνα 53

¹⁰¹ Χρησιμοποιήθηκε η μεγαλύτερη διάσταση καθόσον το πρόγραμμα εκπαίδευσης απέριπτε το παράθυρο 29X12

CASCADE ΣΤΑΔΙΑ		HAAR MODE	TOTAL POSITIVE CONSUMED	NUMBER OF UNIQUE FEATURES	ACCEPTANCE RATIO
HAAR	20	BASIC	677:731	59700	8,41624*10 ⁻⁶
-/-	-/-	CORE	677:737	76856	4,811194*10 ⁻⁶
-/-	-/-	ALL	677:735	87779	3,77607*10 ⁻⁶
-/-	25	ALL	677:753	-/-	1,27905*10 ⁻⁷
LBP	-/-	-----	677:733	2970	2,24697*10 ⁻⁵
-/-	25	-----	677:772	2970	1,46939*10 ⁻⁶
-/-	-/-	-----	653:749	2970	7,19376*10 ⁻⁹
HOG	20	-----	677:734	8	2,0571*10 ⁻⁵

Πίνακας 15. Χρησιμοποιούμενες παράμετροι για την εκπαίδευση των HAAR/GAB, LBP/GAB και HOG/GAB

FEATURE EXTRACTION	LEARNING ALGORITHM	ΑΡΙΘΜΟΣ ANNOTATED EIKONΩΝ	ΣΥΝΟΛΟ AN-NOTATED ΟΧΗΜΑΤΩΝ	ΔΙΑΡΚΕΙΑ ΕΚΠ/ΣΗΣ	ΠΑΡ/ΣΕΙΣ
HOG	SVM	184	852	---	Διακόπηκε λόγω μη επαρκούς μνήμης
-/-	-/-	160	691	---	-/-
-/-	-/-	130	563	---	-/-
-/-	-/-	115	464	1h47m	Ολοκληρώθηκε

Πίνακας 16. Χρησιμοποιούμενες παράμετροι για την εκπαίδευση του HOG/SVM

5.2 Χρησιμοποιούμενο Λογισμικό – Αλγόριθμοι Ανίχνευσης/Ιχνηλάτησης – Αξιολόγηση Μοντέλων

Το λογισμικό που χρησιμοποιήθηκε για την ανάπτυξη των προγραμμάτων ανίχνευσης – ιχνηλάτησης των οχημάτων ήταν συνδυασμός κώδικα γραμμένου σε Python, με χρησιμοποίηση βιβλιοθηκών του OpenCV και της Dlib.

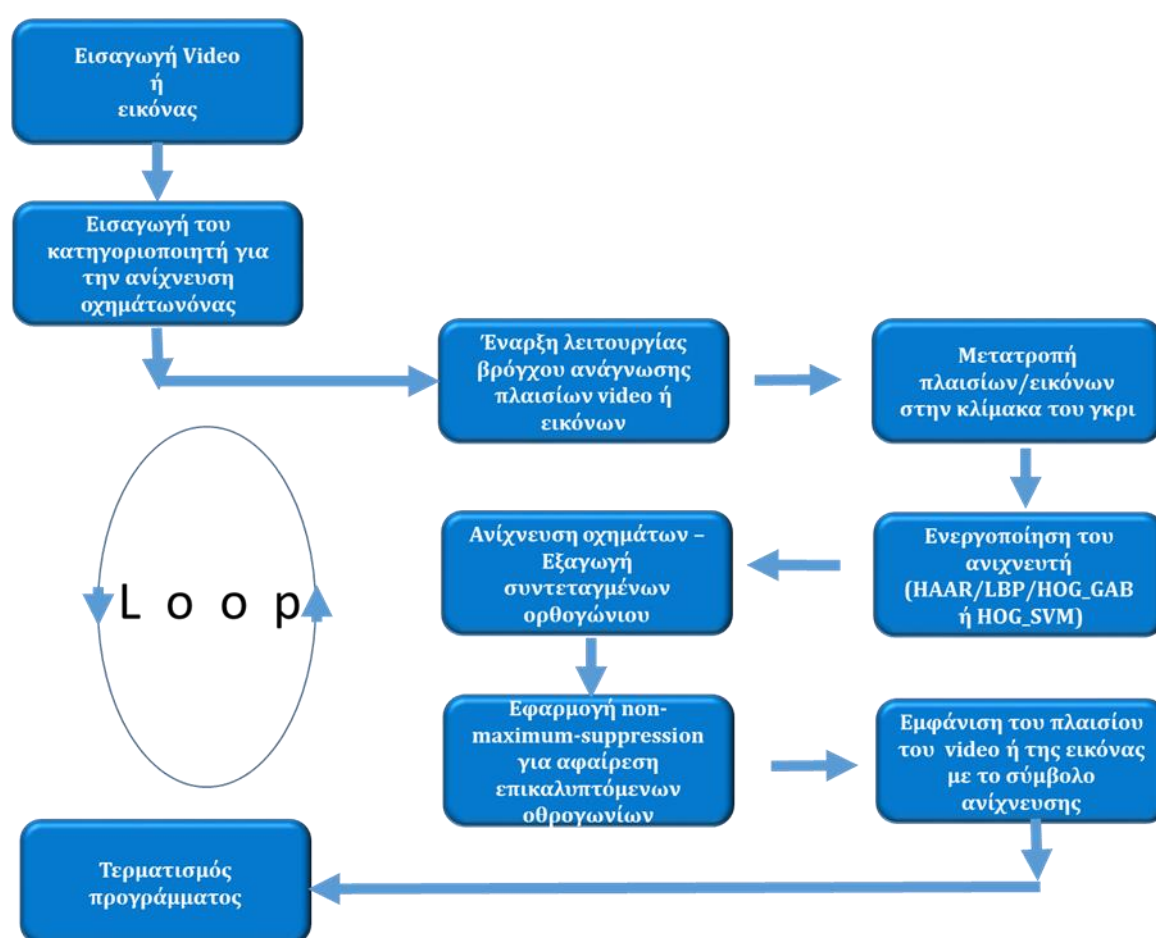
- Στο desktop, το οποίο χρησιμοποιήθηκε για την εκπαίδευση και αξιολόγηση ανίχνευσης των μοντέλων HAAR/GAB, LBP/GAB και HOG/GAB, εγκαταστάθηκε η έκδοση Python 2.7.13 και OpenCV 2.4 και ως IDE για τα προγράμματα της Python χρησιμοποιήθηκε το Pycharm έκδοσης 2017 2.4.
- Για την εκπαίδευση και αξιολόγηση ανίχνευσης του HOG/SVM καθώς επίσης και την αξιολόγηση της ιχνηλάτησης όλων των αλγορίθμων χρησιμοποιήθηκε το laptop Lenovo T510. Για τη συγγραφή σε Python έγινε εκμετάλλευση του SPYDER¹⁰² (Scientific Python Development EnviRonment) IDE¹⁰³. Το SPYDER ενσωματώνει βασικές βιβλιοθήκες όπως οι NumPy, SciPy, Matplotlib και

¹⁰² Χρησιμοποιήθηκε η έκδοση 3.1.4

¹⁰³ Το συγκεκριμένο λογισμικό για τα Windows είναι διαθέσιμο μέσω του Anaconda η οποία αποτελεί πλατφόρμα ελεύθερου και ανοικτού κώδικα για προγραμματισμό σε Python και R.

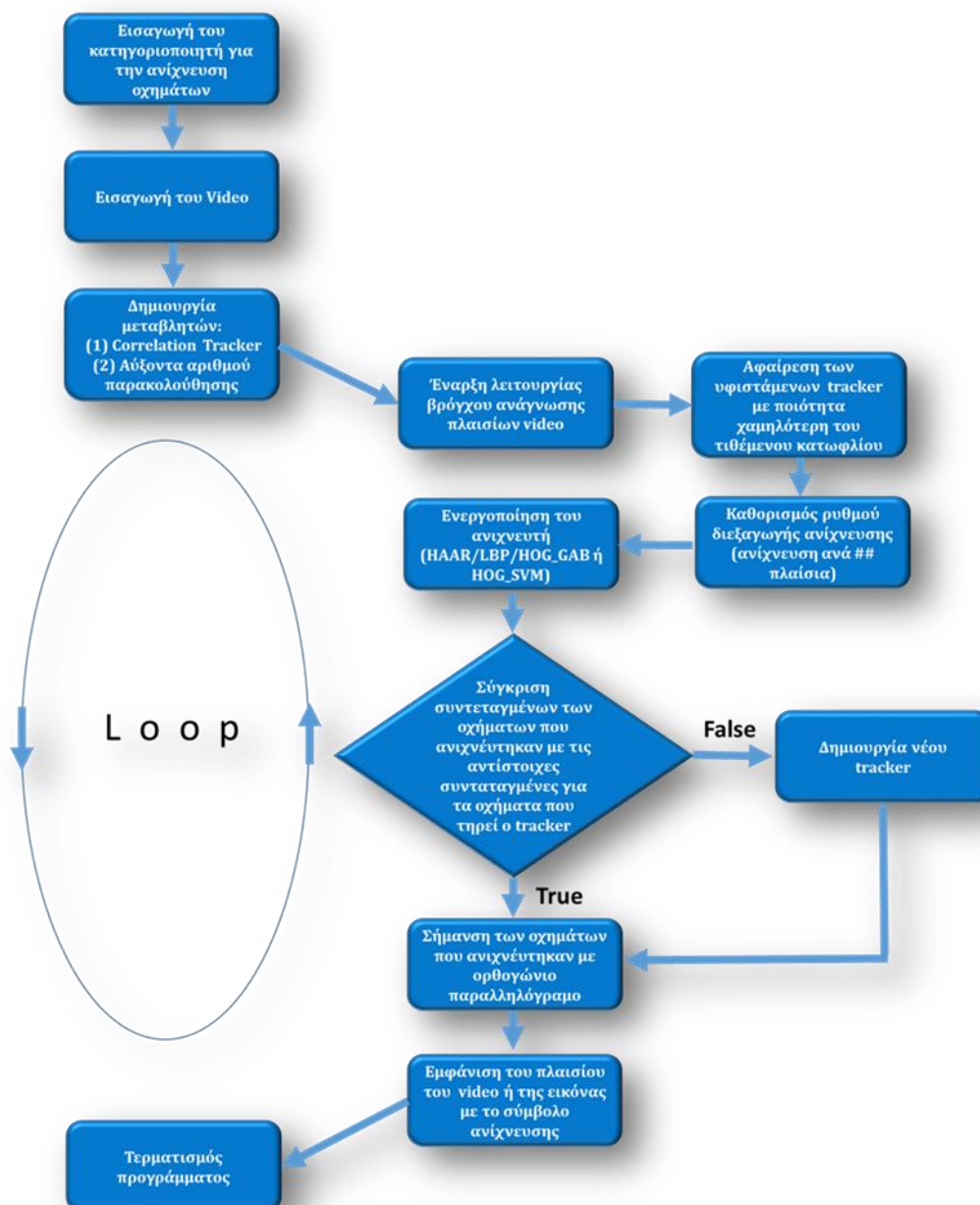
IPython.

Η OpenCV (Open Computer Vision) αποτελεί βιβλιοθήκη ανοιχτού κώδικα η οποία εκδίδεται με άδεια κλάσης Berkeley Software Distribution (BSD) και είναι ελεύθερη για ακαδημαϊκή και επιστημονική χρήση. Η OpenCV είναι γραμμένη σε C/C++ και στοχεύει στην ανάπτυξη εφαρμογών μηχανικής όρασης πραγματικού χρόνου (real time). Διαθέτει διεπαφές σε C++, Python και Java, ενώ υποστηρίζεται από λειτουργικά Windows, Mac OS, iOS και Android. Η βιβλιοθήκη Dlib, που χρησιμοποιήθηκε για την εκπαίδευση σε HOG/SVM, διαθέτει ένα μεγάλο εύρος αλγορίθμων μηχανικής μάθησης. Η εκτέλεσή της γίνεται μέσω C++ API.



Εικόνα 51. Αλγόριθμος ανίχνευσης

Για τους σκοπούς της μεταπτυχιακής διατριβής δημιουργήθηκαν 2 αλγόριθμοι σε Python. Ο πρώτος (εικόνα 51) στοχεύει στην ανίχνευση των οχημάτων ενώ ο δεύτερος (εικόνα 52), για τους λόγους που αναλύθηκαν στην ενότητα 3.1 συνδυάζει την ανίχνευση με την ιχνηλάτηση. Οι εν λόγω αλγόριθμοι υλοποιήθηκαν ξεχωριστά για τους HAAR/GAB, LBP/GAB, HOG/GAB και HOG_SVM.

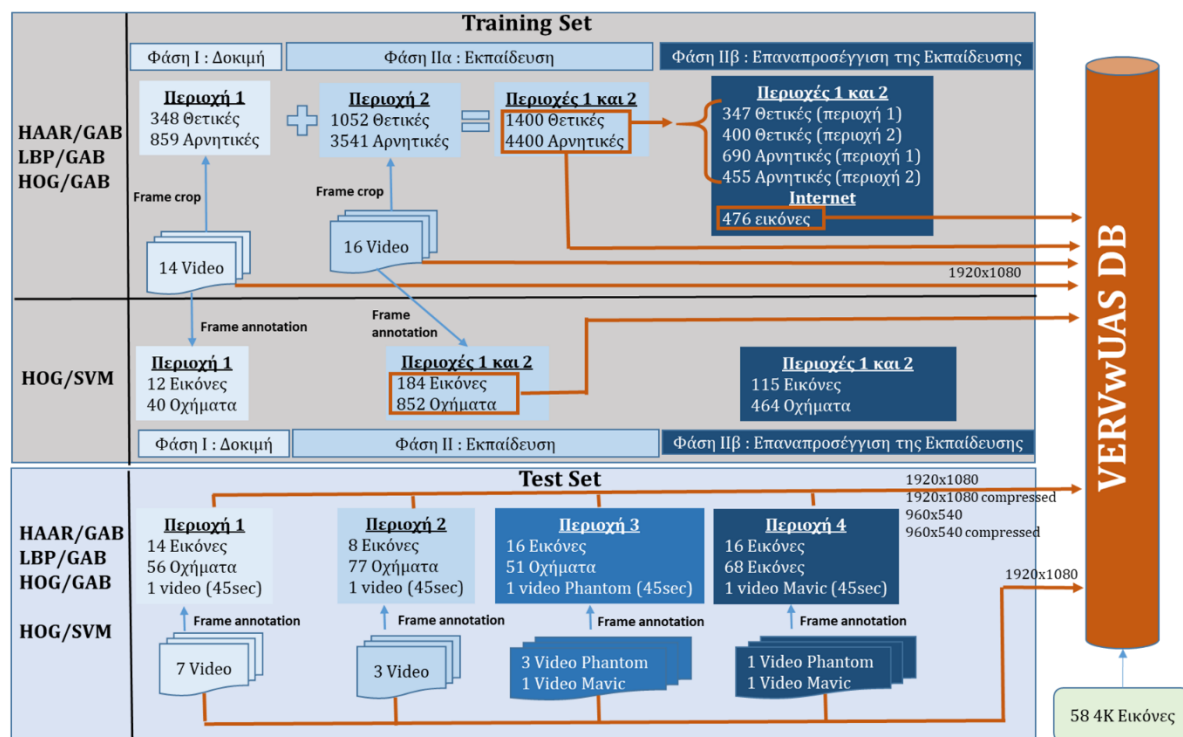


Εικόνα 52. Αλγόριθμος ιχνηλάτησης οχημάτων

Για την αξιολόγηση των αλγορίθμων ανίχνευσης, χρησιμοποιήθηκαν 252 εικόνες προερχόμενες από πλαίσια των video που τηρήθηκαν για την εργασία του ελέγχου. Αυτές, μετά τη διεργασία της επισήμανσης των οχημάτων¹⁰⁴, εισήχθησαν στον αλγόριθμο ανίχνευσης (εικόνα 51) και καταγράφηκαν τόσο οι συντεταγμένες των αντικειμένων που ανιχνεύτηκαν (true positive, false positive) όσο και εκείνων που δεν ανιχνεύτηκαν (false negative). Στη συνέχεια, αφού οι παραπάνω συντεταγμένες καταχωρήθηκαν στο ίδιο φύλλο excel με αυτό του annotation, χρησιμοποιήθηκε

¹⁰⁴ Τα δεδομένα του annotation (x,y, width, heigh) τοποθετήθηκαν σε ένα αρχείο excel.

κώδικας σε Matlab για τον υπολογισμό της ακρίβειας της ανίχνευσης. Το πρόγραμμα έκανε εισαγωγή των συντεταγμένων του ορθογωνίου του annotation (A) σε σχέση με αυτό που εξάγει ο αλγόριθμος ανίχνευσης (B) ώστε να υπολογίσει το λόγο IoU^{105} της τομής των A,B προς την ένωσή τους : $IoU = \frac{A \cap B}{A \cup B}$. Ως true positive λογίστηκαν οι ανιχνεύσεις οχημάτων με $IoU \geq 0,5$. Η διαδικασία αξιολόγησης ολοκληρώθηκε σε τρία στάδια:



Εικόνα 53. Διαδικασία δημιουργίας της VERVwUAS DB μέσω της εκπαίδευσης και ελέγχου των κατηγοριοποιητών

- Στο **πρώτο στάδιο**, αφού λήφθηκαν υπόψη οι τιμές των TP, TN και FN, υπολογίστηκαν για κάθε μοντέλο οι δείκτες accuracy, error rate, recall, precision, F καθώς και ο συνολικός χρόνος που απαιτήθηκε για τη σάρωση των 54 εικόνων του test set. Οι συγκεκριμένες μετρήσεις σε αυτό το πρώτο επίπεδο, εφαρμόστηκαν για τα μοντέλα των πινάκων 14 και 15. Οι διαστάσεις των 54 φωτογραφιών που υποβλήθηκαν στην διαδικασία της ανίχνευσης ήταν 1920x1080 και 960x540¹⁰⁶.
- Στο **δεύτερο στάδιο** αξιολόγησης, τα δύο μοντέλα που είχαν την καλύτερη

¹⁰⁵ Intersection over Union

¹⁰⁶ Αναλυτικά η σύνθεση του test set φαίνεται στην εικόνα 53

απόδοση κατά το πρώτο στάδιο, βάσει του δείκτη F και του χρόνου ανίχνευσης, υποβλήθηκαν ξανά στην διαδικασία της ανίχνευσης οχημάτων, χρησιμοποιώντας τις φωτογραφίες του πρώτου σταδίου, οι οποίες όμως είχαν υποστεί συμπίεση της τάξης του 40%. Η ελάττωση της ποιότητας των εικόνων έγινε προκειμένου να εξομοιωθεί η περίπτωση μείωσης του ρυθμού μετάδοσης δεδομένων (bit rate), κατά το video streaming από την εναέρια πλατφόρμα στο σταθμό εδάφους¹⁰⁷.

- Στο **τρίτο στάδιο** της αξιολόγησης, εξετάστηκε η απόδοση των δύο παραπάνω μοντέλων με τη χρησιμοποίηση αυτή τη φορά των αντίστοιχων video του test set (4 video διάρκειας 0,45sec έκαστο, προερχόμενα από τις 4 περιοχές λήψης). Σκοπός του συγκεκριμένου σταδίου ήταν η καταμέτρηση του ποσοστού των οχημάτων τα οποία αφού ανιχνεύτηκαν, μεταβιβάστηκαν στη συνέχεια στον tracker. Η μεθοδολογία που σχεδιάστηκε να εφαρμοστεί στο συγκεκριμένο στάδιο περιλάμβανε την μετάδοση ροής video εντός του οικιακού δικτύου (από το desktop στο laptop) με χρήση του VLC server. Στη συνέχεια, η ροή video στο δέκτη (laptop) διαβιβάστηκε στο πρόγραμμα ανίχνευσης και ιχνηλάτησης οχημάτων, για την αξιολόγηση της ευρωστίας του συνολικού μοντέλου. Το bit rate μετάδοσης καθορίστηκε βάσει των αποτελεσμάτων που προέκυψαν από τη μέτρηση της ταχύτητας του δικτύου, κατά τις πτήσεις με Ε/Π.

5.3 Μέτρηση Ταχύτητας Δικτύου

Από τη σύγκριση των δικτύων κινητής τηλεφωνίας στην Ελλάδα, αναφορικά με την γενιά στην οποία εντάσσονται, όπως αναλύθηκαν στην υποενότητα 3.3.5, φαίνεται ότι για την περιοχή της Μαγνησίας το δίκτυο της Cosmote είναι αυτό που εξασφαλίζει ταχύτητες 3G και 4G¹⁰⁸. Έχοντας αυτό ως δεδομένο, πραγματοποιήθηκε προμήθεια του router HUAWEI 4G USB Stick E3372h – 153 στο οποίο τοποθετήθηκε κάρτα sim της Cosmote (τεχνικά χαρακτηριστικά όπως φαίνονται στον πίνακα 16).

Στη συνέχεια το 4G USB Stick συνδέθηκε στο laptop Lenovo T510 και πραγματοποιήθηκαν πτήσεις με Ε/Π προκειμένου να μετρηθεί η ταχύτητα (upload) του δικτύου. Οι πτήσεις έλαβαν χώρα σε δύο διαφορετικές περιοχές, στην περιοχή των

¹⁰⁷ Η απαίτηση για μείωση του bit rate συνδέεται με το διαθέσιμο εύρος του δικτύου κινητής τηλεφωνίας και θα καθοριστεί από τις τιμές upload που θα μετρηθούν.

¹⁰⁸ Υπενθυμίζεται ότι η Vodafone εντάσει το EDGE στα δίκτυα 3G.

Δήμων Ρήγα Φεραίου -Κιλελέρ και στην περιοχή της πόλης του Βόλου. Η κάθε περιοχή κάλυπτε μία έκταση περίπου 150Km². Η επιλογή τους έγινε λόγω του ότι σύμφωνα με το χάρτη της εικόνας 30, η πρώτη καλύπτεται από 3G δίκτυο και η δεύτερη από 4G, παρέχοντας έτσι τη δυνατότητα πραγματοποίησης συγκρίσεων. Σε κάθε περιοχή σχεδιάστηκε και υλοποιήθηκε πτήση με ταχύτητες 50 Knots ($25,72ms^{-1}$) και 100 Knots ($51,44 ms^{-1}$) και σε ύψη πτήσης 500ft (133m), 1000ft (333m) και 3000ft (1000m). Με αυτό τον τρόπο δίνεται η ευκαιρία να εξεταστεί η ταχύτητα του δικτύου σε μία ευρεία σχετικά περιοχή τριών διαστάσεων, ενώ ταυτόχρονα οι διαφορετικές διευθύνσεις παρέχουν τη δυνατότητα εξαγωγής συμπερασμάτων για ύπαρξη επιπτώσεων που οφείλονται στο Doppler spread.

Υποστηριζόμενα Επικοινωνιακά Συστήματα	FDD: DD800 / 900 / 1800 / 2100 / 2600 UMTS: 900 / 2100 GSM: 850 / 900 / 1800 / 1900
Ταχύτητες	LTE FDD: Cat4 DL: 150 Mbps / UL: 50 Mbps @20 M BW UMTS: DCHSPA+: 42 / 5.76 Mbps; 21 M / 5.76 Mbps; 14 M / 5.76 M HSUPA: 7.2 M / 5.76 M 2G: EDGE packet data service of up to 236.8 kbps
Λοιπές Γενικές Πληροφορίες	Υποστηρίζει τοποθέτηση εξωτερικής κεραίας TS-5*2 Ο δέκτης του διαθέτει δυνατότητα διαφορικής απολαβής (receiver diversity)

Πίνακας 17. Τεχνικά χαρακτηριστικά HUAWEI 4G USB Stick E3372h – 153

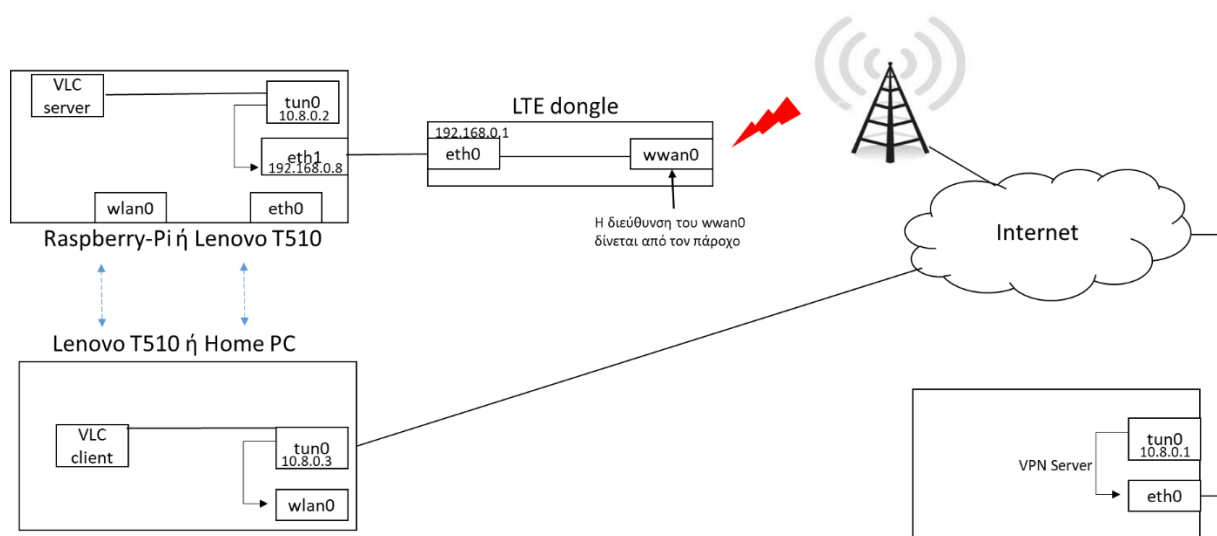


Εικόνα 54. Περιοχές πτήσης για τη μέτρηση της ταχύτητας του δικτύου 3G/4G

5.4 Video Streaming από Ε/Π

Η μέτρηση της upload ταχύτητας του δικτύου κινητής τηλεφωνίας αποτελεί και τον οδηγό για τον καθορισμό του bit rate του video που θα επιχειρηθεί να αποσταλεί από το Ε/Π στο σταθμό εδάφους. Η επιδίωξη είναι να διαπιστωθεί εάν είναι εφικτή η ροή video,

σε σχεδόν πραγματικό χρόνο, με τιμή bit rate η οποία δε θα υπερβαίνει την ελάχιστη ταχύτητα που προσφέρει το δίκτυο. Συναφώς επισημαίνεται ότι το bit rate του video streaming θα πρέπει να είναι τέτοιο ώστε, όταν εφαρμόζεται στο σταθμό εδάφους το μοντέλο ανίχνευσης και ιχνηλάτισης οχημάτων που επιλέχθηκε, αυτό να παρουσιάζει ικανοποιητική απόδοση.



Εικόνα 55. Διάταξη VPN για τη ροή video μέσω του δικτύου κινητής τηλεφωνίας

Το video streaming δοκιμάστηκε σε δύο φάσεις:

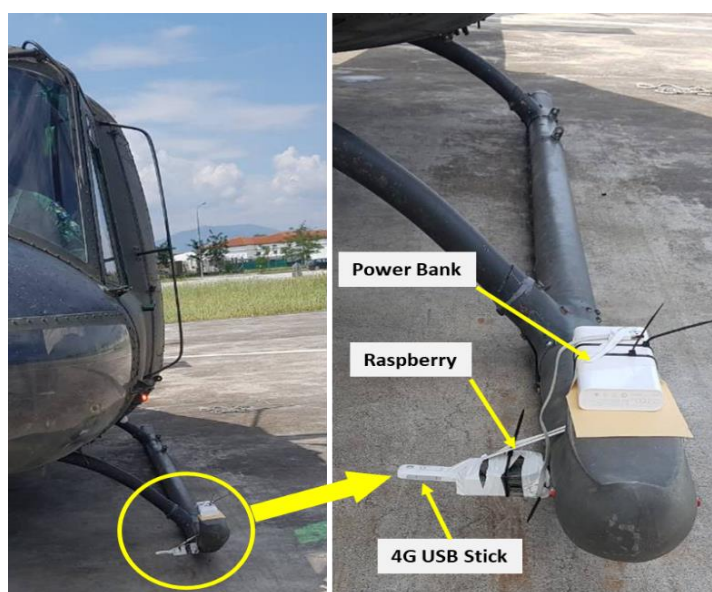
- Σε πρώτη φάση επιχειρήθηκε η ροή δεδομένων video από το Ε/Π προς το σταθμό εδάφους χρησιμοποιώντας video συγκεκριμένων και προεπιλεγμένων διαστάσεων και bit rate. Για την ασύρματη δοκιμή διακίνησης του video μέσω του δικτύου κινητής τηλεφωνίας, λόγω της μη δυνατότητας port forwarding στο 4G USB stick της HUAWEI, δημιουργήθηκε ένα VPN δίκτυο μεταξύ του laptop Lenovo T510, που τοποθετήθηκε εντός του Ε/Π, και του οικιακού desktop. Το laptop συνδέθηκε στο mobile internet της Cosmote μέσω του HUAWEI 4G USB Stick, ενώ το desktop διέθετε ενσύρματη σύνδεση στο internet¹⁰⁹. Στους δύο Η/Υ εγκαταστάθηκε το λογισμικό OpenVPN¹¹⁰ το οποίο αποτελεί ένα ολοκληρωμένο SSL VPN που υλοποιεί τα επίπεδα 2 και 3 ασφαλούς προέκτασης OSI δικτύου, εκμεταλλευόμενο το πρωτόκολλο SSL/TSL. Η εγκατάσταση του Open VPN στα Windows περιλαμβάνει και το αντίστοιχο γραφικό περιβάλλον διεπαφής με το χρήστη (OpenVPN GUI) για τη διαχείριση της σύνδεσης. Ο φορέας μέσω του

¹⁰⁹ Αν και η ονομαστική ταχύτητα ήταν 24Mbps, η μέση πραγματική (download) που μετρήθηκε κυμαίνονταν μεταξύ 2,8 και 4,7 Mbps.

¹¹⁰ Το OpenVPN GUI αποτελεί Copyright © 2002-2018 OpenVPN Inc. (<https://openvpn.net/>)

οποίου πραγματοποιήθηκε η ροή δεδομένων του video ήταν ο αναπαραγωγός πολυμέσων VLC¹¹¹. Μέσω αυτού πραγματοποιήθηκε επίσης και η μετατροπή του αποσταλμένου video στις επιθυμητές διαστάσεις και bit rate (πρωτόκολλο συμπίεσης H.264).

- Μετά τη διαπίστωση των αποτελεσμάτων της ροής video και τον καθορισμό του κατάλληλου bit rate, υλοποιήθηκε η δεύτερη φάση της δοκιμής. Σύμφωνα με αυτή, στο Ε/Π τοποθετήθηκε η διάταξη που φαίνεται στην εικόνα 56. Αυτή περιλάμβανε τη χρησιμοποίηση ενός Raspberry Pi3 B+ στο οποίο συνδέθηκε το HUAWEI 4G USB Stick 3372h-153. Η τροφοδοσία του συστήματος επιτεύχθηκε με τη σύνδεσή του σε power bank. Για τη λήψη video συνδέθηκε, στην αντίστοιχη υποδοχή, η Raspberry Pi Camera Module v2 η οποία διαθέτει αισθητήρα IMX219 8 megapixel της Sony. Μετά την εγκατάστασή της, η κάμερα δύναται να υποστηρίξει τη λήψη video 1080p30, 760p60 και VGA90. Στη συνέχεια, μέσω του δικτύου VPN που περιγράφεται στην εικόνα 55 αποστάλθηκε στο laptop Lenovo T510 (διέθετε ενσύρματη σύνδεση στο internet), near real time video streaming. Αυτό δρομολογήθηκε προς το μοντέλο ανίχνευσης και ιχνηλάτησης οχημάτων προκειμένου να δοκιμαστεί η εφικτότητα αλλά και αποτελεσματικότητα της όλης διάταξης (end-to-end δοκιμή).

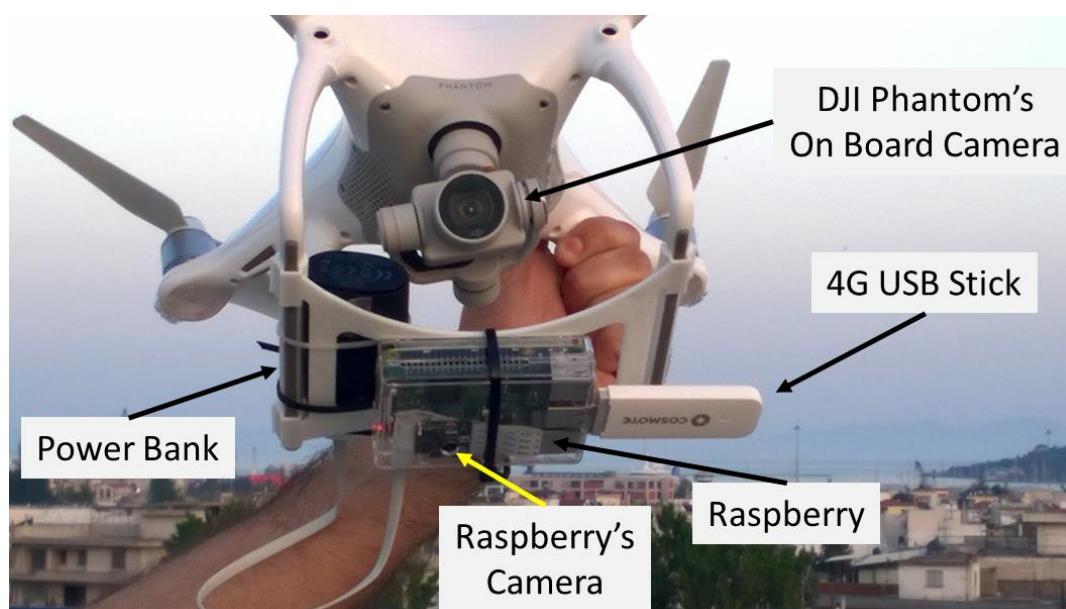


Εικόνα 56. Τοποθέτηση Raspberry Pi3 B+ επί του Ε/Π

¹¹¹ Η έκδοση που χρησιμοποιήθηκε ήταν η 2.2.8 Weatherwax

5.5 Ανίχνευση και Ιχνηλάτηση σε Video Streaming Μεταδιδόμενο από UAV

Το σημείο ολοκλήρωσης της μεταπτυχιακής διατριβής αποτελεί η εφαρμογή των αλγόριθμων ανίχνευσης και ιχνηλάτησης οχημάτων σε ροή video η οποία προέρχονταν από το UAS DJI Phantom. Επειδή το εν λόγω drone δε διαθέτει δυνατότητα ασύρματης αποστολής, μέσω του δικτύου της κινητής τηλεφωνίας, της ροής του video που καταγράφει η δική του (on board) κάμερα¹¹², χρησιμοποιήθηκε η διάταξη που αναφέρθηκε στην προηγούμενη παράγραφο και φαίνεται στην εικόνα 57 (Raspberry Pi3 B+, HUAWEI 4G USB Stick 3372h-153, power bank, Raspberry Pi Camera Module v2). Η λύση της IMX219 προτιμήθηκε, έναντι της τοποθέτησης μίας action camera τύπου go pro, για λόγους βάρους και την επίτευξη μιας πιο "συμπαγούς κατασκευής".



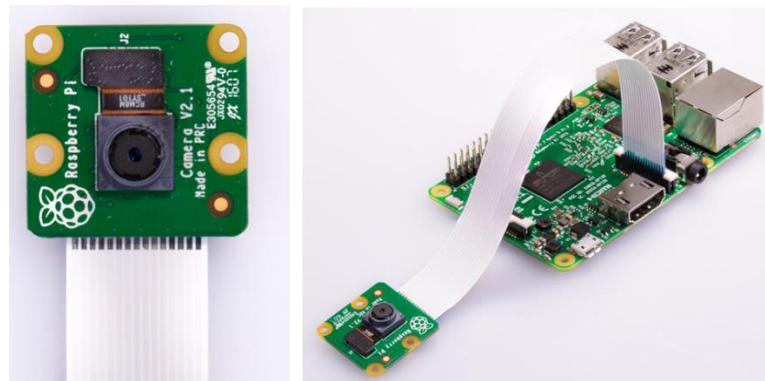
Εικόνα 57. Τοποθέτηση του Raspberry στο DJI Phantom

Για τη ροή των δεδομένων του video από το Raspberry (UAV) προς τον προσωπικό υπολογιστή (Lenovo T510 Laptop) χρησιμοποιήθηκε το μεταξύ τους VPN δίκτυο.

Λόγω του ότι παρασιτικές εκπομπές, μία πιθανή πηγή των οποίων είναι οι σερβοκινητήρες των drone, δύνανται να επηρεάσουν την μετάδοση υψηλού ρυθμού δεδομένων (Asadrour et al., 2014) πριν την απογείωση του DJI Phantom

¹¹² Το DJI Phantom αποστέλει ασύρματα το video που καταγράφει η on board κάμερα μέσω WiFi ασύρματης ζεύξης με το σταθμό εδάφους (2,4GHz)

πραγματοποιήθηκαν μετρήσεις της ταχύτητας download/upload, σε τρεις διαφορετικές καταστάσεις λειτουργίας του. Οι μετρήσεις πραγματοποιήθηκαν στην οροφή ενός κτιρίου και σε ύψος 12m προκειμένου οι συνθήκες να προσομοιάζουν το δυνατόν περισσότερο αυτές της πραγματικής πτήσης του UAS. Οι τρεις καταστάσεις λειτουργίας που αναφέρθηκαν συνδέονται με την ταχύτητα περιστροφής των πτερύγων και είναι : (1) Απενεργοποιημένο (off), (2) Λειτουργία μειωμένης ταχύτητας (idle) και (3) Λειτουργία πλήρους ταχύτητας (fly). Μετά την πραγματοποίηση των μετρήσεων το αερόχημα απογειώθηκε και εκτέλεσε αιώρηση σε ύψος 45m.



Εικόνα 58. Raspberry Pi Camera Module v2 IMX219

Κεφάλαιο

6

Αποτελέσματα

6.1 1ο Στάδιο Αξιολόγησης: Απόδοση των Μοντέλων Ανίχνευσης Οχημάτων

Όπως αναφέρθηκε στο προηγούμενο κεφάλαιο η αξιολόγηση των μοντέλων ανίχνευσης αντικειμένων θα υλοποιηθεί σε τρία στάδια. Στο πρώτο από αυτά τα στάδια θα πραγματοποιηθεί σύγκριση, βάσει ενός συνόλου 54 εικόνων προερχόμενο και από τις τέσσερις περιοχές της πόλης του Βόλου, των παρακάτω μοντέλων¹¹³:

- Haar, Classifier:GAP, Feature Sets : BASIC, Stages : 20, MHR:0,995, MaxFAR:0,5
- Haar, Classifier:GAP, Feature Sets : CORE, Stages : 20, MHR:0,995, MaxFAR:0,5
- Haar, Classifier:GAP, Feature Sets : ALL, Stages : 20, MHR:0,995, MaxFAR:0,5
- Haar, Classifier:GAP, Feature Sets : ALL, Stages : 25, MHR:0,995, MaxFAR:0,5
- LBP, Classifier:GAP, Stages : 20, MHR:0,995, MaxFAR:0,5
- LBP, Classifier:GAP, Stages : 25, MHR:0,995, MaxFAR:0,5
- LBP, Classifier:GAP, Stages : 25, MHR:0,995, MaxFAR:0,45
- HOG, Classifier:GAP, Stages : 20, MHR:0,995, MaxFAR:0,5
- HOG, Classifier:SVM

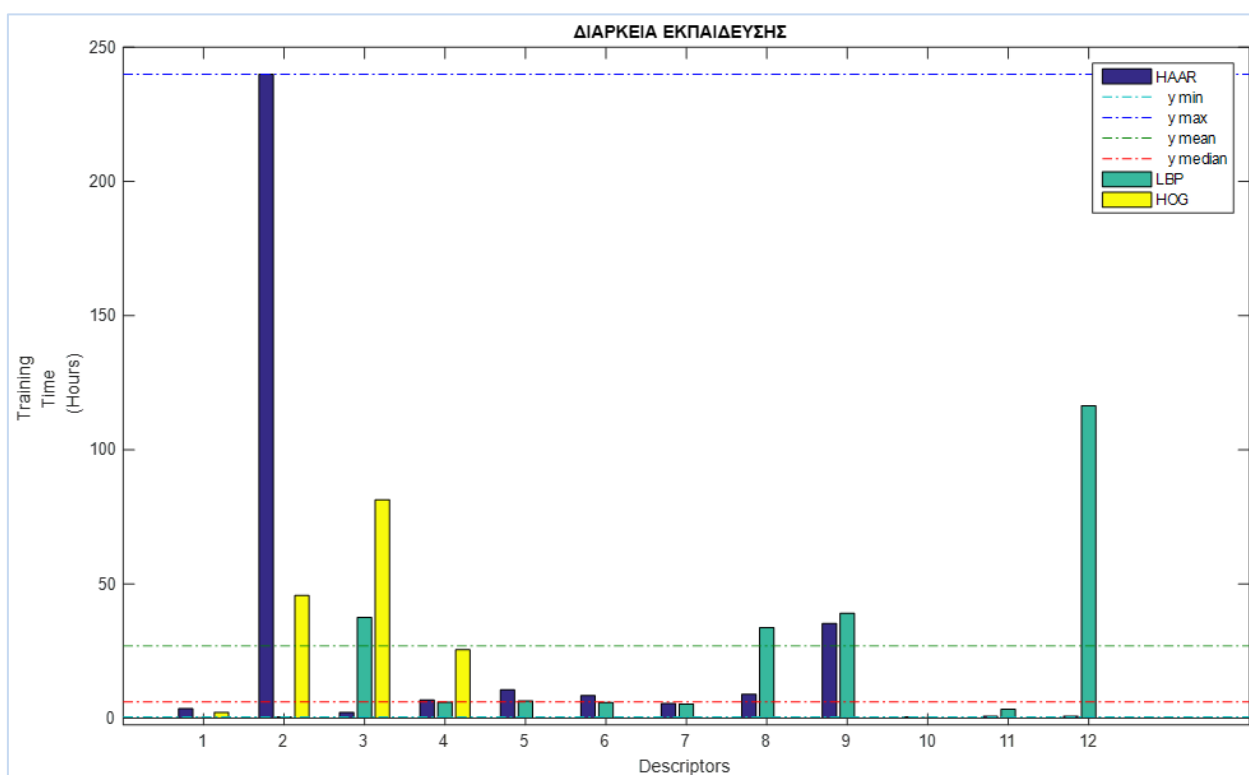
6.1.1 Διάρκεια Χρόνου Εκπαίδευσης

Στο διάγραμμα 1 απεικονίζεται με τη μορφή bar chart ο χρόνος που απαιτήθηκε για την εκπαίδευση των μοντέλων της φάσης Ια της διαδικασίας εκπαίδευσης (Πίνακες 9, 10, 11 και 12), όπως αυτή προσδιορίζεται στην υποενότητα 5.1.3 και τοποθετείται χρονικά στην εικόνα 53. Από τη μελέτη του συγκεκριμένου διαγράμματος δεν δύναται

¹¹³ Τα συγκεκριμένα μοντέλα προέκυψαν από τη φάση ΙΙβ της εκπαίδευσης

να εξαχθεί κάποιο ασφαλές συμπέρασμα σχετικά με την ταχύτητα υλοποίησης της εκπαίδευσης κάθε κατηγοριοποιητή. Αυτό οφείλεται σε συνδυασμό ενός ή περισσότερων από τους παρακάτω παράγοντες¹¹⁴:

- Χρησιμοποίηση διαφορετικών παραμέτρων εκπαίδευσης
- Ολοκλήρωση της εκπαίδευσης πριν την επίτευξη του καθοριζόμενου αριθμού σταδίων
- Τερματισμός της εκπαίδευσης λόγω του αριθμού των θετικών εικόνων που δόθηκαν για εκπαίδευση
- Μη καταγραφή της διάρκειας της εκπαίδευσης που ολοκληρώθηκε με μικρό αριθμό σταδίων (μικρότερα των 17 ή 18 κατά περίπτωση)



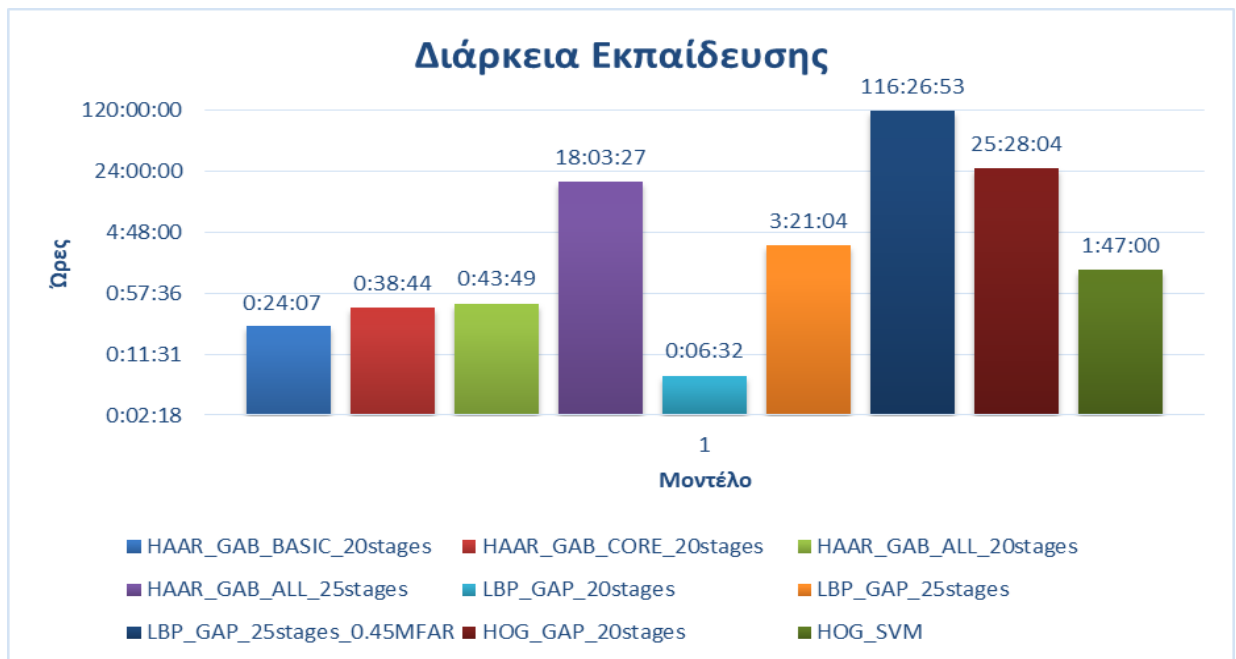
Διάγραμμα 1. Διάρκεια εκπαίδευσης του συνόλου των μοντέλων που εκπαιδεύτηκαν κατά τη Φάση Ια

Η επαναπροσέγγιση της εκπαιδευτικής διαδικασίας (Φάση ΙΙβ, Εικόνα 53) για τους λόγους που αναλύθηκαν στην υποενότητα 5.1.3, οδήγησε στην δημιουργία πιο ομοιογενών μοντέλων (πίνακας 13), αναφορικά με τις τιμές των παραμέτρων εκπαίδευσης (αριθμό σταδίων, τιμές MFAR, MHR, numPos κλπ), γεγονός που προσέδωσε πλέον την δυνατότητα εξαγωγής συμπερασμάτων σχετικά με το χρόνο

¹¹⁴ Λεπτομέρειες σχετικά με την εκπαίδευση, όπως φαίνονται στους πίνακες 10, 11 και 12.

εκπαίδευσή τους¹¹⁵. Συγκεκριμένα, όπως φαίνεται και στο διάγραμμα 2:

- Επαληθεύεται το γεγονός ότι για την εκπαίδευση των Haar-like features (BASIC/20 stages) απαιτείται τουλάχιστον τριπλάσιος χρόνος σε σχέση με τον αντίστοιχο χρόνο εκπαίδευσης των LBP (20 stages).
- Απαιτείται περισσότερος χρόνος για την εκπαίδευση των Haar/CORE και ALL (20 stages) σε σχέση με το BASIC (20 stages) feature set.
- Η εκπαίδευση 25 σταδίων για τα μοντέλα Haar/ALL και LBP απαιτήσε 25 και 30 φορές περισσότερο χρόνο αντίστοιχα, σε σχέση με την εκπαίδευση των 20 σταδίων.
- Η εκπαίδευση 20 σταδίων του HOG/GAB μοντέλου, απαιτήσε μεσοσταθμικά περίπου 40 φορές περισσότερο χρόνο σε σχέση με τα αντίστοιχα Haar-like features 20 σταδίων.
- Μείωση του Max False Alarm Rate κατά 0,05 (0,5→0,45), επέφερε αύξηση του χρόνου εκπαίδευσης του LBP μοντέλου (25 στάδια) περίπου κατά 33 φορές¹¹⁶.



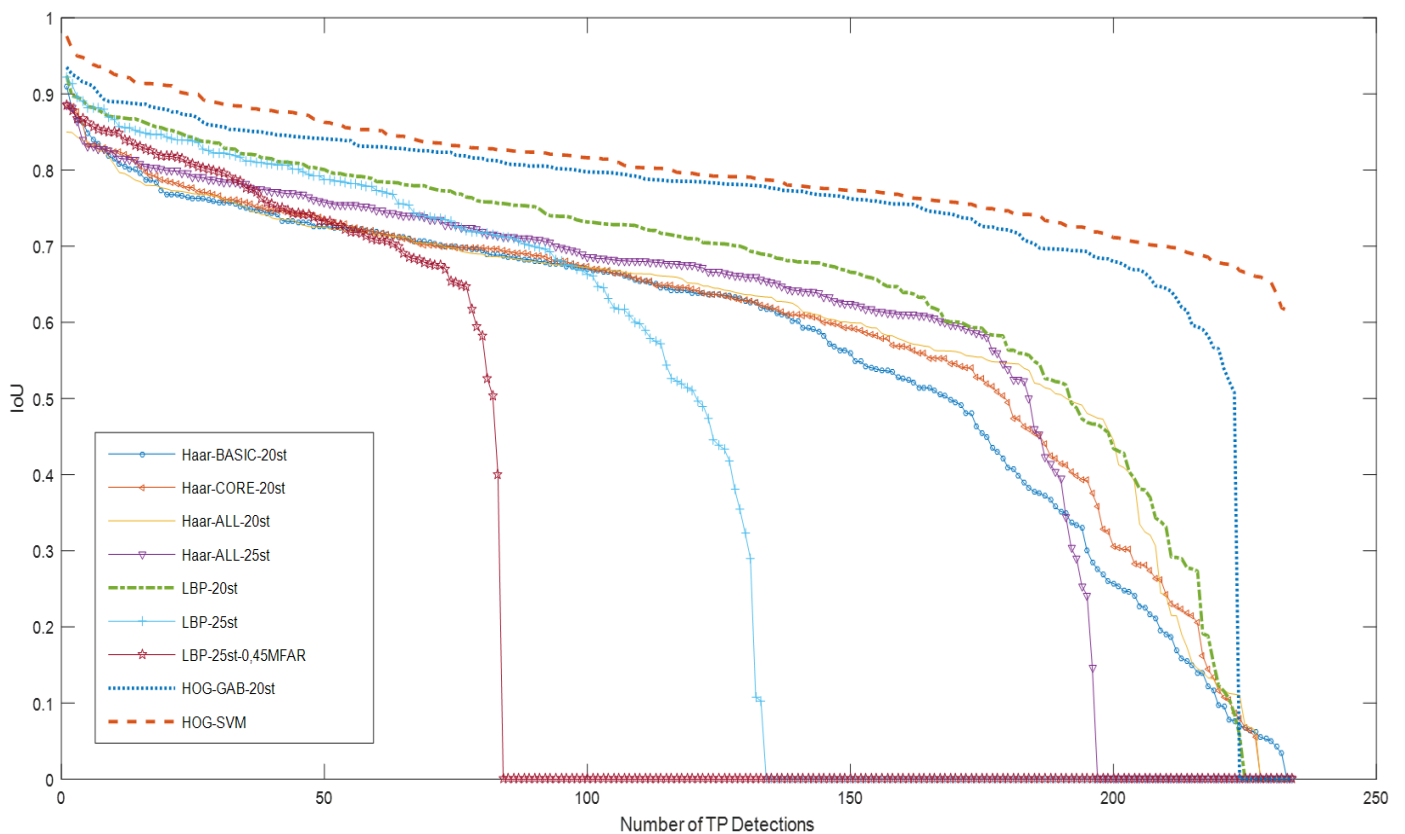
Διάγραμμα 2. Διάρκεια εκπαίδευσης των μοντέλων της Φάσης ΙΙβ (απεικόνιση σε λογαριθμική κλίμακα)

¹¹⁵ Τα συγκεκριμένα 9 μοντέλα που προέκυψαν από τη φάση ΙΙβ της εκπαίδευσης, είναι και αυτά που τέθηκαν στην διαδικασία αξιολόγησης

¹¹⁶ Ανάλογα αύξηση του χρόνου εκπαίδευσης, λόγω μείωσης της τιμής MFAR, παρατηρήθηκε και για την εκπαίδευση 11 σταδίων του Haar/GAB με MFAR=0,3 (πίνακας 8), όπου απαιτήθηκε χρόνος μεγαλύτερος των 10 ημερών.

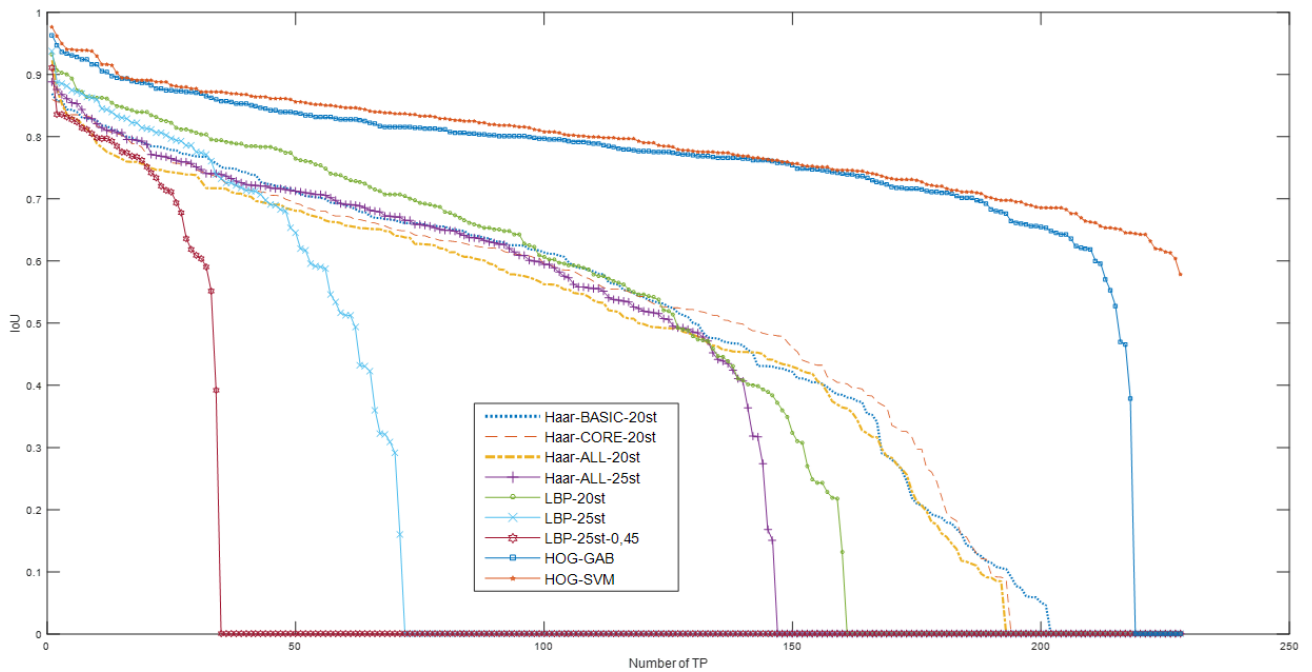
6.1.2 Αξιολόγηση της Ακρίβειας Εντοπισμού (Localization) των Οχημάτων

Ο δείκτης που χρησιμοποιήθηκε για την ακρίβεια του προσδιορισμού της θέσης των οχημάτων στις εικόνες του test set ήταν ο λόγος της τομής προς την ένωση, του ορθογωνίου (bounding box - bb) που εξάγεται από το μοντέλο ανίχνευσης και αυτού του ground truth (Intersection over Union - $IoU = \frac{A \cap B}{A \cup B}$).¹¹⁷ Όσο μεγαλύτερη η τιμή αυτού του λόγου τόσο καλύτερη και η ποιότητα προσδιορισμού της θέσης. Στα διαγράμματα 3 και 4 από τη σύγκριση των τιμών του IoU ($IoU > 0$) που πέτυχε το κάθε μοντέλο για εικόνες διαστάσεων 1920x1080 και 960x540 αντίστοιχα, φαίνεται ότι την καλύτερη απόδοση και στις δύο περιπτώσεις είχαν τα δύο μοντέλα HOG (GAB και SVM). Επίσης, τα μοντέλα Haar που εκπαιδεύτηκαν σε 20 στάδια παρουσιάζουν σε όλο το εύρος τους, μία ομαλότητα στην μείωση της τιμής του IoU. Στον αντίποδα, αυτά που υστέρησαν χαρακτηριστικά ήταν τα LBP 25 σταδίων εκπαίδευσης με τιμές MFAR 0,5 και 0,45, αντίστοιχα.



Διάγραμμα 3. Ακρίβεια προσδιορισμού της θέσης του οχήματος (διαστάσεις εικόνας 1920x1080)

¹¹⁷ A = Bounding box που εξάγεται από το μοντέλο ανίχνευσης, B = Ground Truth Bounding box

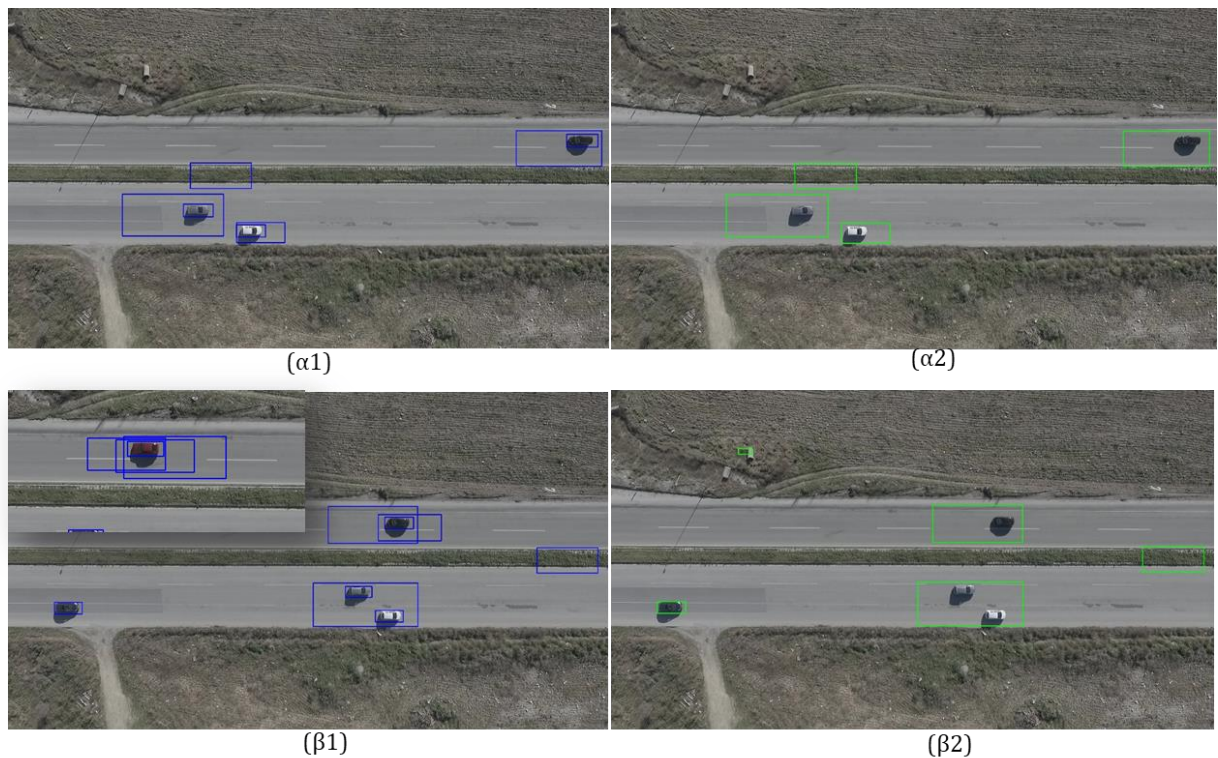


Διάγραμμα 4. Ακρίβεια προσδιορισμού της θέσης του οχήματος (διαστάσεις εικόνας 960x540)

Αν και τα παραπάνω αποτελέσματα επιβεβαιώνονται και οπτικά με μελέτη των αντίστοιχων φωτογραφιών, δείγμα των οποίων παρουσιάζεται στην εικόνα 59, η πραγματική αιτία πίσω από το μέγεθος των bounding boxes είναι οι πολλαπλές FP ανιχνεύσεις, ειδικά για τα Haar/BASIC-CORE-ALL μοντέλα, οι οποίες λόγω της εφαρμογής της τεχνικής Non-Maximum_Suppression, συγχωνευόταν όλα τα bounding boxes σε ένα μεγάλο (εικόνα 60).



Εικόνα 59. Παραδείγματα από το μέγεθος των Bounding Boxes για κάθε ένα από τα μοντέλα.

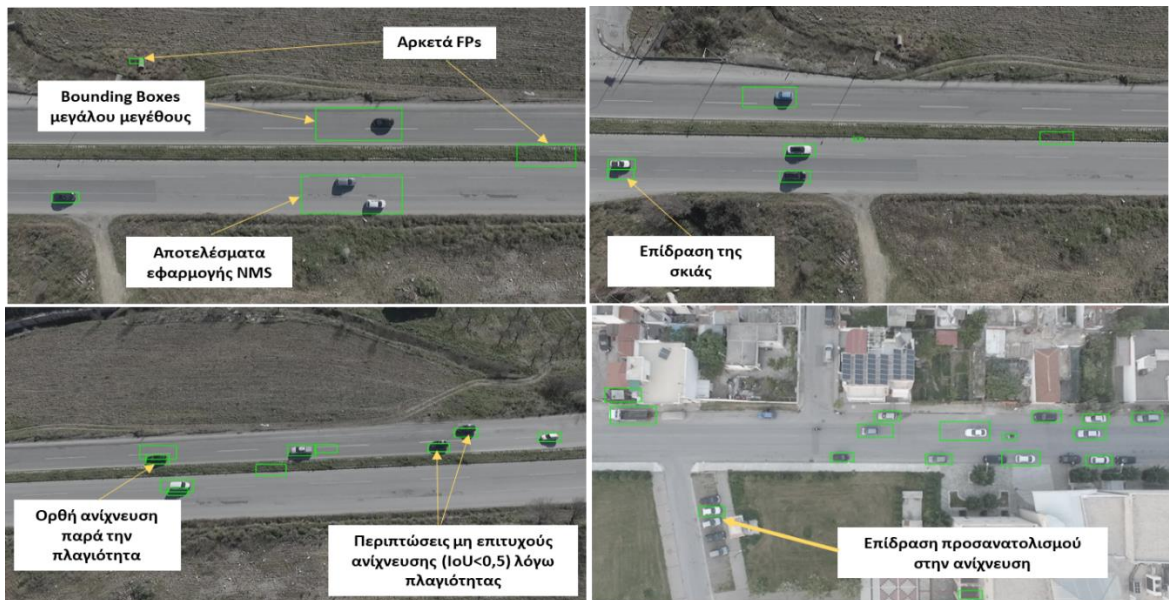


Εικόνα 60. Παράδειγμα εφαρμογής NMS. Στις εικόνες (α1) και (β1) φαίνεται η εφαρμογή του μοντέλου Haar χωρίς τη χρησιμοποίηση του αλγορίθμου NMS (μπλε bb). Στις εικόνες (α2) και (β2) φαίνονται τα αποτελέσματα εφαρμογής του αλγορίθμου NMS (πράσινα bb). Χαρακτηριστική είναι η τοποθέτηση των δύο οχημάτων στην εικόνα (β2) εντός του ίδιου bb. Επίσης στις εικόνες (α1) και (β1) φαίνεται ότι κάθε όχημα περικλείεται από bb με τιμή IoU>0,5.

6.1.3 Ποιοτικά Χαρακτηριστικά της Απόδοσης κάθε Μοντέλου

Haar, Classifier:GAP, Feature Sets : BASIC, Stages : 20, MHR:0,995, MaxFAR:0,5

Το συγκεκριμένο μοντέλο εμφάνισε μεγάλο αριθμό FP. Επίσης λόγω των πολλαπλών bounding boxes που σχηματίζονταν στην εγγύτητα των οχημάτων, αυτά συγχωνεύονταν σε ένα μεγάλο. Υπήρξαν περιπτώσεις όπου λόγω αυτού του φαινομένου βρέθηκαν 2 οχήματα εντός του ίδιου ορθογωνίου. Το μοντέλο επηρεάστηκε σε μεγάλο βαθμό από την σκιά που σχηματιζόταν από τα οχήματα, εμφανίζοντάς την, εσφαλμένα, ως όχημα. Επέδειξε πολύ καλή συμπεριφορά σε ότι αφορά στην ανίχνευση με λήψη 25° ως προς το κατακόρυφο. Σε αυτή την περίπτωση, αν και υπήρξε ορθή ανίχνευση αριθμού οχημάτων, εντούτοις δεν καταχωρήθηκαν ως TP λόγω της χαμηλής τιμής IoU που πέτυχαν (<0,5). Τέλος, χαμηλή ήταν η απόδοση του μοντέλου αναφορικά με την ικανότητά του να ανιχνεύει οχήματα η κατεύθυνση των οποίων (διαμήκης άξονας) σχημάτιζε γωνία σε σχέση με τον διαμήκη άξονα της εικόνας.



Εικόνα 61. Χαρακτηριστικά παραδείγματα της απόδοσης του Haar/BASIC_20stages

Haar, Classifier:GAP, Feature Sets : CORE, Stages : 20, MHR:0,995, MaxFAR:0,5

Γενικά, το συγκεκριμένο μοντέλο «υπέφερε» από τα ίδια προβλήματα, όπως αυτά αναλύθηκαν για την περίπτωση του Haar/BASIC, πλην όμως εμφάνισε ελαφρώς καλύτερη απόδοση. Επίσης, αξίζει να επισημανθεί ότι εμφάνισε την καλύτερη απόδοση, έναντι των υπολοίπων, αναφορικά με την ανίχνευση των οχημάτων σε μη κάθετες λήψεις.



Εικόνα 62. Χαρακτηριστικά παραδείγματα της απόδοσης του Haar/CORE_20stages. (α) Επίδραση της σκιάς (β) Απόδοση σε μη κάθετες λήψεις (γ) Επίδραση του προσανατολισμού των οχημάτων στην ανίχνευση (δ) Δημιουργούμενο bb λόγω εφαρμογής του NMS

Haar, Classifier:GAP, Feature Sets : ALL, Stages : 20, MHR:0,995, MaxFAR:0,5

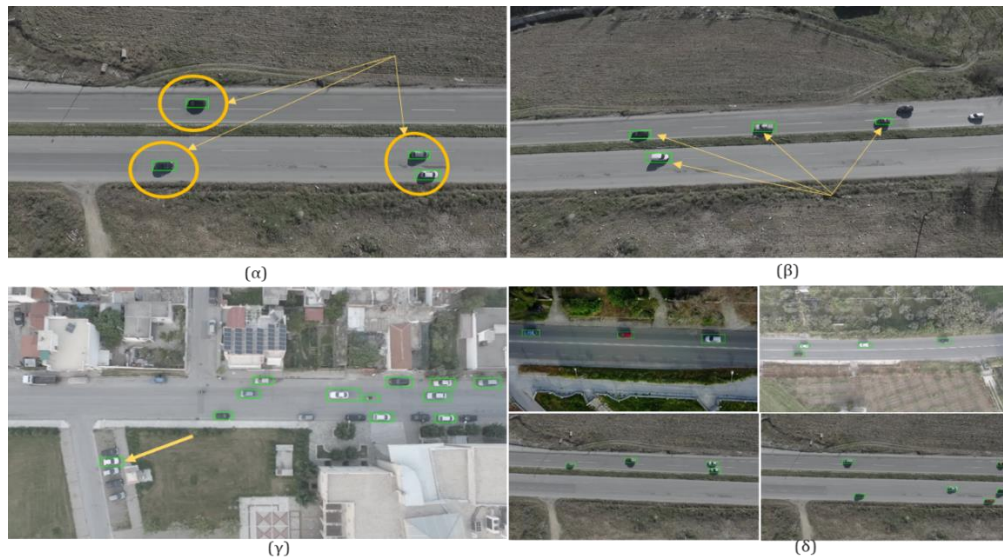
Η συμπεριφορά του, ως προς τα προβλήματα που παρουσίασε (επίδραση σκιάς, εφαρμογή του NMS, τιμή IoU, επίδραση προσανατολισμού), προσομοίωσε σε μεγάλο βαθμό αυτή του Haar/BASIC.



Εικόνα 63. Χαρακτηριστικά παραδείγματα της απόδοσης του Haar/ALL_20stages. (α) Δημιουργούμενο bb λόγω εφαρμογής του NMS (β) Πολύ καλή η απόδοσή του σε μη κάθετες λήψεις (γ) Επίδραση της σκιάς (δ) Επίδραση του προσανατολισμού των οχημάτων στην ανίχνευση

Haar, Classifier:GAP, Feature Sets : ALL, Stages : 25, MHR:0,995, MaxFAR:0,5

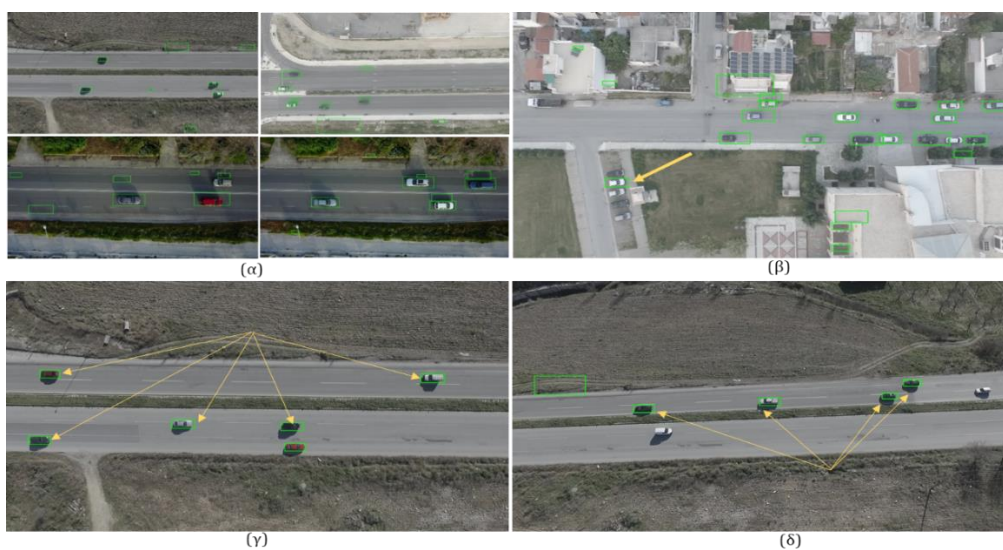
Το συγκεκριμένο μοντέλο εμφάνισε χαρακτηριστική διαφορά σε σχέση με τα προηγούμενα. Από τις εικόνες φαίνεται ότι σε ελάχιστες περιπτώσεις απαιτήθηκε η εφαρμογή του NMS με αποτέλεσμα τα bounding boxes να είναι μικρά σε μέγεθος. Επίσης, μειώθηκαν σε μεγάλο βαθμό οι περιπτώσεις αναγνώρισης της σκιάς ως TP όχημα. Κατά τα λοιπά, εμφάνισε τα ίδια προβλήματα στην ανίχνευση οχημάτων ο προσανατολισμός των οποίων ήταν στα όρια των δυνατοτήτων αναγνώρισης του αλγόριθμου ($\pm 10^\circ$), ενώ πολύ καλή ήταν και η απόδοσή του στην αναγνώριση των οχημάτων σε πλευρικές λήψεις.



Εικόνα 64. Χαρακτηριστικά παραδείγματα της απόδοσης του Haar/ALL_25stages. (α) Μεγαλύτερη ακρίβεια στα bb (β) Πολύ καλή η απόδοσή του σε μη κάθετες λήψεις (γ) Επίδραση του προσανατολισμού των οχημάτων στην ανίχνευση (δ) Πολύ μικρή επίδραση της σκιάς

LBP, Classifier:GAP, Stages : 20, MHR:0,995, MaxFAR:0,5

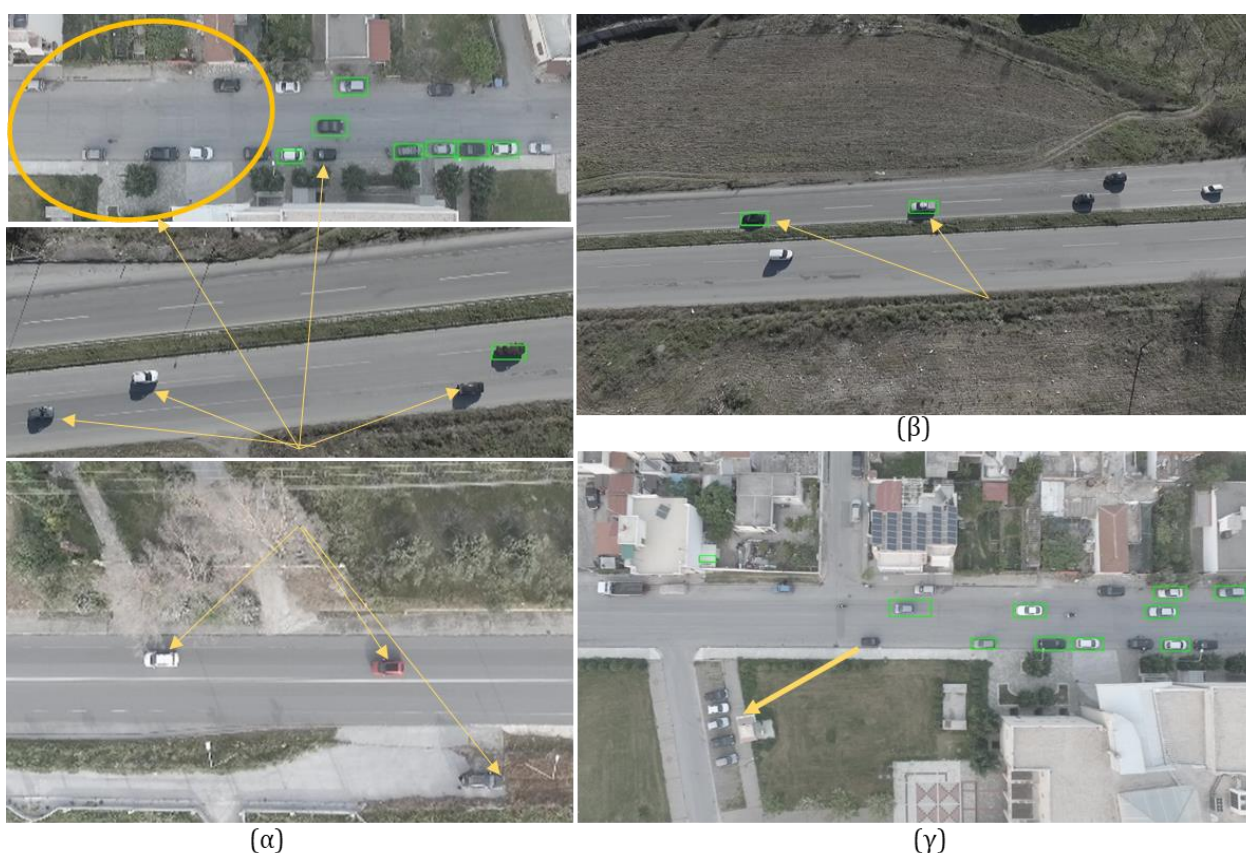
Από την ποιοτική μελέτη των εικόνων του εν λόγω μοντέλου το πλέον χαρακτηριστικό σημείο είναι αυτό που μεγάλοι αριθμού εσφαλμένων θετικών κατηγοριοποιήσεων (FPs). Από την άλλη, τα bounding boxes, για τις TP αναγνώρισεις ήταν αρκετά ακριβή. Η απόδοση στην ανίχνευση οχημάτων από πλευρικές λήψεις, ήταν ανάλογη του Haar/ALL_25stages.



Εικόνα 65. Χαρακτηριστικά παραδείγματα της απόδοσης του LBP_20stages. (α) Μεγάλος αριθμός FP (β) Επίδραση του προσανατολισμού των οχημάτων στην ανίχνευση (γ) Ακρίβεια των bb (δ) Πολύ καλή η απόδοσή του σε μη κάθετες λήψεις

LBP, Classifier:GAP, Stages : 25, MHR:0,995, MaxFAR:0,5

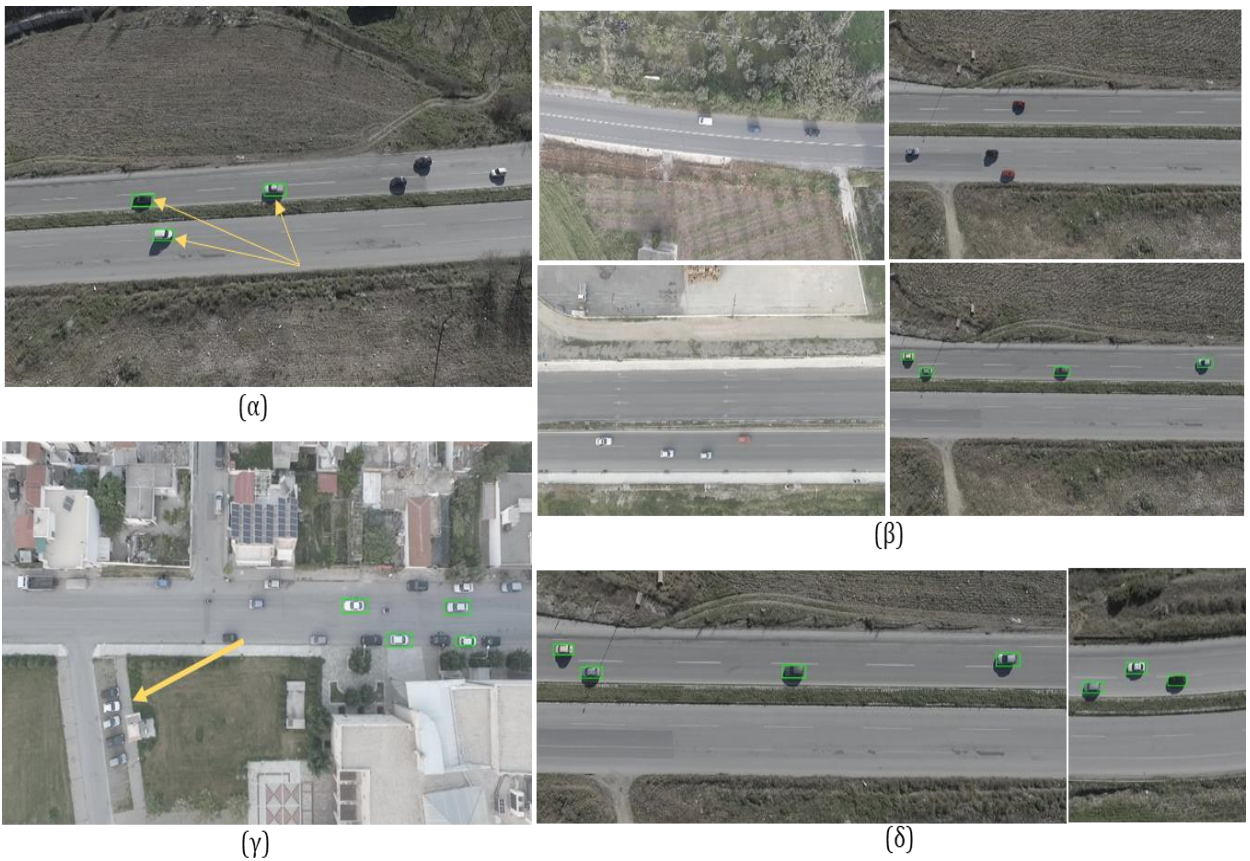
Στο συγκεκριμένο μοντέλο, αν και μειώθηκε ο αριθμός των FP, εντούτοις, μείωση παρουσίασε και η τιμή των TP. Επίσης, χαμηλή ήταν η απόδοσή του στην περίπτωση των λήψεων με γωνία ως προς την καθετότητα. Στο ζήτημα του προσανατολισμού των οχημάτων, εκτιμάται ότι δεν δύναται να εξαχθούν ασφαλή συμπεράσματα ως προς τις δυνατότητες του μοντέλου καθόσον, η αδυναμία ανίχνευσης έστω του ενός από τα πέντε οχήματα που βρίσκονταν υπό γωνία $\approx 10^\circ$ (επετεύχθη από τα προηγούμενα μοντέλα), πιθανώς εντάσσεται στο γενικότερο θέμα εμφάνισης χαμηλής τιμής TP ανιχνεύσεων.



Εικόνα 66. Χαρακτηριστικά παραδείγματα της απόδοσης του LBP_25stages. (α) Μείωση του αριθμού των FP και ταυτόχρονη μείωση των TP (β) Χαμηλή απόδοσή σε μη κάθετες λήψεις (γ) Η αδυναμία ανίχνευσης των συγκεκριμένων οχημάτων πιθανώς οφείλεται στην χαμηλή απόδοση της ανίχνευσης, παρά στον προσανατολισμό τους.

LBP, Classifier:GAP, Stages : 25, MHR:0,995, MaxFAR:0,45

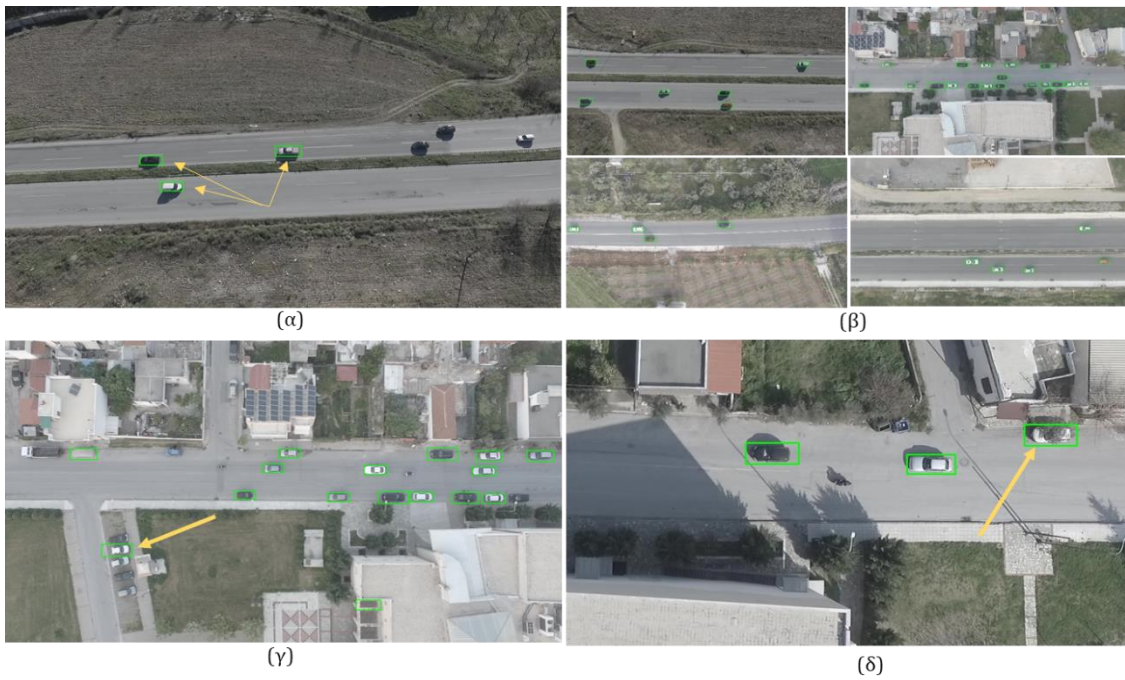
Παρουσίασε ανάλογα χαρακτηριστικά με αυτά του LBP_25stages. Η σημαντικότερη διαφοροποίηση μεταξύ τους έγκειται στο μηδενικό αριθμό FP και στην περαιτέρω μείωση της τιμής των TP ανιχνεύσεων.



Εικόνα 67. Χαρακτηριστικά παραδείγματα της απόδοσης του LBP_25stages_0,45MFAR. (α) Μέτρια απόδοσή σε μη κάθετες λήψεις (β) Ο αριθμός των TP ήταν αρκετά μικρός ενώ δεν παρουσίασε κανένα FP (γ) Η αδυναμία ανίχνευσης των συγκεκριμένων οχημάτων πιθανώς οφείλεται στην χαμηλή απόδοση της ανίχνευσης, παρά στον προσανατολισμό τους.

HOG, Classifier:GAP, Stages : 20, MHR:0,995, MaxFAR:0,5

Το συγκεκριμένο μοντέλο, σε σχέση με τα προηγούμενα, παρουσίασε θεαματική βελτίωση σε ό,τι αφορά στις TP και FP ανιχνεύσεις. Επίσης, βελτιωμένη ήταν και η απόδοσή του ως προς την τιμή του IoU. Επιπρόσθετα, η σκιά των οχημάτων δεν προκάλεσε FP ανιχνεύσεις, ενώ αναφορικά με το ζήτημα της επίδρασης του προσανατολισμού στην ανίχνευσή τους, δεν τα κατάφερε καλύτερα από τα μοντέλα Haar ή το LBP_20stages. Βαρύνουσας σημασίας είναι το γεγονός ανίχνευσης οχήματος, μέρος του οποίου αποκρύπτονταν από το φυσικό περιβάλλον. Τέλος, το μοναδικό σημείο στο οποίο επέδειξε μέτρια επίδοση, αφορούσε την ανίχνευση των οχημάτων σε λήψεις από πλάγια θέση.



Εικόνα 68. Χαρακτηριστικά παραδείγματα της απόδοσης του HOG_GAB_20stages. (α) Μέτρια απόδοσή σε μη κάθετες λήψεις και κανένα FP λόγω σκιάς (β) Υψηλό ποσοστό TP ανιχνεύσεων (γ) Επηρεάστηκε και το συγκεκριμένο μοντέλο από τον προσανατολισμό των οχημάτων (δ) Ανίχνευση παρά τη μερική απόκρυψη του οχήματος.

HOG, Classifier:SVM



Εικόνα 69. Χαρακτηριστικά παραδείγματα της απόδοσης του HOG_SVM. (α) Καλή απόδοσή σε μη κάθετες λήψεις και κανένα FP λόγω σκιάς (β) Πολύ υψηλό ποσοστό TP ανιχνεύσεων (γ) Καλύτερη απόδοση, συγκριτικά με τα υπόλοιπα μοντέλα στο ζήτημα του προσανατολισμού των οχημάτων (δ) Επιτυχής και ανεπιτυχής ανίχνευση οχήματος σε απόκρυψη.

Αποτελεί το μοντέλο με την καλύτερη απόδοση. Χαρακτηριστικά σημεία είναι η ανίχνευση των 3 από τα 5 οχήματα με προσανατολισμό $\cong 10^\circ$ ως προς το οριζόντιο. Επιπλέον ανίχνευσε οχήματα μέρος των οποίων καλύπτονταν από φυσικά εμπόδια. Το μοναδικό σημείο στο οποίο δεν επέδειξε βελτίωση (αντιθέτως παρουσίασε μειωμένη απόδοση), ήταν αυτό της ανίχνευσης οχημάτων σε λήψεις με γωνία 25° ως προς την κατακόρυφο άξονα.



Εικόνα 70. Ανίχνευση σε συνθήκες χαμηλού φωτισμού

Τα περισσότερα μοντέλα επέδειξαν αντοχή στην ανίχνευση σε συνθήκες χαμηλού φωτισμού. Η συγκεκριμένη περίπτωση αποτελεί ιδιαίτερη πρόκληση όταν συνδυάζεται με κίνηση των οχημάτων οπότε η ποιότητα ειδώλου των οχημάτων μειώνεται αντιστρόφως ανάλογα με την ταχύτητά τους. Τέλος ανάλογη ευρωστία εμφάνισαν τα μοντέλα και στις περιπτώσεις αλλαγής της διάστασης (scale) των οχημάτων (μεγαλύτερη διάσταση σε σχέση με αυτή που χρησιμοποιήθηκε κατά την εκπαίδευση).



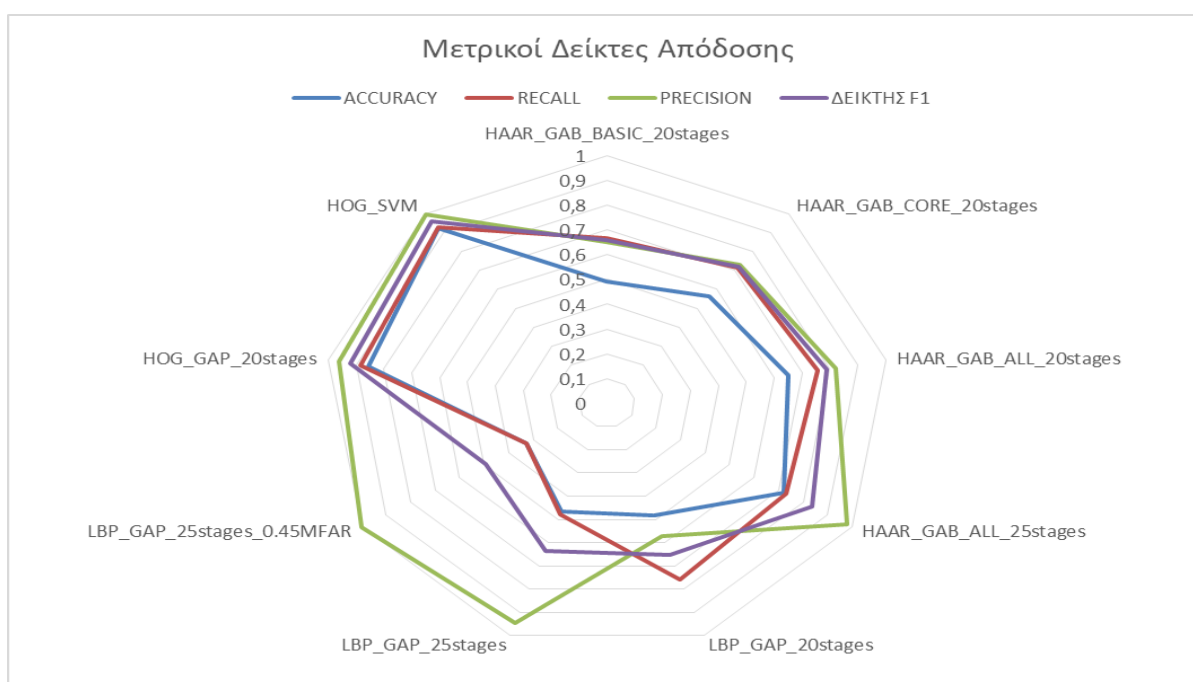
Εικόνα 71. Ανίχνευση οχημάτων σε διαφορετική κλίμακα

6.1.4 Ποσοτική Σύγκριση της Απόδοσης κάθε Μοντέλου Χρησιμοποιώντας Μετρικούς Δείκτες

Τα 9 μοντέλα ανίχνευσης αξιολογήθηκαν με τους μετρικούς δείκτες που αναλύθηκαν στο 2ο κεφάλαιο, και πιο συγκεκριμένα στην συνολική ορθότητα (accuracy), στο ρυθμό σφάλματος (error rate), στην ακρίβεια (precision) στον αριθμό εύστοχων αποκρίσεων (recall) και στο δείκτη F1. Στη συνέχεια παρατίθενται οι πίνακες τιμών και τα αντίστοιχα διαγράμματα για τις δύο διαφορετικές διαστάσεις εικόνων που χρησιμοποιήθηκαν για την αξιολόγησή τους (1920x1080 και 960x540).

ΜΟΝΤΕΛΟ	ACCURACY	ERROR RATE	RECALL	PRECISION	ΔΕΙΚΤΗΣ F1
HAAR_GAB_BASIC_20stages	0,49123	0,50877	0,66667	0,65116	0,65882
HAAR_GAB_CORE_20stages	0,56467	0,43533	0,716	0,72764	0,72177
HAAR_GAB_ALL_20stages	0,64966	0,35034	0,75794	0,81974	0,78763
HAAR_GAB_ALL_25stages	0,72047	0,27953	0,732	0,97861	0,83753
LBP_GAP_20stages	0,48354	0,51646	0,75794	0,57186	0,65188
LBP_GAP_25stages	0,46512	0,53488	0,47809	0,94488	0,63492
LBP_GAP_25stages_0.45MFAR	0,32669	0,67331	0,32669	1,0	0,49249
HOG_GAP_20stages	0,85441	0,14559	0,88492	0,96121	0,92149
HOG_SVM	0,9249	0,0751	0,92857	0,99574	0,96099

Πίνακας 18. Μετρήσεις δεικτών απόδοσης των μοντέλων (ανίχνευση οχημάτων στις εικόνες του test set διαστάσεων 1920x1080 της VERVwUAS DB).



Διάγραμμα 5. Μετρήσεις δεικτών απόδοσης των μοντέλων (ανίχνευση οχημάτων στις εικόνες του test set διαστάσεων 1920x1080 της VERVwUAS DB).

Από τη μελέτη των πινάκων 17 και 18 διαπιστώνεται ότι:

- Από τα μοντέλα Haar 20 σταδίων, αυτό με το feature set ALL επέδειξε την καλύτερη συμπεριφορά καθόσον είχε το μεγαλύτερο αριθμό TP και ταυτόχρονα τις λιγότερες FP και FN ανιχνεύσεις.
- Αναφορικά με τα μοντέλα LBP το καθένα επέδειξε διαφορετικές αδυναμίες. Συγκεκριμένα, το LBP_20 stages είχε τον μεγαλύτερο αριθμό FP (1,5 φορές περισσότερα από το δεύτερο χειρότερο). Τα LBP_25stages_0,45MFAR και LBP_25stages είχαν μακράν τη χειρότερη απόδοση σε ότι αφορά στα FN.
- Τα μοντέλα HOG (GAB και SVM) παρουσίασαν τις καλύτερες τιμές αναφορικά με τα TP, FP και FN. Αυτή η συμπεριφορά απεικονίζεται και στους μετρικούς δείκτες όπου ο δείκτης F1 κυμάνθηκε άνω του 0,9.

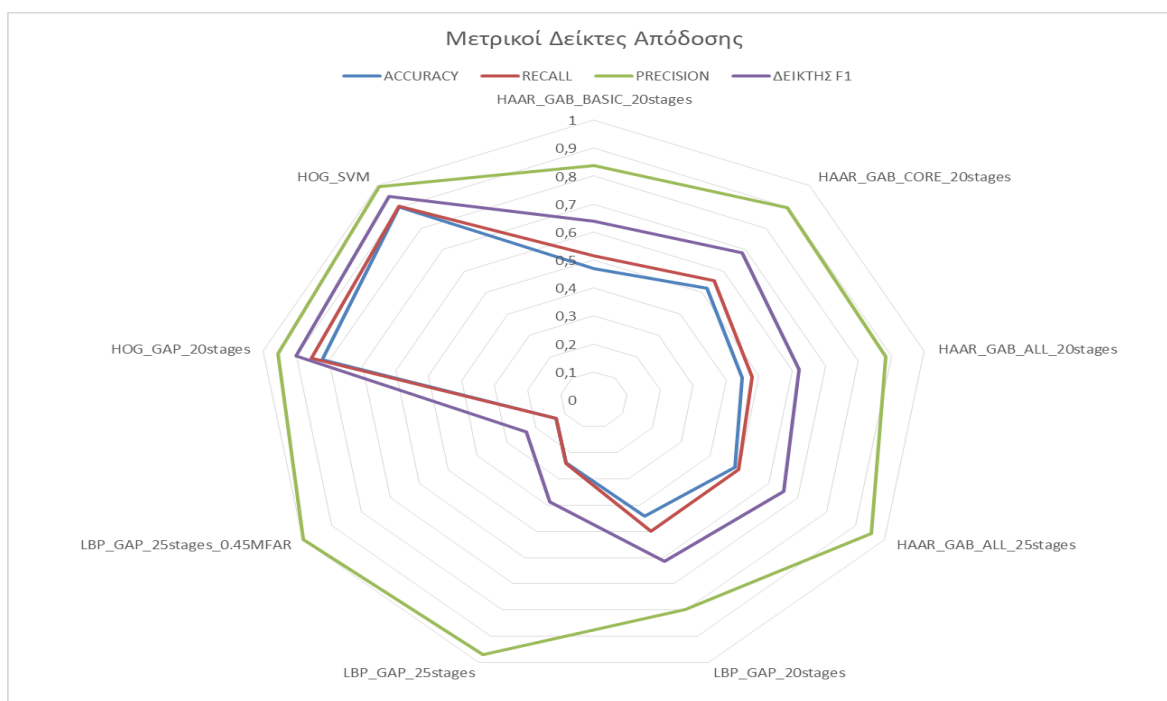
ΜΟΝΤΕΛΟ	TP	FP	FN (λόγω IoU<0,5)	FN (λόγω αδυναμίας ανίχνευσης)	Σύνολο FN
HAAR_GAB_BASIC_20stages	168	90	64	20	84
HAAR_GAB_CORE_20stages	179	67	48	23	71
HAAR_GAB_ALL_20stages	191	42	36	25	61
HAAR_GAB_ALL_25stages	183	4	13	54	67
LBP_GAP_20stages	191	143	33	28	61
LBP_GAP_25stages	120	7	13	118	131
LBP_GAP_25stages_0.45	82	0	1	168	169
HOG_GAP_20stages	223	9	0	29	29
HOG_SVM	234	1	0	18	18
Υπόμνημα	Π>200 200>ΚΤ>150 ΚΚ<150	Π<20 20<ΚΤ<50 ΚΚ>50	Π<20 20<ΚΤ<50 ΚΚ>50	Π<20 20<ΚΤ<50 ΚΚ>50	Π<40 40<ΚΤ<100 ΚΚ>100

Πίνακας 19. Σύγκριση των τιμών TP, FP και FN μεταξύ των μοντέλων (εικόνες του test set διαστάσεων 1920x1080 της VERVwUAS DB)

- Η τιμή του δείκτη F1 του μοντέλου Haar_ALL_25stages ήταν καλύτερη από αυτή του Haar_ALL_20stages. Αυτό οφείλεται στο γεγονός ότι το πρώτο είχε πολύ μικρό αριθμό FP και ελαφρώς μεγαλύτερη συνολική τιμή FN σε σχέση με το δεύτερο. Μία όμως πιο αναλυτική παρατήρηση του πίνακα 18 δείχνει ότι το μοντέλο των 20 σταδίων εκπαίδευσης είχε περισσότερες ανιχνεύσεις οχημάτων από το Haar_ALL_25stages οι οποίες όμως, επειδή δεν πέρασαν το κατώφλι της τιμής του IoU (0,5), θεωρήθηκαν ως FN. Επίσης, το μοντέλο Haar_ALL_20stages είχε σχεδόν το μισό αριθμό FN (λόγω αδυναμίας ανίχνευσης), σε σχέση με το Haar_ALL_25stages.

ΜΟΝΤΕΛΟ	ACCURACY	ERROR RATE	RECALL	PRECISION	ΔΕΙΚΤΗΣ F1
HAAR_GAB_BASIC_20stages	0,46931	0,53069	0,51587	0,83871	0,63882
HAAR_GAB_CORE_20stages	0,52256	0,47744	0,556	0,89677	0,68642
HAAR_GAB_ALL_20stages	0,44906	0,55094	0,47791	0,88148	0,61979
HAAR_GAB_ALL_25stages	0,4845	0,5155	0,49603	0,9542	0,65274
LBP_GAP_20stages	0,44366	0,55634	0,5	0,79747	0,61463
LBP_GAP_25stages	0,24016	0,75984	0,24206	0,96825	0,3873
LBP_GAP_25stages_0.45MFAR	0,13095	0,86905	0,13095	1,0	0,23158
HOG_GAP_20stages ¹¹⁸	0,82061	0,17939	0,85317	0,95556	0,90147
HOG_SVM ¹¹⁹	0,90119	0,09881	0,90476	0,99563	0,94802

Πίνακας 20. Μετρήσεις δεικτών απόδοσης των μοντέλων (ανίχνευση οχημάτων στις εικόνες του test set διαστάσεων 960x540 της VERVwUAS DB).



Διάγραμμα 6. Μετρήσεις δεικτών απόδοσης των μοντέλων (ανίχνευση οχημάτων στις εικόνες του test set διαστάσεων 960x540 της VERVwUAS DB).

¹¹⁸ Κατά την εισαγωγή των εικόνων στον αλγόριθμο επεξεργασίας, τροποποιήθηκαν οι διαστάσεις τους από 960x540 σε 1920x1080. Υπενθυμίζεται ότι η εκπαίδευση του αλγορίθμου έγινε στο παράθυρο 58x24 που σχετίζεται με διάσταση εικόνας 1920x1080.

¹¹⁹ Κατά την εισαγωγή των εικόνων στον αλγόριθμο επεξεργασίας, τροποποιήθηκαν οι διαστάσεις τους από 960x540 σε 1280x720. Αν και η εκπαίδευση έγινε με annotation εικόνων διαστάσεων 960x540 το μοντέλο άρχισε να αποδίδει σε ανίχνευση εικόνων διαστάσεων 1280x720.

Από τη μελέτη των πινάκων 19 και 20 διαπιστώνεται ότι:

- Από τα μοντέλα Haar 20 σταδίων, αυτή το φορά το feature set CORE ήταν αυτό που επέδειξε την καλύτερη συμπεριφορά καθόσον είχε το μεγαλύτερο αριθμό TP και τα λιγότερα FP και FN.
- Αναφορικά με τα μοντέλα LBP το καθένα, όπως και στην περίπτωση των εικόνων διαστάσεων 1920x1080 επέδειξε διαφορετικές αδυναμίες. Συγκεκριμένα, το LBP_20 stages είχε τον μεγαλύτερο αριθμό FP. Τα LBP_25stages_0,45MFAR και LBP_25stages είχαν, και σε αυτή την περίπτωση, τη χειρότερη απόδοση σε ότι αφορά στα FN.
- Τα μοντέλα HOG (GAB και SVM) παρουσίασαν ξανά τις καλύτερες τιμές αναφορικά με τα TP, FP και FN, ενώ ο F1 διατηρήθηκε άνω του 0,9.

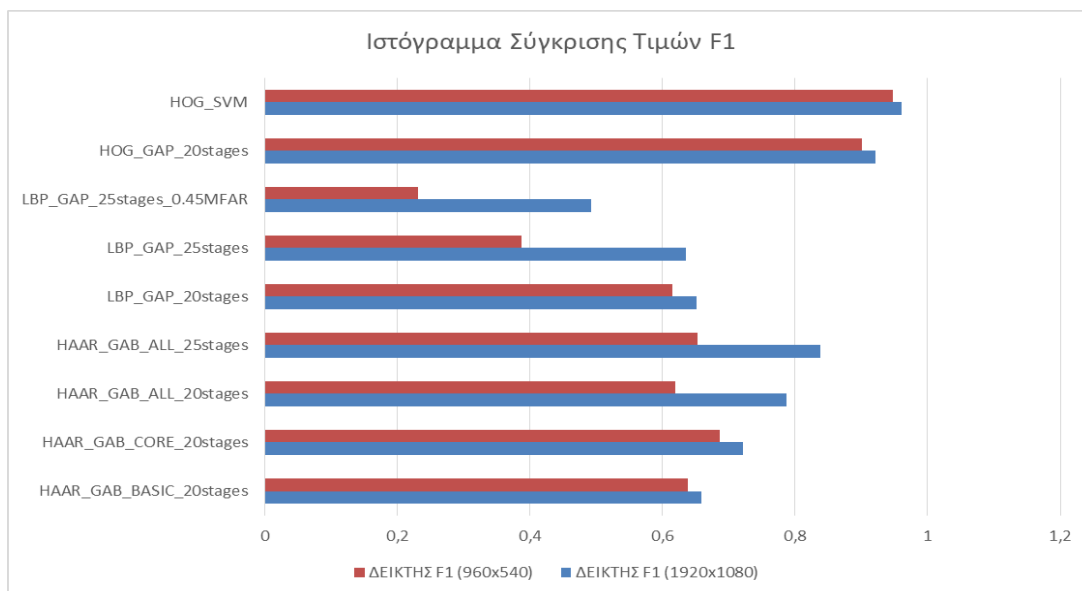
ΜΟΝΤΕΛΟ	TP	FP	FN (λόγω IoU<0,5)	FN (λόγω αδυναμίας ανίχνευσης)	Σύνολο FN
HAAR_GAB_BASIC_20stages	130	25	71	51	122
HAAR_GAB_CORE_20stages	139	16	54	57	111
HAAR_GAB_ALL_20stages	119	16	73	57	130
HAAR_GAB_ALL_25stages	125	6	21	106	127
LBP_GAP_20stages	126	32	34	92	126
LBP_GAP_25stages	61	2	10	181	191
LBP_GAP_25stages_0.45	33	0	1	218	219
HOG_GAP_20stages	215	10	3	34	37
HOG_SVM	228	1	0	24	24
Υπόμνημα	Π>200 200>ΚΤ>150 ΚΚ<150	Π<20 20<ΚΤ<50 ΚΚ>50	Π<20 20<ΚΤ<50 ΚΚ>50	Π<20 20<ΚΤ<50 ΚΚ>50	Π<40 40<ΚΤ<100 ΚΚ>100

Πίνακας 21. Σύγκριση των τιμών TP, FP και FN μεταξύ των μοντέλων (εικόνες 960x540 του test set της VERVwUAS DB)

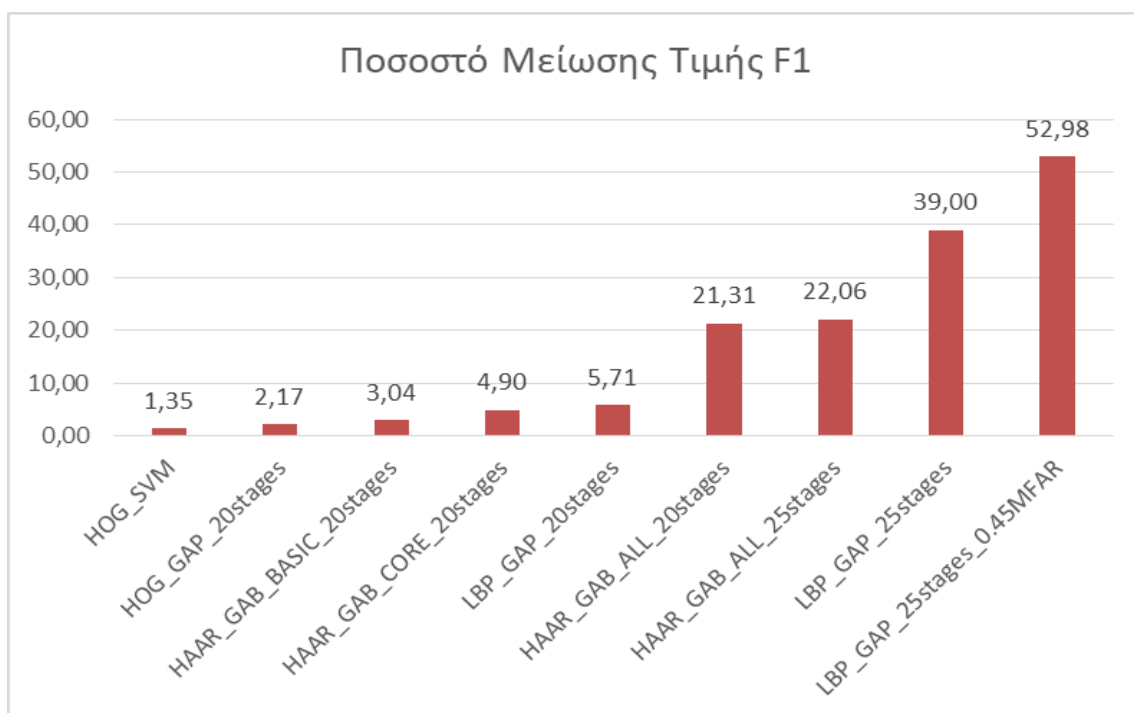
- Σε αυτή την περίπτωση, η τιμή του δείκτη F1 ενός μοντέλου 20 σταδίων εκπαίδευσης (Haar_CORE) επικράτησε έναντι αυτής των 25 σταδίων (Haar_ALL). Αυτό οφείλεται στο γεγονός ότι το πρώτο είχε μεγαλύτερο αριθμό TP και μικρότερο αριθμό συνολικών FN. Αξίζει επίσης προσοχής και το γεγονός ότι: (1) το Haar_CORE είχε σχεδόν 2,5 φορές περισσότερες ανιχνεύσεις οι οποίες όμως θεωρήθηκαν ως FN επειδή δεν πέρασαν το κατώφλι της τιμής του IoU (0,5) και (2) το μοντέλο Haar_CORE είχε σχεδόν το μισό αριθμό FN (λόγω αδυναμίας ανίχνευσης), σε σχέση με το Haar_ALL_25stages.

Μελετώντας τα διαγράμματα 7 και 8, αφενός μεν συγκρίνεται η απόδοση των μοντέλων ανίχνευσης (τιμή F1) μεταξύ εικόνων διαφορετικών διαστάσεων, αφετέρου δε

διακρίνεται και ο βαθμός στον οποίο επηρεάστηκε το κάθε ένα από αυτά. Έτσι, εξαιρουμένου του HOG_SVM, καθίσταται εμφανές ότι τα cascade μοντέλα 20 σταδίων εκπαίδευσης, επηρεάστηκαν πολύ λιγότερο από τα αντίστοιχα των 25 σταδίων, στην ανίχνευση οχημάτων σε εικόνες μικρότερης διάστασης.



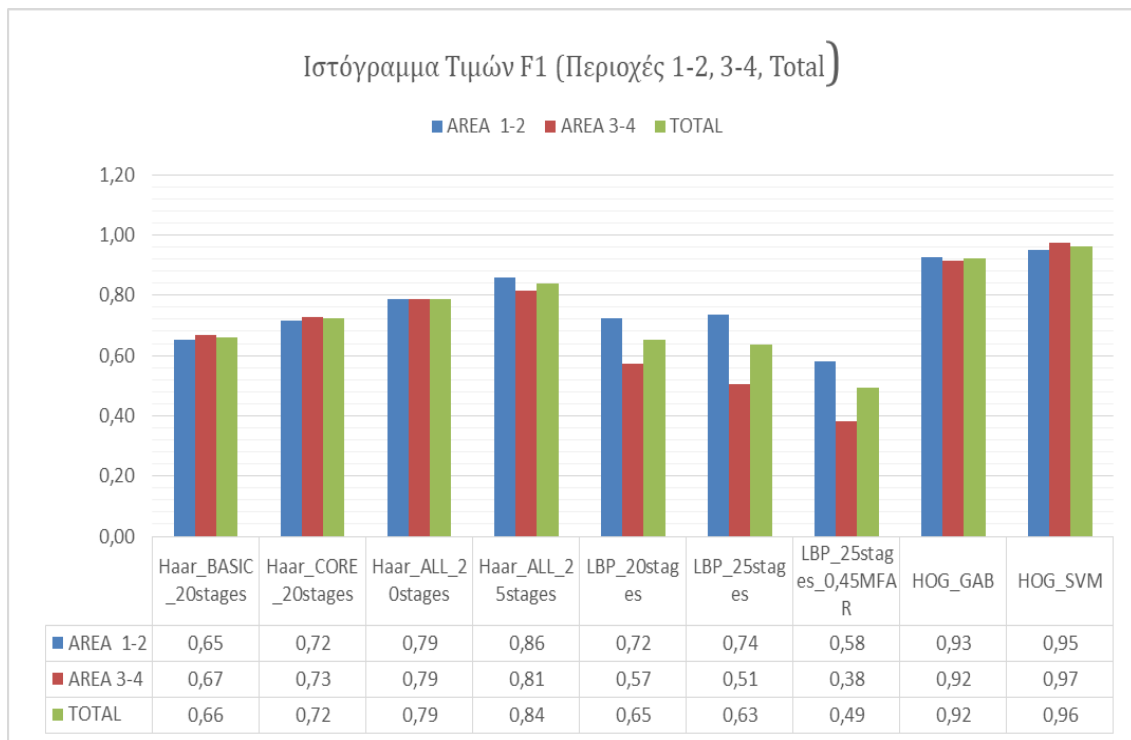
Διάγραμμα 7. Ιστόγραμμα σύγκρισης τιμών F1



Διάγραμμα 8. Ιστόγραμμα απεικόνισης του ποσοστού μείωσης της τιμής F1 (από εικόνες διαστάσεων 1920x1080 σε 960x540, του test set της VERVwUAS DB)

6.1.5 Σύγκριση Απόδοσης κάθε Μοντέλου μεταξύ των Περιοχών 1-2 και 3-4¹²⁰

Στο διάγραμμα του ακολουθεί συγκρίνεται η απόδοση κάθε μοντέλου (δείκτης F1) ξεχωριστά, για τις περιοχές 1-2 και 3-4. Μέσω αυτής της σύγκρισης είναι δυνατή η εξαγωγή συμπερασμάτων σχετικά με το πόσο προσαρμοσμένο είναι το κάθε μοντέλο, στο δείγμα εκπαίδευσης (εικόνες περιοχών 1-2). Η σύγκριση έγινε με εκμετάλλευση των αποτελεσμάτων που προέκυψαν από την εφαρμογή των μοντέλων στις εικόνες διαστάσεων 1080X960, καθόσον σε αυτές στηρίχθηκε η διαδικασία της εκπαίδευσης.



Διάγραμμα 9. Ιστόγραμμα τιμών F1 στις περιοχές 1-2, 3-4 και στο συνολικό δείγμα

Στο διάγραμμα διακρίνεται ότι για τα μοντέλα Haar_BASIC/CORE/ALL 20 σταδίων εκπαίδευσης και HOG_GAB/SVM, η τιμή του δείκτη F1 είναι ανεξάρτητη της περιοχής στην οποία πραγματοποιείται η ανίχνευση. Για τα μοντέλα 25 σταδίων εκπαίδευσης, παρατηρείται ότι επηρεάζεται η απόδοσή τους από την περιοχή στην οποία δοκιμάζονται. Αυτή η επίδραση φαίνεται να λιγότερο σημαντική για το Haar_ALL_25stages και πολύ μεγαλύτερη για τα μοντέλα LBP feature. Τα τελευταία επηρεάζονται περισσότερο και ως feature type καθόσον διαφορά στις τιμές F1 δεν επέδειξαν μόνο τα μοντέλα των 25 σταδίων αλλά και το LBP_20stages.

¹²⁰ Η περιγραφή των περιοχών έγινε στην υποενότητα 4.2.2 όπου υπάρχει αντίστοιχα χάρτης και φωτογραφικό υλικό (εικόνα 54)

6.1.6 Χρόνος Ανίχνευσης

Στη συνέχεια εξετάστηκε ο παράγοντας χρόνος και πως αυτός επηρεάζει το κάθε μοντέλο. Στον πίνακα 21 φαίνεται η χρονική διάρκεια επεξεργασίας του test set των 54 εικόνων¹²¹. Η πλέον εμφανής διαπίστωση είναι αυτή της μείωσης του απαιτούμενου χρόνου επεξεργασίας για τις εικόνες διαστάσεων 960x540. Ο χρόνος¹²² μειώθηκε μεσοσταθμικά κατά έναν παράγοντα 4 ο οποίος εκφράζει ταυτόχρονα και τον λόγο μεταξύ των μεγεθών των εικόνων. Αυτή η παρατήρηση δεν ισχύει για τα μοντέλα HOG_GAB/SVM καθόσον, όπως έχει ήδη αναφερθεί, τροποποιούνταν οι διαστάσεις των εικόνων 960x540, πριν την εισαγωγή τους στον αλγόριθμο ανίχνευσης.

MONTELO	ΧΡΟΝΟΣ ΑΝΙΧΝΕΥΣΗΣ (sec)	
	1920x1080	960x540
HAAR_GAB_BASIC_20stages	47.104	11.721
HAAR_GAB_CORE_20stages	42.443	10.591
HAAR_GAB_ALL_20stages	41.142	11.100
HAAR_GAB_ALL_25stages	38.292	9.672
LBP_GAP_20stages	36.952	11.361
LBP_GAP_25stages	33.342	9.530
LBP_GAP_25stages_0.45MFAR	26.781	8.170
	1920x1080	960x540 → 1920x1080
HOG_GAP_20stages	55.023	52.781
	1920x1080	960x540 → 1280x720
HOG_SVM	168.236	78.296

Πίνακας 22. Διάρκεια χρόνου επεξεργασίας του test set εικόνες του test set της VERVwUAS DB

Κατά την επεξεργασία των εικόνων διαπιστώθηκε ότι τα μοντέλα δαπανούσαν περισσότερο χρόνο σε φωτογραφίες οι οποίες περιλάμβαναν μεγαλύτερο αριθμό οχημάτων και ως εκ τούτου πραγματοποιούνταν περισσότερες ανιχνεύσεις. Χαρακτηριστικά παραδείγματα είναι οι εικόνες 15 και 16 του test set της VERVwUAS DB, οι οποίες απεικονίζουν 20 και 21 οχήματα αντίστοιχα (οριζόντιου προσανατολισμού). Λαμβάνοντας το παραπάνω δεδομένο, διαβιβάστηκαν στα μοντέλα

¹²¹ Σε αυτούς τους χρόνους περιλαμβάνεται και ο χρόνος αποθήκευσης των εικόνων με τα bounding boxes από την ανίχνευση. Εφόσον εξαληφθεί αυτή η απαίτηση, τότε οι εν λόγω τιμές μειώνονται κατά 15-20%.

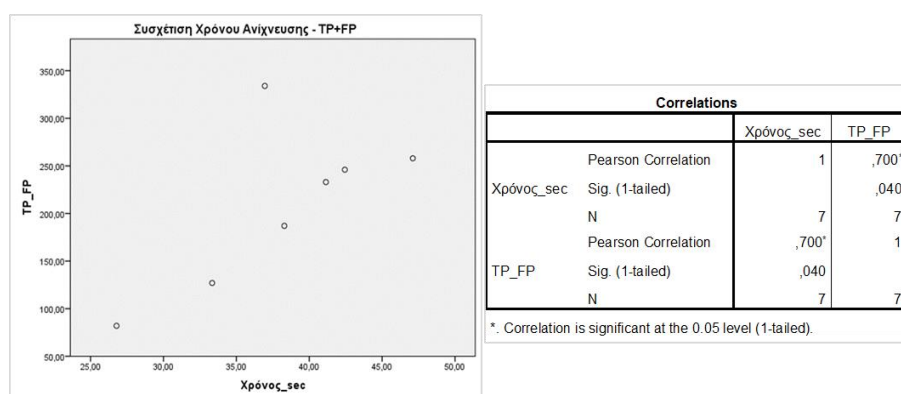
¹²² Οι χρόνοι του πίνακα 21 προήλθαν από το μέσο όρο 5 μετρήσεων

ανίχνευσης 3 εικόνες με διαφορετικό αριθμό οχημάτων σε κάθε μία από αυτές (μεγάλο – μέσο – μικρό). Τα αποτελέσματα του πίνακα 22 επιβεβαίωσαν την αρχική υποψία καθώς φαίνεται ότι ο χρόνος για την επεξεργασία των εικόνων 16, 21 και 24, ακολουθεί φθίνουσα πορεία και άμεσα συσχετιζόμενη με τον αριθμό των οχημάτων.

ΜΟΝΤΕΛΟ	ΧΡΟΝΟΣ ΑΝΙΧΝΕΥΣΗΣ (sec) (TP – FP)		
	Εικόνα 16 (21 οχήματα)	Εικόνα 21 (10 οχήματα)	Εικόνα 24 (2 οχήματα)
HAAR_GAB_BASIC_20stages	1.73(12-5)	1.52 (10 - 3)	0.68 (2-1)
HAAR_GAB_CORE_20stages	1.71 (12-6)	1.41 (8-4)	0.53 (2-0)
HAAR_GAB_ALL_20stages	1.32 (11-1)	1.29 (8-2)	0.56 (2-0)
HAAR_GAB_ALL_25stages	1.34 (11-1)	1.15 (8-0)	0.58 (2-0)
LBP_GAP_20stages	1.31 (14-8)	1.12 (10-8)	0.57 (2-0)
LBP_GAP_25stages	1.14 (9-1)	1.11 (7-1)	0,49 (1-0)
LBP_GAP_25stages_0.45MFAR	0.59 (4-0)	0.76 (5-0)	0,43 (1-0)
HOG_GAP_20stages	1.85 (15-1)	1.68 (10-1)	0.85(2-0)
HOG_SVM	4.49 (17-1)	3.68(9-0)	2.85(2-0)

Πίνακας 23. Συσχέτιση χρόνου επεξεργασίας εικόνας και αριθμού ανιχνεύσεων (από τους αριθμούς στην παρένθεση ο πρώτος αναφέρεται σε TP και ο δεύτερος σε FP ανιχνεύσεις – οι 3 εικόνες προήλθαν από το test set διαστάσεων 1920x1080 της VERVwUAS DB)

Ανάλογα αποτελέσματα δίνει και η εξέταση του δείκτη συσχέτισης Pearson αναφορικά με το χρόνο επεξεργασίας του test set (εικόνες 1920x1080) κάθε μοντέλου (αφορά στις τιμές του πίνακα 22 εξαιρουμένων των μοντέλων HOG_GAB/SVM) και του αθροίσματος TP και FP που πέτυχε κάθε ένα από αυτά. Σύμφωνα με την εικόνα 72 και τον παρατιθέμενο σε αυτή πίνακα, υπάρχει ισχυρή θετική συσχέτιση μεταξύ τους (δείκτης Pearson:0,7)



Εικόνα 72. Συσχέτιση μεταξύ χρόνου επεξεργασίας και του αθροίσματος TP και FP των μοντέλων (προέρχεται από τις τιμές του πίνακα 22 εξαιρουμένων των μοντέλων HOG_GAB και HOG_SVM)

Μετά τα αποτελέσματα των παραπάνω μετρήσεων, και προκειμένου να εξαληφθεί ο παράγοντας "ανίχνευση" από τις μετρήσεις για το χρόνο επεξεργασίας, αναζητήθηκαν εικόνες του test set στις οποίες το σύνολο των μοντέλων πέτυχαν τις ίδιες τιμές TP, FP και FN. Κατά περίπτωση οι εικόνες αυτές ήταν οι υπ' αριθμό 4 και 27. Ο χρόνος που απαιτήθηκε για την επεξεργασία τους φαίνεται στον πίνακα 23.

ΜΟΝΤΕΛΟ	ΕΙΚΟΝΑ	ΧΡΟΝΟΣ ΑΝΙΧΝΕΥΣΗΣ (sec)
HAAR_GAB_BASIC_20stages	4	0,633
HAAR_GAB_CORE_20stages	27	0,644
HAAR_GAB_ALL_20stages	4	0,640
HAAR_GAB_ALL_25stages	4	0,638
LBP_GAP_20stages	27	0,670
LBP_GAP_25stages	4	0,665
LBP_GAP_25stages_0.45MFAR	4	0,595
HOG_GAP_20stages	4	0,984
HOG_SVM	4	2,964

Πίνακας 24. Χρόνος επεξεργασίας εικόνων (οι εικόνες 4 και 27 προήλθαν από το αντίστοιχο test set διαστάσεων 1920x1080 της VERVwUAS DB)¹²³

Οι διαπιστώσεις που προκύπτουν είναι ότι για το σύνολο των μοντέλων, πλην των HOG, ο χρόνος επεξεργασίας δε διαφέρει σημαντικά και κυμαίνεται στα $0,65 \pm 0,02 \text{sec}$ (μόνο για το LBP_GAP_25stages_0,45MFAR έπεσε κάτω των 0,6sec). Για τα HOG_GAB/SVM φαίνεται ότι η εξαιρετική τους απόδοση στην ανίχνευση δεν έρχεται χωρίς κόστος. Έτσι το μοντέλο HOG_GAB χρειάστηκε σχεδόν 1sec για την επεξεργασία της εικόνας ενώ το HOG_SVM επέδειξε τον απογοητευτικό χρόνο των 3sec περίπου. Για τα δύο τελευταία μοντέλα, λόγω των πολύ καλών επιδόσεών τους στον τομέα της ανίχνευσης, μετρήθηκαν οι αντίστοιχοι χρόνοι ανίχνευσης και σε εικόνες διαστάσεων 960x540 (υπενθυμίζεται η μετατροπή τους από το πρόγραμμα ανίχνευσης στη διάσταση 1280x720). Και τα δύο μοντέλα μείωσαν κατά το ήμισυ τους χρόνους με το πρώτο να επιτυγχάνει 0,425sec και το δεύτερο 1,47sec.

Ένα ακόμα ενδιαφέρον στοιχείο αφορά στη μέτρηση του χρόνου επεξεργασίας με την

¹²³ Οι συγκεκριμένες τιμές αποτελούν το μέσο όρο 10 μετρήσεων.

προσθήκη ή όχι στο βρόγχο επεξεργασίας της συνάρτησης του Non-Maximum-Suppression. Αναμένονταν, ότι λόγω της επιβάρυνσης της επεξεργασίας που προέρχεται από την εφαρμογή της NMS, θα απαιτούνταν περισσότερος χρόνος επεξεργασίας. Οι μετρήσεις έδειξαν το ακριβώς αντίθετο. Συγκεκριμένα, το πρόγραμμα που δεν περιλάμβανε τη συνάρτηση NMS, απαιτούσε περίπου 3 έως 5 δευτερόλεπτα περισσότερο για την ολοκλήρωση της επεξεργασίας του συνόλου των εικόνων (πχ για το HAAR_BASIC απαιτήθηκαν 51,61sec για την επεξεργασία των 54 εικόνων χωρίς NMS).

6.2 2ο Στάδιο Αξιολόγησης: Περαιτέρω Μελέτη των Μοντέλων που Διακρίθηκαν

Από την μελέτη της απόδοσης του συνόλου των μοντέλων που αξιολογήθηκαν κατά το πρώτο στάδιο, το HOG_GAB 20 σταδίων εκπαίδευσης και το HOG_SVM ήταν αυτά με τις καλύτερες επιδόσεις. Στη συνέχεια, χρησιμοποιώντας το ίδιο test set των 54 εικόνων (οι οποίες όμως συμπίεστηκαν στο 40% του αρχικού τους μεγέθους¹²⁴), αξιολογήθηκε η απόδοση των δύο αυτών μοντέλων στην ανίχνευση οχημάτων. Οι διαστάσεις που μελετήθηκαν, και για τα δύο μοντέλα, ήταν: (1) 1920x1080 και (2) 960x540. Η τελευταία, με την είσοδό της στο πρόγραμμα ανίχνευσης μετατρέπονταν σε 1280x720 και 1920x1080. Στη συνέχεια της συγκεκριμένης ενότητας, θα γίνει αναφορά και στη δυνατότητα των μοντέλων να εντοπίσουν οχήματα σε φωτογραφίες που προέρχονται από τις βάσεις δεδομένων που παρουσιάστηκαν στο 4ο κεφάλαιο.

6.2.1 Αξιολόγηση της Απόδοσης σε Συμπιεσμένες Εικόνες

Από τη μελέτη των τιμών για τα TP, FP και FN, καθώς και των μετρικών δεικτών που προκύπτουν από αυτές, για τα μοντέλα HOG_GAB 20 σταδίων εκπαίδευσης και HOG_SVM, διαπιστώνεται σε γενικές γραμμές:

- Τάση μείωσης των TP με αντίστοιχη αύξηση του αριθμού των FN, μετακινούμενοι προς εικόνες μικρότερων διαστάσεων και ανάλυσης.
- Αντίστοιχη μείωση της τιμής του δείκτη F1.
- Διατήρηση του δείκτη F1, και για τα δύο μοντέλα στο σύνολο του dataset, άνω του 0,8.
- Σημαντική μείωση της τιμής των TP για τις συμπιεσμένες εικόνες διαστάσεων

¹²⁴ Για τη συμπίεση χρησιμοποιήθηκε το πρόγραμμα το πρόγραμμα Light Image Resizer.

960x540.

ΔΕΙΓΜΑ ΕΙΚΟΝΩΝ	TP	FP	FN	ACCURACY	RECALL	PRECISION	ΔΕΙΚΤΗΣ F1
1920x1080	223	9	29	0,85441	0,88492	0,96121	0,92149
960x540 (→1920x1080)	215	10	37	0,82061	0,85317	0,95556	0,90147
1920x1080 (Compressed)	204	3	48	0,8	0,80952	0,98551	0,88889
960x540 (Compressed) (→1280x720)	181	2	71	0,71259	0,71825	0,98907	0,83218

Πίνακας 25. Μετρικοί δείκτες του HOG_GAB στο σύνολο των εικόνων του test set της VERVwUAS DB

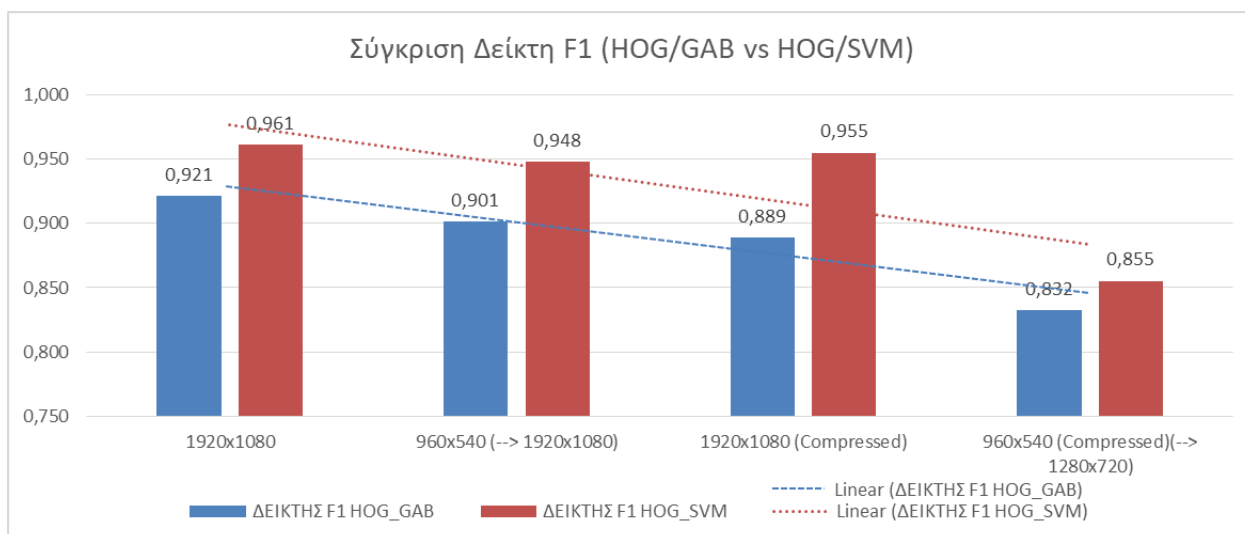
ΔΕΙΓΜΑ ΕΙΚΟΝΩΝ	TP	FP	FN	ACCURACY	RECALL	PRECISION	ΔΕΙΚΤΗΣ F1
1920x1080	234	1	18	0,9249	0,92857	0,99574	0,96099
960x540 (→1920x1080)	228	1	24	0,90119	0,90476	0,99563	0,94802
1920x1080 (Compressed)	232	2	20	0,91339	0,92063	0,99145	0,95473
960x540 (Compressed) (→1280x720)	189	1	63	0,74703	0,75	0,99473	0,85520

Πίνακας 26. Μετρικοί δείκτες του HOG_SVM στο σύνολο των εικόνων του test set της VERVwUAS DB

- Επεξεργασία εικόνας video σε πραγματικό ή μη χρόνο (αναλόγως των υπολογιστικών δυνατοτήτων) για την μελέτη των υφιστάμενων τεχνικών αυτόματης ανίχνευσης και ιχνηλάτησης στατικών ή κινούμενων οχημάτων από εναέρια μέσα, για σκοπούς επιτήρησης περιοχών με στόχο την πρόληψη πυρκαγιών (πχ εντοπισμό και ιχνηλάτηση οχημάτων που εισέρχονται σε δασική περιοχή) ή για σκοπούς ενεργοποίησης των μηχανισμών έρευνας και διάσωσης σε περίπτωση φυσικών καταστροφών/ανθρωπιστικών καταστροφών (αυτόματη παρακολούθηση συμβάντος και μεταφορά εικόνας σε επίγειο σταθμό διαχείρισης κρίσεων για λήψη μέτρων αντιμετώπισης).
- Μελέτη των υφιστάμενων δυνατοτήτων μετάδοσης εικόνας (video streaming), από τις αναφερθείσες εναέρια πλατφόρμες προς επίγειους σταθμούς, με τη χρήση του δικτύου κινητής τηλεφωνίας, σύμφωνα πάντα με το διαθέσιμο σε πραγματικό χρόνο bandwidth που προσφέρεται άνωθεν του ηπειρωτικού ή του θαλάσσιου χώρου.

Συγκρίνοντας του δείκτες F1 των δύο μοντέλων στο διάγραμμα 10, φαίνεται ότι ο

HOG_SVM ουσιαστικά επηρεάζεται σημαντικά μόνο στην περίπτωση των συμπιεσμένων εικόνων διαστάσεων 960x540 οπότε και υφίσταται μείωση της τάξης του 10% περίπου. Στον αντίποδα, ο δείκτης F1 του HOG_GAB παρουσιάζει μία συνεχή γραμμική μείωση



Διάγραμμα 10. Ιστόγραμμα σύγκρισης τιμής δείκτη F1 μεταξύ HOG_GAB και HOG_SVM

6.2.2 Εφαρμογή των HOG_GAB/SVM σε Διαφορετικά Datasets

Τα datasets, εικόνες των οποίων χρησιμοποιήθηκαν για να διαπιστωθεί η γενίκευση των 2 μοντέλων ήταν : (1) η Overhead Imagery Research Data Set (OIRDS), (2) η Dutch UAS Dataset 001, (3) η DLR Munich Vehicle Dataset (DLR-MVD) και (4) η Vehicle Detection in Aerial Imagery (VEDAI). Η εφαρμογή των μοντέλων στην Dutch UAS Dataset 001 ήταν ανεπιτυχής σε ότι αφορά στην ανίχνευση οχημάτων. Επ’ αυτού επισημαίνεται ότι ο αριθμός οχημάτων στη συγκεκριμένη βάση ήταν ελάχιστος και οι λήψεις στα περισσότερα από αυτά έγιναν με γωνία περίπου 30° ως προς το κατακόρυφο.

Ορισμένα χαρακτηριστικά σημεία για τις εικόνες αυτών των datasets που μελετήθηκαν είναι:

- Παρά την ήδη μεγάλη διάσταση των φωτογραφιών της DLR Munich Vehicle Dataset (5616x3744) και προκειμένου τα οχήματα να αποκτήσουν τις διαστάσεις οι οποίες χρησιμοποιήθηκαν για την εκπαίδευση, οι εν λόγω φωτογραφίες μεγεθύνθηκαν κατά ένα παράγοντα $\times 1,3^{125}$ για τον HOG_GAB και $\times 1,5$ φορές για τον HOG_SVM. Για την επεξεργασία της φωτογραφίας των εικόνων 73 και 74, ο πρώτος χρειάστηκε 61,294sec και ο δεύτερος 640,091sec

¹²⁵ Η αρχική προσπάθεια για μεγένθυση $\times 1,5$ δεν κατέστη επιτυχής λόγω μη επαρκούς μνήμης.

αντίστοιχα. Στις συγκεκριμένες φωτογραφίες ο φακός ήταν τοποθετημένος σε Α/Φ που πετούσε σε ύψος 1000μ, ενώ για την VERNwUAS DB το DJI Phantom βρισκόταν σε ύψος 70m άνωθεν του εδάφους. Παρά τις διαφορές που παρουσιάζουν οι δύο βάσεις δεδομένων και το γεγονός ότι η διαδικασία της εκπαίδευσης των κατηγοριοποιητών δεν τροφοδοτήθηκε με εικόνες από την DLR-MVD, πέτυχαν ικανό αριθμό TP αναγνώρισεων (οχήματα με προσανατολισμό παράλληλο στο διαμήκη άξονα της φωτογραφίας)



Εικόνα 73. Εφαρμογή του μοντέλου HOG_GAB σε φωτογραφία της DLR-MVD



Εικόνα 74. Εφαρμογή του μοντέλου HOG_SVM σε φωτογραφία της DLR-MVD

- Στον αντίποδα από πλευράς διαστάσεων είναι η OIRDS dataset. Οι φωτογραφίες που χρησιμοποιήθηκαν (εικόνες 75 και 76 για κάθε ένα από τα δύο μοντέλα), επιδιώχθηκε να απεικονίζουν οχήματα σε οριζόντιο προσανατολισμό. Οι αρχικές τους διαστάσεις ήταν 257x257 και κατά τη διαδικασία επεξεργασίας τους από

τα μοντέλα μετατράπηκαν σε 1000x1000.



Εικόνα 75. Εφαρμογή του μοντέλου HOG_GAB σε φωτογραφίες της OIRDS dataset



Εικόνα 76. Εφαρμογή του μοντέλου HOG_SVM σε φωτογραφίες της OIRDS dataset

- Τέλος, το χαρακτηριστικό της VEDAI είναι ότι διαθέτει εικόνες που προέρχονται τόσο από το οπτικό, όσο και από το υπέρυθρο (IR) φάσμα. Οι φωτογραφίες μετατράπηκαν από την αρχική διάσταση των 512x512 σε αυτή των 1500x1500¹²⁶. Στις εικόνες 77 και 78 αντίστοιχα φαίνονται τα αποτελέσματα της εφαρμογής των μοντέλων ανίχνευσης.



Εικόνα 77. Εφαρμογή του μοντέλου HOG_GAB σε φωτογραφία της VEDAI dataset (η

¹²⁶ Η μεγένθυση έγινε προκειμένου το μέγεθος των οχημάτων στις εικόνες (διαστάσεις του οχήματος σε pixel), να υπερβεί τις αντίστοιχες διαστάσεις του παραθύρου εκπαίδευσης (58x24).

δεύτερη σειρά εικόνων είναι στο IR φάσμα)



Εικόνα 78. Εφαρμογή του μοντέλου HOG_SVM σε φωτογραφία της VEDAI dataset (η δεύτερη σειρά εικόνων είναι στο IR φάσμα)

Οι βάσεις δεδομένων που εξετάστηκαν στην συγκεκριμένη υποενότητα διέθεταν annotation file πλην όμως δεν χρησιμοποιήθηκαν για την πραγματοποίηση ποσοτικής αξιολόγησης των μοντέλων HOG_GAB και HOG_SVM λόγω:

- Του προσανατολισμού των οχημάτων στα εν λόγω datasets ο οποίος ήταν τυχαίος. Επελέγησαν εικόνες στις οποίες η διεύθυνση των οχημάτων ήταν 0° - 180° ως προς τον διαμήκη άξονα της φωτογραφίας, προκειμένου να συμφωνεί με το δείγμα εκπαίδευσης.
- Του υφιστάμενου χρονικού περιορισμού για την ολοκλήρωση της μεταπτυχιακής διατριβής.

6.3 3ο Στάδιο Αξιολόγησης: Απόδοση Αλγόριθμου Ιχνηλάτισης

Το συγκεκριμένο στάδιο αξιολόγησης, περιλαμβάνει την μέτρηση της απόδοσης των 2 παραπάνω μοντέλων ανίχνευσης, σε συνδυασμό με τον αλγόριθμο ιχνηλάτισης¹²⁷. Η υλοποίηση της διάταξης συνίσταται στην λειτουργία ενός βρόγχου μεταξύ του μοντέλου ανίχνευσης και του αλγόριθμου ιχνηλάτισης (το μοντέλο ανίχνευσης ερευνά περιοδικά

¹²⁷ Η μεθοδολογία που ακολουθήθηκε για την αξιολόγησή τους, περιγράφεται στην ενότητα 5.2

την εικόνα για ύπαρξη οχημάτων και οι αντίστοιχες συντεταγμένες ανίχνευσης διαβιβάζονται στον tracker είτε για ανανέωση υφιστάμενης ιχνηλάτησης είτε για έναρξη νέας).

Από το σύνολο των εννέα μοντέλων, επελέγησαν τα HOG_GAB και HOG_SVM. Από αυτά τα δύο, το πρώτο, αν και είναι πιο γρήγορο, εν τούτοις, αποδείχτηκε υποδεέστερο στην απόδοση σε σχέση με το HOG_SVM (ειδικά στην περίπτωση ανίχνευσης οχημάτων σε εικόνες χαμηλής ποιότητας).

Παρά την πρόκριση των 2 παραπάνω μοντέλων, το μοναδικό που αξιολογήθηκε ήταν το HOG_SVM καθώς δεν κατέστη δυνατή η λειτουργία του μοντέλου HOG_GAB σε συνδυασμό με την εφαρμογή του αλγορίθμου ιχνηλάτησης. Συγκεκριμένα, ο autocorrelation tracker ο οποίος λειτουργούσε σε python (SPYDER IDE) και χρησιμοποιούσε τις βιβλιοθήκες του OpenCV 3.3, για τη λειτουργία της ανίχνευσης και της Dlib 19.4, για τη λειτουργία της ιχνηλάτησης, εμφάνισε σφάλμα. Μετά από έρευνα στο διαδίκτυο διαπιστώθηκε ότι αυτό οφείλονταν σε αφαίρεση της υλοποίησης των HOG cascades από τις βιβλιοθήκες του OpenCV εκδόσεως 3.0 ή μεταγενέστερων αυτής¹²⁸.

Τα αποτελέσματα για τον HOG_SVM, από την εφαρμογή του στα τέσσερα test video που δημιουργήθηκαν για τους σκοπούς της αξιολόγησης, φαίνονται στον πίνακα 26.

ΠΕΡΙΟΧΗ ΛΗΨΗΣ VIDEO	ΑΝΑΛΥΣΗ	BIT RATE	ΟΧΗΜΑΤΑ (GROUND TRUTH)	TP DETECTED	TP TRACKED	FN	FP
Περιοχή 1	1280x720	512Kbps	17	14	11	3	---
Περιοχή 2	-/-	-/-	36	26	17	10	1
Περιοχή 3	-/-	-/-	7	7	7	---	---
Περιοχή 4	-/-	-/-	11	11	11	---	---

Πίνακας 27. Αποτελέσματα αξιολόγησης ανίχνευσης – ιχνηλάτησης του μοντέλου HOG_SVM στα test video.

Λόγω της χρονικής καθυστέρησης του μοντέλου ανίχνευσης, και προκειμένου να αποφευχθεί χρονική υστέρηση, προγραμματίστηκε η εφαρμογή του αλγορίθμου

¹²⁸ Σύμφωνα με τον Vadim Pisarevsky (2015-05-28 14:24): «We decided to drop the current HOG cascades in OpenCV 3.x. The implemented HOG features are quite weird - different from Dalal's interpretation of HOG, different from P. Dollar integral channel features» (<https://github.com/opencv/opencv/issues/4989>) Πρόσβαση:6/5/18

ανίχνευσης ανά 75 frames. Αυτή η ρύθμιση, οδήγησε στην αριθμητική διαφορά μεταξύ ανιχνευόμενων (detected) και ιχνηλατούμενων (tracked) οχημάτων, σε σχέση με το ground truth. Συγκεκριμένα, για τα οχήματα που παρέμεναν για μικρό χρονικό διάστημα εντός του οπτικού πεδίου της κάμερας¹²⁹ (π.χ. μικρότερο των 3 δευτερολέπτων για video των 25 fps), δεν παρέχονταν επαρκής χρόνος για την περιοδική εφαρμογή του αλγορίθμου ανίχνευσης. Σε ότι αφορά στην περιοχή 2, τα 4 από τα 10 οχήματα που δεν ανιχνεύτηκαν (FN) ήταν εκείνα των οποίων η διεύθυνση διέφερε περίπου 10°, σε σχέση με την οριζόντια. Τέλος, η αναπαραγωγή του video κατά την λειτουργία του προγράμματος ανίχνευσης – ιχνηλάτησης, παρουσιάζει χρονική υστέρηση ανάλογη με τον αριθμό των οχημάτων που έχουν ανιχνευτεί.

6.4 Μετρήσεις Ταχύτητας Δικτύου

Οι μετρήσεις της ταχύτητας μεταφοράς δεδομένων του ασύρματου δικτύου κινητής τηλεφωνίας, αφορούν σε δύο περιπτώσεις, σε αυτή του UAS και σε αυτή του Ε/Π. Στην πρώτη περίπτωση μετρήθηκε¹³⁰ η αμφίδρομη ταχύτητα ροής δεδομένων, στις τρεις διαφορετικές καταστάσεις λειτουργίας του drone. Τα αποτελέσματα των μετρήσεων έχουν καταγραφεί στο Παράρτημα Β ενώ η στατιστική τους ανάλυση παρουσιάζεται στον πίνακα 27 που ακολουθεί. Από τη μελέτη τους διαπιστώνεται ότι η λειτουργία του UAV δε φαίνεται να επηρεάζει το ρυθμό ροής δεδομένων.

	MINIMUM	MAXIMUM	MEAN	STD. DEVIATION
OFF_Download	3,37	4,01	3,5560	0,26350
OFF_Upload	2,11	2,57	2,3480	0,20921
IDLE_Download	3,24	4,21	3,8460	0,36596
IDLE_Upload	2,36	2,54	2,4760	0,06950
FLY_Download	3,66	4,13	3,8760	0,21102
FLY_Upload	2,22	2,66	2,4460	0,19591

Πίνακας 28. Στατιστικά δεδομένα ρυθμού ροής δεδομένων δικτύου κινητής τηλεφωνίας (χρήση UAS)

Στην δεύτερη περίπτωση, η μέτρηση του ρυθμού ροής δεδομένων πραγματοποιήθηκε κατά την διάρκεια πτήσης του Ε/Π, σύμφωνα με όσα αναφέρθηκαν στην ενότητα 5.3.

¹²⁹ Για παράδειγμα, στην περιοχή 1 που η κάμερα κινείται παράλληλα στο δρόμο, τα οχήματα που κινούνται στην αντίθετη διεύθυνση μένουν για μικρότερο χρονικό διάστημα εντός του οπτικού πεδίου, σε σχέση με εκείνα που κινούνται στην ίδια διεύθυνση με αυτή.

¹³⁰ Οι μετρήσεις πραγματοποιήθηκαν εντός της πόλης του Βόλου, με το DJI Phantom προσγειωμένο στην οροφή ενός κτιρίου (ύψος 12m).

Τα αποτελέσματα των μετρήσεων έχουν καταγραφεί στο Παράρτημα Β, ενώ η στατιστική τους ανάλυση παρουσιάζεται στους πίνακες 28 και 29 που ακολουθούν (γίνεται αναφορά μόνο στην ταχύτητα upload).

	MINIMUM	MAXIMUM	MEAN	STD. DEVIATION
500ft / 50Knots	1,98	3,99	2,6190	0,55042
500ft / 100Knots	1,05	3,17	1,9800	0,58867
1000ft / 50Knots	1,89	3,20	2,4920	0,39160
1000ft / 100Knots	0,86	2,67	1,4460	0,67732
3000ft / 50Knots	0,59	0,87	0,7240	0,09980
3000ft / 100Knots	0,68	2,97	1,5110	0,84100

Πίνακας 29. Στατιστικά δεδομένα ρυθμού ροής δεδομένων δικτύου κινητής τηλεφωνίας από Ε/Π (περιοχή πτήσης : Κιλελέρ-Αρμένιο)

	MINIMUM	MAXIMUM	MEAN	STD. DEVIATION
500ft / 50Knots	0,72	2,88	1,6710	0,71612
500ft / 100Knots	0,88	2,10	1,2640	0,44647
1000ft / 50Knots	0,56	2,86	1,4080	0,77086
1000ft / 100Knots	0,45	1,16	0,6850	0,23908
3000ft / 50Knots	0,46	1,51	0,8650	0,29669
3000ft / 100Knots	0,44	0,95	0,7720	0,17731

Πίνακας 30. Στατιστικά δεδομένα ρυθμού ροής δεδομένων δικτύου κινητής τηλεφωνίας από Ε/Π (περιοχή πτήσης : Βόλος – με κόκκινο χρώμα οι συνθήκες στις οποίες μπορεί να μην επιτραπεί η μετάδοση video 512bps)

Από τη μελέτη των δύο τελευταίων πινάκων διαπιστώνεται ότι:

- Η αρχική υπόθεση για μεγαλύτερη ταχύτητα του δικτύου κινητής τηλεφωνίας στην περιοχή του Βόλου, ο οποίος καλύπτεται από δίκτυο 4G, δεν επαληθεύτηκε (όλες οι μέσες τιμές ταχύτητας δικτύου, πλην αυτής που αφορά τα 3000ft/50Knots, ήταν χαμηλότερες στην περιοχή του Βόλου).
- Δε δύναται να εξαχθούν ασφαλή συμπεράσματα για το αν υπάρχει συσχέτιση μεταξύ του ύψους - ταχύτητας πτήσης και του ρυθμού μετάδοσης δεδομένων
- Διαφαίνεται ότι το δίκτυο επιτρέπει πάντα τη ροή video 384bps και σχεδόν πάντα σε 512bps.

6.5 Near Real Time Video Streaming

Η αξιολόγηση του video streaming αφορά 3 επίπεδα λειτουργίας. Το πρώτο σχετίζεται με την μελέτη της δυνατότητας αποστολής video συγκεκριμένου bit rate, μέσω του δικτύου κινητής τηλεφωνίας, από Ε/Π προς ένα σταθμό εδάφους. Το δεύτερο και τρίτο

επίπεδο αφορά στην πειραματική υλοποίηση της διάταξης για ανίχνευση και ιχνηλάτηση οχημάτων από ροή video η οποία προέρχεται είτε από Ε/Π είτε από σύστημα UAS, σε σχεδόν πραγματικό χρόνο.

6.5.1 Video Streaming από Ε/Π προς Σταθμό Εδάφους

Με βάση τις μετρήσεις ταχύτητας του δικτύου κινητής τηλεφωνίας που αναφέρθηκαν στην προηγούμενη ενότητα, αποφασίστηκε η δοκιμή μετάδοσης ροής video από το Ε/Π προς το σταθμό εδάφους (οικιακός Η/Υ), σε τρεις ταχύτητες : (1) 384bps (2) 512 bps και (3) 1024bps¹³¹. Η περιοχή πτήσης, από την οποία έγινε η αποστολή, περιλάμβανε την εγγύτητα του Α/Δ Στεφανοβικείου. Το video που χρησιμοποιήθηκε είχε resolution 1920x1080 και διάρκεια 1min20sec σε 25 frames/s. Η αποστολή του έγινε μέσω του VLC player, σύμφωνα με τις ρυθμίσεις του πίνακα 30 και πρωτόκολλο συμπίεσης H.264.

Χαρακτηριστικά Αποστολής				Ποιοτικά Χαρακτηριστικά Αναπαραγωγής στο Δέκτη	
Bit Rate (Kbps)	Αριθμός video	Frame Rate (fps)	Resolu-tion	Υστέρηση	Πάγωμα Εικόνας
384	2	25	960x540	≅ 30sec	Όχι
512	2	25	960x540	-/-	-/-
1024	1	25	960x540	-/-	Ναι

Πίνακας 31. Χαρακτηριστικά αποστολής και λήψης video streaming.

Η υστέρηση που αναφέρεται στον πίνακα 30 σχετίζεται με το χρόνο που μεσολάβησε από την στιγμή της αποστολής μέχρι και την έναρξη αναπαραγωγής του video στο δέκτη. Για τα video με bit rate 384 και 512 Kbps, η αναπαραγωγή τους στο δέκτη ήταν ομαλή και συνεχής. Το video που απεστάλη με bitrate 1024Kbps ενώ αρχικά (στα πρώτα 20sec), είχε ομαλή ροή, στη συνέχεια πάγωσε και ο VLC player σταμάτησε την αναπαραγωγή. Για τη συγκεκριμένη περίπτωση επισημαίνονται δύο βασικά στοιχεία που σχετίζονται με τη ροή δεδομένων :

- Η διακοπή της ομαλής αναπαραγωγής της ροής του video σχετίζεται άμεσα με την περιοχή πτήσης του Ε/Π και την κάλυψη του δικτύου. Τη στιγμή της διακοπής, το Ε/Π εκτελούσε πτήση εγγύς του εδάφους και συγκεκριμένα σε περιοχή με έντονες χαραδρώσεις του ανάγλυφου. Οπότε, ο παράγοντας κάλυψη δικτύου διετέλεσε σημαντικό ρόλο στην αδυναμία μετάδοσης του video.
- Παρά το γεγονός ότι ο VLC player δεν αναπαρήγαγε το video στην οθόνη του

¹³¹ Ανάλυση της μεθοδολογίας που ακολουθήθηκε για την ασύρματη αποστολή video, γίνεται στην ενότητα 5.4

δέκτη, εν τούτοις, στο background συνεχίστηκε η ροή του και αποθηκεύτηκε ολόκληρο στον Η/Υ.

6.5.2 Εφαρμογή Μοντέλου Ανίχνευσης και Ιχνηλάτησης Οχημάτων με Video Streaming από Ε/Π

Για την αποστολή του video streaming από το Ε/Π προς το σταθμό εδάφους (Lenovo T510 με ενσύρματη σύνδεση στο διαδίκτυο), το λαμβανόμενο video από την κάμερα του Raspberry 1080p30fps με ψηφιακό zoom x4, μετατρέπονταν από το VLC σε 1080p25fps και μεταδίδονταν με bit rate 512 Kbps. Το ύψος πτήσης του Ε/Π ήταν 300-350ft και η ταχύτητά του 50Knots.

Μετά την αποκατάσταση της διασύνδεσης μέσω VPN με το σταθμό εδάφους άρχισε η ασύρματη ροή video η οποία στη συνέχεια δρομολογήθηκε στο πρόγραμμα ιχνηλάτησης. Λόγω του χρόνου που καταναλώνει το μοντέλο HOG_SVM για την λειτουργία της ανίχνευσης και προκειμένου να υπάρχει όσο το δυνατόν πιο ομαλή near real time ροή video, ρυθμίστηκε η λειτουργία του αλγόριθμου ανίχνευσης κάθε 75 πλαίσια.



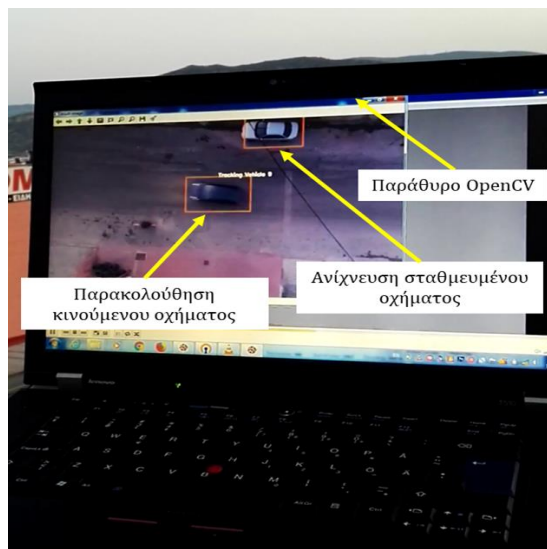
Εικόνα 79. Ανίχνευση και ιχνηλάτηση οχήματος μέσω video streaming από Ε/Π

Τα αποτελέσματα ήταν αρκετά ενθαρρυντικά καθόσον όπως φαίνεται στην εικόνα 79 αλλά και στα video που έχουν καταγραφεί, παρά την κακή ποιότητα του λαμβανόμενου video, η συνολική διάταξη λειτούργησε επιτυχάνοντας το σκοπό της. Το μοντέλο αντιμετώπισε προβλήματα μόνο στην περίπτωση ανίχνευσης των οχημάτων τα οποία διατηρούνταν εντός του οπτικού πεδίου της κάμερας για χρόνο μικρότερο των 3sec. Αυτό οφείλονταν στη ρύθμιση του μοντέλου να διεξάγει ανίχνευση κάθε 75 πλαίσια¹³².

¹³² Λόγω της ροής του video με 25fps, η περίοδος ανίχνευσης ήταν 3sec ($75:25=3\text{sec}$)

6.5.3 Εφαρμογή Μοντέλου Ανίχνευσης και Ιχνηλάτησης Οχημάτων με Video Streaming από UAS

Αναφορικά με την αποστολή ροής video από το DJI Phantom Pro, αυτή κατέστη εφικτή με την υλοποίηση της διάταξης που περιγράφεται στην ενότητα 5.5. Το video ανάλυσης 1080p30 που καταγράφηκε από την Raspberry Pi Camera Module v2, μετατράπηκε μέσω του VLC 960x540/10fps και απεστάλη στο Lenovo T510 όπου έτρεχε το μοντέλο ανίχνευσης και ιχνηλάτησης οχημάτων HOG_GAB. Αν και η δοκιμή ήταν επιτυχής, στιγμιότυπο της οποίας φαίνεται στην εικόνα 80, εν τούτοις δεν κατέστη εφικτή η μεθοδική μελέτη των αποτελεσμάτων καθόσον απαιτήθηκε η Π/Γ του DJI Phantom, για λόγους ασφαλείας, λόγω δυσχερειών που παρουσίαζε ο έλεγχός του. Ως πιθανά αίτια θεωρούνται:



Εικόνα 80. Ανίχνευση και ιχνηλάτηση οχήματος μέσω video streaming από UAS

- Η αλλαγή του κέντρου βάρους με την προσθήκη της διάταξης Raspberry – power bank στη βάση του drone. Η συγκεκριμένη θέση, η οποία είναι και η πλέον απομακρυσμένη από το KB του αεροχήματος, αυξάνει ανάλογα και τη δημιουργούμενη ροπή στρέψης, περιορίζοντας ταυτόχρονα τα όρια ελέγχου ως προς το pitch και το roll.
- Παρεμβολή της συχνότητας ελέγχου του drone¹³³ από το την εκπομπή του 4G router¹³⁴.

¹³³ Ο μηχανισμός ελέγχου λειτουργεί στη περιοχή συχνοτήτων : 2.400 - 2.483 MHz

¹³⁴ Το LTE δίκτυο της Cosmote στην Ελλάδα για το upload εκμεταλλεύεται τις μπάντες: (1) B20 (800 DD)/832 – 862 MHz, B3(1800+) /1710 – 1785 MHz και B7 (2600)/2500 – 2570 MHz

Κεφάλαιο

7

Συμπεράσματα

Αναφορικά με το χρόνο που δαπανήθηκε για την εκπαίδευση των μοντέλων cascade ιδίου αριθμού σταδίων, επιβεβαιώθηκαν τα συμπεράσματα και άλλων ερευνητικών εργασιών που αναφέρουν αύξηση της χρονικής διάρκειας εκπαίδευσης, κατά τη μετάβαση από το feature type LBP στο Haar (OpenCV by Doxygen, 2018) και τέλος στο HOG. Επίσης, ένας ακόμα παράγοντας που επηρέασε την διάρκεια της εκπαίδευσης είναι και το αρχικά επιλεγέν παράθυρο 58x24. Η ελάττωσής του κατά ένα παράγοντα 4 (29x12) επιτάχυνε την όλη διαδικασία. Συναφώς επισημαίνεται ότι η εκπαίδευση του HOG_GAB_20stages με μέγεθος παραθύρου 58x24 αποτέλεσε και το αίτιο της μεγάλης απόκλισης που παρουσίασε ο χρόνος εκπαίδευσης του σε σχέση με τα υπόλοιπα cascades. Επεκτείνοντας τη σύγκριση μεταξύ των συγκεκριμένων cascade μοντέλων, μέσω της συσχέτισης του παράγοντα χρόνου με αυτόν της απόδοσης, παρατηρήθηκε ότι η μεταξύ τους σχέση είναι ανάλογη. Αυξανομένης της διάρκειας εκπαίδευσης αυξάνεται και η απόδοση του μοντέλου.

Σε κάθε στάδιο, ο αλγόριθμος της εκπαίδευσης προσπαθεί να εντοπίσει αρνητικά δείγματα τα οποία παρέχουν επαρκή διαφορετικότητα σε σχέση με τα ήδη εκπαιδευμένα. Αυτό σημαίνει ότι θα πρέπει να αποδοθεί μεγάλη σημασία στη διαδικασία δημιουργίας του αρνητικού δείγματος εικόνων που θα χρησιμοποιηθεί για την εκπαίδευση των cascades. Πρακτικά και μέσα από την παρούσα έρευνα, συμπεραίνεται ότι η αρχική προσπάθεια εκπαίδευσης (εισαγωγή 3000 έως 3500 εικόνων ως αρνητικό δείγμα), δεν ολοκληρώνονταν λόγω της σχετικής ομοιότητας που παρουσίαζαν αυτές μεταξύ τους (εικόνες προερχόμενες μόνο από τις περιοχές 1 και 2). Η επαναπροσέγγιση της διαδικασίας και ανανέωση του δείγματος εκπαίδευσης με επιλεγμένες εικόνες από το διαδίκτυο, οι οποίες προσέδιδαν μεγάλο βαθμό

διαφορετικότητας στο δείγμα, οδήγησε στην επιτυχή ολοκλήρωση της διαδικασίας με τη χρησιμοποίηση αρκετά μικρότερου αριθμού αρνητικών δειγμάτων (1600). Η τελευταία αυτή παρατήρηση θα πρέπει να συνδυαστεί και με το χρόνο εκπαίδευσης ο οποίος για το σύνολο των cascades, των οποίων ολοκληρώθηκε η εκπαίδευση και οδηγήθηκαν στη διαδικασία της αξιολόγησης, ήταν σημαντικά μικρότερος σε σχέση με τις αρχικές προσπάθειες εκπαίδευσης των αντίστοιχων μοντέλων όπου χρησιμοποιήθηκε μεγάλος αριθμός αρνητικού δείγματος, προερχόμενο μόνο από τις περιοχές 1 και 2.

Ο HOG_SVM παρουσίασε την καλύτερη απόδοση αναφορικά με την ποιότητα προσδιορισμού της θέσης των οχημάτων (localization), ακολουθούμενος από ένα cascade μοντέλο (HOG_GAP_20stages). Τα υπόλοιπα 4 cascade μοντέλα 20 σταδίων εκπαίδευσης παρουσίασαν παρόμοια συμπεριφορά. Η αύξηση των σταδίων εκπαίδευσης στα 25, επηρέασε αρκετά τη συμπεριφορά των μοντέλων. Συγκεκριμένα, αν και αύξησε ελαφρώς μεσοσταθμικά τη ποιότητα ανίχνευσης από την άλλη οδήγησε σε αύξηση των FN (σε μεγαλύτερο βαθμό για το LBP και λιγότερο για το Haar). Επίσης, η αύξηση του αριθμού σταδίων εκπαίδευσης μείωσε και τον αριθμό των ανιχνεύσεων με χαμηλό δείκτη IoU (<0,5).

Η αντιμετώπιση των πολλαπλών bounding boxes, ειδικά για τα Haar μοντέλα, με χρήση του αλγορίθμου non-maximum-suppression δεν κρίνεται ως ιδιαίτερα επιτυχής καθώς αν και επιτυγχάνει την απαλοιφή των πολλαπλών boxes, εν τούτοις ο συγκερασμός του σε ένα μεγαλύτερων διαστάσεων, οδηγεί στη συνέχεια σε απόρριψη μίας επιτυχούς ανίχνευσης, λόγω της χαμηλής τιμής του δείκτη IoU.

Τα μοντέλα Haar και LBP, τα πρώτα σε μεγάλο βαθμό και το δεύτερο σε μικρότερο, επηρεάζονται από την σκιά που σχηματίζει το όχημα στο οδόστρωμα, αναγνωρίζοντάς τη εσφαλμένα ως όχημα, γεγονός που οδήγησε σε αύξηση των FP. Το συγκεκριμένο πρόβλημα φαίνεται να αμβλύνεται μέσω της αύξησης των σταδίων εκπαίδευσης (επιφέροντας όμως ταυτόχρονα τις αρνητικές επιπτώσεις της μείωσης των TP με παράλληλη αύξηση των FN). Επίσης δεν πρέπει να διέλθει της προσοχής ότι το ζήτημα της σκίασης δεν επηρέασε τα μοντέλα HOG_GAB και HOG_SVM. Αναφορικά με την καθετότητα της λήψης και το πώς αυτή επιδρά στην ανίχνευση, φαίνεται ότι το μοντέλο Haar_CORE είναι αυτό που επηρεάζεται λιγότερο. Λόγω όμως του μικρού

αριθμού δειγμάτων που εξετάστηκαν και αφορούν στη συγκεκριμένη περίπτωση, εκτιμάται ότι δεν δύναται να εξαχθούν ασφαλή συμπεράσματα και το ζήτημα πρέπει να διερευνηθεί περαιτέρω.

Τα αποτελέσματα του 6ου κεφαλαίου επιβεβαιώνουν δύο ακόμα στοιχεία που αναφέρονται στη βιβλιογραφία και τα οποία σχετίζονται με τα μοντέλα που εξετάστηκαν. Το πρώτο αφορά στο γεγονός ότι η δυνατότητα ανίχνευσης, δεν επηρεάζεται από την κλίμακα των οχημάτων (αριθμό pixel που καταλαμβάνουν οι διαστάσεις των οχημάτων), εφόσον το μέγεθος αυτών είναι μεγαλύτερο από το παράθυρο που χρησιμοποιήθηκε για την εκπαίδευσή τους (Beyeler, 2015). Έτσι, το σύνολο των μοντέλων ανταποκρίθηκε επιτυχώς ανιχνεύοντας οχήματα ποικίλων διαστάσεων (scale invariant). Το δεύτερο αναφέρεται στην επιβεβαίωση του μειονεκτήματος των cascade και HOG_SVM μοντέλων, αναφορικά με την ανίχνευση αντικειμένων η διεύθυνση των οποίων είναι διαφορετική από αυτή που χρησιμοποιήθηκε στο δείγμα εκπαίδευσης (ibid). Το μοναδικό μοντέλο που διατήρησε δυνατότητα ανίχνευσης των οχημάτων στις οριακές τιμές αυτής της διαφοράς ($\pm 10^\circ$ σύμφωνα με τη βιβλιογραφία), ήταν το HOG_SVM. Αν και στη βιβλιογραφία αναφέρεται η απαίτηση εκπαίδευσης σε 18 διαφορετικές διευθύνσεις, προκειμένου να μην εξαρτάται η ανίχνευση του αντικειμένου από τη διεύθυνσή του (orientation invariance), τα αποτελέσματα των μετρήσεων δείχνουν ότι οι αντίθετες μεταξύ τους διευθύνσεις μπορούν να συγχωνευτούν στην ίδια ομάδα εκπαίδευσης¹³⁵, παρέχοντας ικανοποιητική απόδοση, όπως φαίνεται και από τα αποτελέσματα των μετρήσεων του 6ου κεφαλαίου.

Η εκπαίδευση όσο και η εφαρμογή της διαδικασίας της ανίχνευσης, των αλγορίθμων που εξετάστηκαν, δεν στηρίζονται στην χρωματική πληροφορία αλλά στην πληροφορία που προκύπτει από τα χαρακτηριστικά (features) των οχημάτων σε σχέση με το φόντο (πχ γωνίες, ένταση της χρωματικής διαφοράς των pixel και ρυθμός μεταβολής της κλπ). Εκτιμάται ότι αυτός είναι και ο λόγος που τα μοντέλα HOG_GAB/SVM ανίχνευσαν οχήματα και σε φωτογραφίες που είχαν ληφθεί στο υπέρυθρο φάσμα, παρά το γεγονός ότι δεν είχε τροφοδοτηθεί η εκπαιδευτική

¹³⁵ Τα οχήματα που χρησιμοποιήθηκαν για την εκπαίδευση των μοντέλων της μεταπτυχιακής διατριβής είχαν διεύθυνση $0^\circ (\pm 10^\circ)$ και $180^\circ (\pm 10^\circ)$.

διαδικασία με ανάλογες εικόνες.

Η απόδοση των μοντέλων στην ανίχνευση αντικειμένων, σχετίζεται με τις διαστάσεις της φωτογραφίας (απαραίτητη προϋπόθεση είναι το μέγεθος που καταλαμβάνει το αντικείμενο να είναι μεγαλύτερο του ελάχιστου κυλιόμενου παραθύρου). Ειδικότερα, μειούμενου του παράγοντα διάσταση, μειώνεται και η απόδοση. Αν και οι μικρότερες διαστάσεις οδηγούν σε μείωση της τιμής των FP, εντούτοις αυτή η μείωση δεν δύναται να αντισταθμίσει τόσο την μείωση του αριθμού των TP όσο και την αύξηση των FN. Τη μεγαλύτερη αντοχή σε αυτή τη μείωση επέδειξαν τα cascade μοντέλα Haar και LBP των 20 σταδίων εκπαίδευσης (πλην του Haar_ALL_20stages ο δείκτης F1 του οποίου μειώθηκε κατά 21,3%). Στον αντίποδα βρίσκονται τα μοντέλα 25 σταδίων εκπαίδευσης τα οποία επέδειξαν μικρότερη αντοχή. Έτσι η απόδοση του LBP_GAB_25stages μειώθηκε κατά 39%, ενώ στο ίδιο μοντέλο, η μείωση της παραμέτρου MFAR κατά 0,05 (MFAR=0,45) οδήγησε σε περαιτέρω μείωση της απόδοσης κατά 13% (52% συνολική μείωση). Τέλος, τα μοντέλα HOG_GAB και SVM αν και επηρεάστηκαν λιγότερο από όλα από τη μείωση των διαστάσεων, εν τούτοις δεν μπορεί να γίνει μία άμεση σύγκριση με τα υπόλοιπα καθόσον, για τους λόγους που σχετίζονται με το παράθυρο εκπαίδευσης και τις διαστάσεις των οχημάτων, τα αποτελέσματα των HOG_GAB και SVM προέρχονται από εικόνες 960x540 οι οποίες όμως κατά την είσοδό τους στον αλγόριθμο επεξεργασίας τροποποιήθηκαν σε 1280x720.

Η τιμή του acceptance ratio, που δίνεται με την ολοκλήρωση της εκπαίδευσης, φαίνεται να παρουσιάζει αποκλίσεις, μεταξύ των διαφορετικών μοντέλων (Haar, LBP, HOG), αναφορικά με το ποια μπορεί είναι η βέλτιστη τιμή της, ώστε να παρέχει ταυτόχρονα όχι μόνο επαρκή βαθμό περιγραφικότητας αλλά και γενίκευσης. Αυτή η τιμή φαίνεται να είναι της τάξης του 10^{-6} για τα Haar cascades και 10^{-5} για τα LBP και HOG cascades. Αυτές οι τιμές επιτεύχθηκαν με 20 στάδια εκπαίδευσης. Τα μοντέλα των 25 σταδίων παρουσίασαν πολύ μικρότερες τιμές, γεγονός που οδήγησε, σε υπερπροσαρμογή τους στο μοντέλο εκπαίδευσης, όπως φαίνεται και από τα αποτελέσματα της σύγκρισης της απόδοσης των μοντέλων μεταξύ των περιοχών 1-2 (δείγμα εκπαίδευσης) και 3-4. Στεκόμενοι στο ζήτημα της σύγκρισης της απόδοσης των μοντέλων μεταξύ των παραπάνω περιοχών, φαίνεται ότι ο LBP_GAB 20 σταδίων, παρά το ότι διέθετε την βέλτιστη τιμή acceptance ratio (της τάξης του 10^{-5} , η οποία και συνίσταται από τη

βιβλιογραφία για αποφυγή υπερπροσαρμογής), η απόδοση του μεταξύ περιοχών 1-2 και 3-4 παρουσίασε διαφορά της τάξης 21%. Αυτό αποτελεί ένδειξη ότι το μοντέλο LBP, σε αντίθεση με τα μοντέλα Haar και HOG, δε διαθέτει χαρακτηριστικά "scene invariant". Ιδιαίτερος τα μοντέλα HOG_GAB και HOG_SVM επέδειξαν τα καλύτερα χαρακτηριστικά γενίκευσης καθόσον ανταπεξήλθαν ικανοποιητικά στη δοκιμασία της ανίχνευσης οχημάτων προερχόμενα από διαφορετικά datasets, κάθε ένα από τα οποία συνιστούσε και διαφορετική πρόκληση. Αυτές οι προκλήσεις αφορούσαν : (1) στην απαίτηση για μεγέθυνση των φωτογραφιών προκειμένου η κλίμακα των οχημάτων να ανταποκρίνεται σε αυτή της εκπαίδευσης, (2) στη χρήση εικόνων ιδιαίτερα χαμηλής ανάλυσης (πχ 257x257) και (3) στη χρήση εικόνων που λήφθηκαν στο IR φάσμα.

Από τα αποτελέσματα καθίσταται εμφανές ότι την βέλτιστη απόδοση μεταξύ των μοντέλων, παρουσίασαν το HOG_GAB 20 σταδίων εκπαίδευσης και το HOG_SVM. Αυτά επέδειξαν την καλύτερη συμπεριφορά σε όλους τους δείκτες και συγκρίσεις πλην αυτής του χρόνου επεξεργασίας της φωτογραφίας. Ιδιαίτερος ο HOG_SVM, παρά τη μικρή χρονική διάρκεια της εκπαίδευσης (107min) αλλά και του μικρού αριθμού δείγματος για την πραγματοποίησή της (annotation 464 οχημάτων από 115 εικόνες), διατήρησε την τιμή του δείκτη F1 άνω του 0,948 σε όλες τις περιπτώσεις πλην αυτής της χρησιμοποίησης δείγματος εικόνων διαστάσεων 960x540, που είχαν υποστεί συμπίεση της τάξης του 40% (F1=0,855). Πάρα τούτα όμως, για την εργασία της ιχνηλάτισης ως προσφορότερο μοντέλο κρίνεται το HOG_GAB λόγω του σημαντικά μικρότερου χρόνου που δαπανά για την ανίχνευση (περίπου 3,3 φορές ταχύτερος). Η αδυναμία χρήσης του από τον αλγόριθμο ιχνηλάτισης, για τους λόγους που αναφέρονται στην ενότητα 6.3, όπως θα αναφερθεί και στη συνέχεια, αποτέλεσε καθοριστικό παράγοντα για το σχεδιασμό λειτουργίας της όλης διάταξης (καταγραφή video από Ε/Π ή drone - streaming σε σταθμό εδάφους - ανίχνευση) σε σχεδόν πραγματικό χρόνο.

Έκπληξη αποτέλεσαν τα αποτελέσματα των μετρήσεων της ταχύτητας ροής δεδομένων (upload) σε διαφορετικά ύψη πτήσης και περιοχές. Η πόλη του Βόλου, παρά το γεγονός ότι υποστηρίζεται από δίκτυο 4G, εν τούτοις, οι μετρηθείσες ταχύτητες του δικτύου ήταν μικρότερες σε σχέση με τις αντίστοιχες στην περιοχή των Δήμων Ρήγα Φεραίου και Κιλλελέρ-Αρμενίου. Εκτιμάται ότι αυτό οφείλεται σε δύο παράγοντες: (1) στην πυκνότητα πληθυσμού της περιοχής του Βόλου η οποία

ευθύνεται για τη μείωση του διαθέσιμου εύρους ζώνης, λόγω των αυξημένων απαιτήσεων για ικανοποίηση των χρηστών αναφορικά με τη διακίνηση δεδομένων και (2) στο διαφορετικό σχεδιασμό κάλυψης με κεραιές κινητής τηλεφωνίας των δύο αυτών περιοχών. Αναφορικά με τη δεύτερη περίπτωση, εκτιμάται ότι λόγω του αστικού ιστού, η περιοχή του Βόλου καλύπτεται σε μεγάλο βαθμό με μικροκυψέλες¹³⁶ ενώ οι περιοχές του Βελεστίου και Κιλκελέρ, λόγω του ότι είναι αραιοκατοικημένες, από μακροκυψέλες. Ο καθ' ύψος προσανατολισμός των δεύτερων είναι πιο ευρύς προκειμένου να καλύπτουν μεγαλύτερη εδαφική έκταση (footprint).

Αναφορικά με το θέμα της πιθανής μείωσης της ταχύτητας μεταφοράς δεδομένων μέσω του δικτύου κινητής τηλεφωνίας, όταν το σύστημα αποστολής και λήψης δεδομένων βρίσκεται στην εγγύτητα πηγών θορύβου όπως οι σερβοκινητήρες των drone, διαπιστώθηκε η μέση τιμή των μετρήσεων upload/download με το DJI Phantom σε λειτουργία, ήταν υψηλότερη από τις αντίστοιχες με το αερόχημα εκτός λειτουργίας. Επομένως, συμπεραίνεται ότι ο θόρυβος που προκαλείται από το μηχανισμό των σερβοκινητήρων δεν επιδρά στην ταχύτητα μεταφοράς δεδομένων.

Η ροή video ρυθμιζόμενη στα 512 Kbps (upload) αποδείχτηκε αρκετά εύρωστη σε όλες τις περιπτώσεις χρησιμοποίησής της (δοκιμές video streaming από το Ε/Π προς το σταθμό εδάφους και end-to-end λειτουργία της όλης διάταξης με video streaming από το drone ή το Ε/Π). Ειδικότερα κατά την τελευταία περίπτωση (ροή video από το Ε/Π), διαπιστώθηκε ότι αυτή ήταν συνεχής καθ' όλη τη διάρκεια της πτήσης, παρά το γεγονός ότι προφίλ αυτής περιλάμβανε αλλαγές ύψους, ταχύτητας, ίχνους πτήσης και συνεχείς απο-προσγειώσεις του Ε/Π. Όμως, δεν θα πρέπει να διαφεύγει της προσοχής, ότι σύμφωνα με τις μετρήσεις το δίκτυο μπορεί να προσφέρει πολύ καλύτερες ταχύτητες (έως και x7). Γι' αυτό θα πρέπει να εξεταστούν και εφαρμοστούν μέθοδοι όπως το adaptive streaming, μέσω του οποίου θα ρυθμίζεται η ταχύτητα ροής του video (upload bitrate), σύμφωνα με το διαθέσιμο εύρος του δικτύου (Antonioni et al., 2015).

Το μοντέλο HOG_SVM, παρά την κακή ποιότητα εικόνας της κάμερας Module v2 IMX219 του Raspberry, η οποία συνοδεύτηκε με χαμηλή ταχύτητα ροής δεδομένων

¹³⁶ Για παράδειγμα, το κέντρο της πόλης του Βόλου σε μία έκταση με κέντρο τη συμβολή των οδών Καρτάλη-Αναλήψεως, καλύπτεται από 18 κεραιές της Cosmote (<https://keraies.eett.gr/>).

(512Kbps), αποδείχτηκε αρκετά ανθεκτικό στην ανίχνευση οχημάτων κατά τις end-to-end δοκιμές με ροή video από το DJI Phantom και το Ε/Π. Ειδικά για την περίπτωση της δοκιμής με το DJI Phantom, αξίζει να σημειωθεί ότι αυτή πραγματοποιήθηκε 30min μετά τη δύση του ηλίου, γεγονός που είχε ως αποτέλεσμα οι χαμηλές συνθήκες φωτισμού να επιδράσουν αρνητικά στην ποιότητα του video.

Όπως αναφέρθηκε σε προηγούμενη παράγραφο, ο χρόνος σάρωσης της εικόνας για ανίχνευση οχημάτων από το μοντέλο HOG_SVM ήταν το μοναδικό μειονέκτημα που διαπιστώθηκε, πλην όμως είναι εξαιρετικής σημασίας καθόσον επιδρά άμεσα στην προσπάθεια επίτευξης ανίχνευσης και ιχνηλάτισης οχημάτων, σε σχεδόν πραγματικό χρόνο. Λόγω του χρόνου που δαπανά για την ανίχνευση οχημάτων το συγκεκριμένο μοντέλο, είναι απαραίτητος ο συνδυασμός του με αλγόριθμο ιχνηλάτησης. Έτσι, προκειμένου να αντιμετωπιστεί το συγκεκριμένο μειονέκτημα του μοντέλου, απαιτείται η ρύθμιση εκτέλεσης ανίχνευσης σε χρονικά διαστήματα η διάρκεια των οποίων είναι τριπλάσια σε σχέση με τον HOG_GAB. Η ρύθμιση του μοντέλου, κατά τις δοκιμές με τα εναέρια μέσα, να εκτελεί ανίχνευση μετά την παρέλευση 75 πλαισίων video, είχε ως αποτέλεσμα να μην εντοπιστεί αριθμός οχημάτων τα οποία παρέμεναν εντός του οπτικού πεδίου της κάμερας για χρόνο μικρότερο των 3sec (αφορά κατά βάση τα οχήματα τα οποία κινούνται σε αντίθεση κατεύθυνση σε σχέση με το εναέριο μέσο που χρησιμοποιείται για τη λήψη video). Επομένως, η εφαρμογή της συγκεκριμένης διάταξης επηρεάζεται σε μεγάλο βαθμό από το εύρος του οπτικού πεδίου της κάμερας, από το ύψος πτήσης του Ε/Π ή του drone καθώς επίσης και από την ταχύτητα των οχημάτων και των εναέριων μέσων που βιντεοσκοπούν.

Εν κατακλείδι, τα αποτελέσματα που έχουν καταγραφεί από τις δοκιμές ανίχνευσης και ιχνηλάτισης οχημάτων, σε σχεδόν πραγματικό χρόνο, μέσω video streaming από Ε/Π ή drone, με εκμετάλλευση του υφιστάμενου δικτύου κινητής τηλεφωνίας για τη μετάδοση δείχνουν τη βιωσιμότητα της συγκεκριμένης διάταξης και επομένως ο στόχος της συγκεκριμένης μεταπτυχιακής διατριβής έχει επιτευχθεί.

Κεφάλαιο

8

Επίλογος

Η συγκεκριμένη μεταπτυχιακή διατριβή, στηρίχτηκε στην ιδέα ότι οι πρόσφατες εξελίξεις στον τομέα της πληροφορικής και της κινητής τηλεφωνίας δύνανται να συνδυαστούν προκειμένου να παρέχουν ένα αξιόπιστο σύστημα το οποίο θα είναι ικανό, μέσω της εναέριας επιτήρησης του επίγειου χώρου, να εκτελεί αυτόματη ανίχνευση και ιχνηλάτηση στατικών ή κινούμενων οχημάτων. Η διάταξη η οποία δημιουργήθηκε για το σκοπό αυτό περιέλαβε : (1) την τοποθέτηση σε εναέριο μέσο (E/Π ή drone) μονάδας Raspberry, στην οποία ενσωματώθηκε κάμερα υψηλής ανάλυσης (2) την ασύρματη μετάδοση του καταγραφόμενου video σε ένα σταθμό εδάφους (3) Την εφαρμογή μοντέλων computer vision, σε Η/Υ του σταθμού εδάφους, για την σε σχεδόν πραγματικό χρόνο ανίχνευση και ιχνηλάτηση των οχημάτων.

Κατά τη διάρκεια της έρευνας, ένας αριθμός μοντέλων μηχανικής όρασης εκπαιδεύτηκε και αξιολογήθηκε στη διαδικασία της ανίχνευσης και ιχνηλάτησης οχημάτων. Τα μοντέλα αυτά εφαρμόζουν όχι μόνο διαφορετικές μεθοδολογίες εντοπισμού και καταγραφής των χαρακτηριστικών σημείων των οχημάτων (Haar, LBP και HOG) αλλά και διαφορετικές τεχνικές κατηγοριοποίησης (Gentle Ada Boost, Support Vector Machine). Για την εκπαίδευση και αξιολόγηση των μοντέλων πραγματοποιήθηκε καταγραφή video από τέσσερις διαφορετικές, ως προς τα χαρακτηριστικά τους, περιοχές (αγροτική περιοχή, αστικός ιστός, προ-αστιακή περιοχή, περιφερειακή οδός). Το σύνολο των καταγραφών έχει αποθηκευτεί και οργανωθεί κατάλληλα προκειμένου να αποτελέσει βάση δεδομένων (Vertical Recorded Vehicles with UAS Data Base - VERVwUAS DB) προς εκμετάλλευση από την ερευνητική κοινότητα.

Τα συγκεκριμένα μοντέλα αφού αξιολογήθηκαν βάσει μίας πληθώρας παραμέτρων, διαπιστώθηκε ότι αυτό που επέδειξε την βέλτιστη απόδοση ήταν το HOG_SVM ακολουθούμενο από το HOG_GAB είκοσι σταδίων. Στη μεταξύ τους σύγκριση,

διαπιστώθηκε ότι ο μόνος παράγοντας στον οποίο υπερέτερουσε το δεύτερο μοντέλο έναντι του πρώτου, ήταν ο χρόνος ανίχνευσης. Αν και διαφορά στην απόδοση μεταξύ των δύο μοντέλων ήταν μικρή, εντούτοις, ο παράγοντας χρόνος, λόγω της απαίτησης δημιουργίας near-real-time μοντέλου, ήταν αυτός που έγειρε τη ζυγαριά υπέρ της επιλογής του HOG_GAB. Όμως, ένα τεχνικής φύσεως ζήτημα που σχετίζεται με την διακοπή υποστήριξης του HOG_GAB σε εκδόσεις OpenCV μεταγενέστερες της 3.4, αποτέλεσε μονόδρομο για την επιλογή του HOG_SVM.

Ταυτόχρονα πραγματοποιήθηκαν και μετρήσεις του δικτύου κινητής τηλεφωνίας προκειμένου να καταγραφεί στατιστικά η ελάχιστη upload ταχύτητα. Στη συνέχεια αφού δοκιμάστηκε η δυνατότητα ροής μαγνητοσκοπημένου video από το Ε/Π προς το σταθμό εδάφους, υλοποιήθηκε η πλήρης ανάπτυξη της διάταξης με εφαρμογή του μοντέλου HOG_SVM που προκρίθηκε από την αξιολόγηση.

Το θετικό πρόσημο που προκύπτει από τα αποτελέσματα των μετρήσεων και οι διαπιστώσεις – συμπεράσματα που εξήχθησαν, παρέχουν γόνιμο έδαφος για πολυδιάστατη συνέχιση της έρευνας στο συγκεκριμένο πεδίο.

Σε πρώτη φάση δύναται να διερευνηθεί η επέκταση των μοντέλων που εξετάστηκαν, στην ανίχνευση οχημάτων σε διαφορετικές διευθύνσεις προκειμένου να υπάρχει δυνατότητα εντοπισμού τους ανεξαρτήτως του προσανατολισμού τους. Η έρευνα θα πρέπει να επικεντρωθεί όχι μόνο στα μοντέλα HOG_GAB και SVM αλλά και στα cascade 20 σταδίων εκπαίδευσης η απόδοση των οποίων δύναται να βελτιωθεί περαιτέρω, με αύξηση του δείγματος εκπαίδευσης.

Η συγκεκριμένη έρευνα δύναται επίσης να συνδυαστεί με τις πρόσφατες εξελίξεις στον τομέα του computer vision, που αφορούν στην είσοδο στο συγκεκριμένο πεδίο των νευρωνικών δικτύων συνέλιξης (Convolutional Neural Networks – CNN), τα οποία διαθέτουν αρχιτεκτονική πολλαπλών στρωμάτων μάθησης (deep learning). Η εν λόγω αρχιτεκτονική δεν επηρεάζεται από μικρές παραμορφώσεις, αλλαγές θέσης και κλίμακας ενώ επιδεικνύει ταυτόχρονα και ένα βαθμό ανθεκτικότητας στο ζήτημα του προσανατολισμού των αντικειμένου (rotation invariance).

Επιπλέον, λόγω της δυνατότητας που υπάρχει να προβλεφθούν οι διαστάσεις των

οχημάτων στο λαμβανόμενο video (WxH pixels), βάσει του σχετικού ύψους πτήσης¹³⁷ της εναέριας πλατφόρμας (UAS, A/Φ-E/Π) σε συνδυασμό με το χρησιμοποιούμενο FoV της κάμερας και την ανάλυση του video, προτείνεται να μελετηθεί η αντικατάσταση του αλγόριθμου NMS με ένα νέο ο οποίος θα απορρίπτει bounding boxes, οι διαστάσεις των οποίων θα ξεπερνούν μία μέγιστη τιμή. Αυτό θα οδηγήσει σε περεταίρω μείωση των FN καθώς και στη βελτίωση της τιμής του IoU.

Επίσης, χρήζει περεταίρω μελέτης η επίδραση του ύψους πτήσης στην ταχύτητα μεταφοράς δεδομένων μέσω του δικτύου κινητής τηλεφωνίας, προκειμένου να εξαχθούν ασφαλή συμπεράσματα. Η διαφαινόμενη μείωση της ταχύτητας μετάδοσης δεδομένων, αυξανόμενου του ύψους πτήσης, σε συνδυασμό με τις μεγάλες μεταβολές που παρουσιάζει αυτή, βάσει της θέσης της πλατφόρμας μετάδοσης αλλά του διαθέσιμου εύρους ζώνης, καθιστά ίσως περισσότερο ελκυστική τη λύση εκμετάλλευσης των δορυφορικών επικοινωνιών.

Τέλος η διάταξη που εξετάστηκε στη συγκεκριμένη μεταπτυχιακή διατριβή δύναται να αναβαθμιστεί μέσω της χρήσης κάμερας με δυνατότητες περιστροφής (gimbal), ο έλεγχος της οποίας θα πραγματοποιείται από το έδαφος. Αυτό συνεπάγεται αμφίδρομη ροή δεδομένων και κατακόρυφη αναβάθμιση των δυνατοτήτων ιχνηλάτισης.

¹³⁷ Ως σχετικό ύψος πτήσης ορίζεται η απόσταση του E/Π από το έδαφος. Η συγκεκριμένη απόσταση εξάγεται είτε με τη χρήση ραδιούψομέτρων είτε με συνδυασμό πληροφοριών μεταξύ του δέκτη του GPS και μοντέλου ψηφιακού χάρτη.

Βιβλιογραφία

- Adobe, S. I. (2004). A digital video primer. Retrieved from http://www.cs.ubc.ca/~cs544/Fall2005/dvprimer_2004.pdf
- Adobe, S. I. (2006). A digital video primer: An introduction to DV production, post-production, and delivery. Retrieved from https://vision.unipv.it/mst/mat/dvprimer_0306.pdf
- Agarwal, S., Awan, A., & Roth, D. (2004). Learning to detect objects in images via a sparse, part-based representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(11), 1475-1490.
- Ahonen, T., Hadid, A., & Pietikainen, M. (2006). Face description with local binary patterns: Application to face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(12), 2037-2041.
- Alpaydin, E. (2010). Introduction to machine learning. sl.
- Antoniou, Z., Panayides, A., Pattichis, M. S., Stavrou, S., Kyriacou, E., Spanias, A., Pattichis, C. S. (2015). (2015). Adaptive emergency scenery video communications using HEVC for responsive decision support in disaster incidents. Paper presented at the *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, 173-176.
- Apostolopoulos, J. G., Tan, W., & Wee, S. J. (2002). Video streaming: Concepts, algorithms, and systems. *HP Laboratories, Report HPL-2002-260*,
- Aragon-Zavala, A. (2008). *Antennas and propagation for wireless communication systems* John Wiley & Sons.
- Asadpour, M., Van den Bergh, B., Giustiniano, D., Hummel, K., Pollin, S., & Plattner, B. (2014). Micro aerial vehicle networks: An experimental analysis of challenges and opportunities. *IEEE Communications Magazine*, 52(7), 141-149.
- Austin, R. (2011). *Unmanned aircraft systems: UAVS design, development and deployment* John Wiley & Sons.
- Balzano, Q., Garay, O., & Manning, T. J. (1995). Electromagnetic energy exposure of simulated users of portable cellular telephones. *IEEE Transactions on Vehicular Technology*, 44(3), 390-403.

- Benes, C. Video streaming primer. Retrieved from http://www.bogotobogo.com/VideoStreaming/Files/Video_streaming_etc/Video%20Streaming%201.12.pdf
- Berners-Lee, T. J. (1992). The world-wide web. *Computer Networks and ISDN Systems*, 25(4-5), 454-459.
- Beyeler, M. (2015). *OpenCV with python blueprints* Packt Publishing Ltd.
- Bhaskar, R. (2017). [Image processing] haar cascades explained in 2 minutes! Retrieved from <http://sra.vjti.info/blog/blog-posts/image-processing-haar-cascades-explained-in-2-minutes>
- Udacity, I. (Producer), & Bobick, A. (Director). (2011–2018). [Video/DVD] USA: Retrieved from <https://classroom.udacity.com/courses/ud810/lessons/3064888962/concepts/30999389800923>
- Boden, M. A. (2006). *Mind as machine: A history of cognitive science* Clarendon Press.
- Breckon, T. P., Barnes, S. E., Eichner, M. L., & Wahren, K. (2009). (2009). Autonomous real-time vehicle detection from a medium-level UAV. Paper presented at the *Proc. 24th International Conference on Unmanned Air Vehicle Systems*, 29.1-29.9.
- Briscoe, B., Odlyzko, A., & Tilly, B. (2006). Metcalfe's law is wrong-communications networks increase in value as they add members-but by how much? *IEEE Spectrum*, 43(7), 34-39.
- Brophy, J. (2012). John francis mitchell biography . Retrieved from <http://www.brophy.net/PivotX/?p=john-francis-mitchell-biography>
- Canu, S. (2014). SVM and kernel machines: Linear and non -linear classification. Paper presented at the *Ocean's Big Data Mining*, Brest (France).
- Cao, X., Lan, J., Yan, P., & Li, X. (2012). Vehicle detection and tracking in airborne videos by multi-motion layer analysis. *Machine Vision & Applications*, 23(5)
- Circular, I. (2011). 328, unmanned aircraft systems (UAS). *Montreal, Canada: International Civil Aviation Organization (ICAO)*,
- Collins, R. Harris corner detector. Retrieved from <http://www.cse.psu.edu/~rtc12/CSE486/lecture06.pdf>
- Colomina, I., & Molina, P. (2014). Unmanned aerial systems for photogrammetry and remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 92, 79-97.

- Cosmote. (2018). Κάλυψη δικτύου. Retrieved from <https://www.cosmote.gr/mobile/cosmoportal/page/T25/section/Cover>
- Dalal, N., & Triggs, B. (2005). (2005). Histograms of oriented gradients for human detection. Paper presented at the *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference On*, , 1 886-893.
- Danelljan, M., Häger, G., Khan, F., & Felsberg, M. (2014). (2014). Accurate scale estimation for robust visual tracking. Paper presented at the *British Machine Vision Conference, Nottingham, September 1-5, 2014*,
- Davies, E. R. (2004). *Machine vision: Theory, algorithms, practicalities* Elsevier.
- Davies, E. R. (2012). *Computer and machine vision: Theory, algorithms, practicalities* Academic Press.
- De Vriendt, J., Lainé, P., Lerouge, C., & Xu, X. (2002). Mobile network evolution: A revolution on the move. *IEEE Communications Magazine*, 40(4), 104-111.
- Divya, J. (2017). Commercial unmanned aerial vehicle (UAV) market analysis - industry trends, companies and what you should know. Retrieved from <http://www.businessinsider.com/commercial-uav-market-analysis-2017-8>
- Do, T. T., Hua, K. A., & Tantaoui, M. A. (2004). (2004). P2VoD: Providing fault tolerant video-on-demand streaming in peer-to-peer environment. Paper presented at the *Communications, 2004 IEEE International Conference On*, , 3 1467-1472.
- Dong, M., Chen, B. M., Cai, G., & Peng, K. (2007). Development of a real-time onboard and ground station software system for a UAV helicopter. *Journal of Aerospace Computing, Information, and Communication*, 4(8), 933-955.
- Doost, H. R. E., & Rahebi, J. (2012). An efficient method for texture classification with local binary pattern based on wavelet transformation. *International Journal of Engineering Science and Technology*, 4(12), 4881-4885.
- Drachman, D. A. (2005). Do we have brain to spare? *Neurology*, 64(12), 2004-2005. doi:64/12/2004 [pii]
- EPIC Systems. (2017). **Quick history of machine vision**. Retrieved from <https://www.epicsysinc.com/blog/machine-vision-history>

- ETSI Mobile Competence Centre. (8 - 10 December, 2004). (8 - 10 December, 2004). Overview of 3GPP release 4 summary of all release 4 features v. TSG #26. Paper presented at the Vouliagmeni Athens, Greece. (RP - 040548)
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861-874.
- Ferreira, A. J., & Figueiredo, M. A. (2012). Boosting algorithms: A review of methods, theory, and applications. *Ensemble machine learning* (pp. 35-85) Springer.
- Fettweis, G., & Alamouti, S. (2014). 5G: Personal mobile internet beyond what cellular did to telephony. *IEEE Communications Magazine*, 52(2), 140-145.
- Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1), 119-139.
- Friedman, J., Hastie, T., & Tibshirani, R. (2000). Additive logistic regression: A statistical view of boosting (with discussion and a rejoinder by the authors). *The Annals of Statistics*, 28(2), 337-407.
- Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The elements of statistical learning* Springer series in statistics New York.
- Gaszczak, A., Breckon, T. P., & Han, J. (2011). Real-time people and vehicle detection from UAV imagery.
- Ghosh, A., Zhang, J., Andrews, J. G., & Muhamed, R. (2010). *Fundamentals of LTE* Pearson Education.
- Gleason, J., Nefian, A. V., Bouysounousse, X., Fong, T., & Bebis, G. (2011). (2011). Vehicle detection from aerial imagery. Paper presented at the *Robotics and Automation (ICRA), 2011 IEEE International Conference On*, 2065-2070.
- Goebel, G. (2017). European helicopter pioneers. Retrieved from http://www.airvectors.net/avheli_1.html
- Great Images in NASA. (2009). First picture from explorer VI satellite. Retrieved from <http://grin.hq.nasa.gov:80/copyright.html>
- Gualdi, G., Prati, A., & Cucchiara, R. (2008). Video streaming for mobile video surveillance. *IEEE Transactions on Multimedia*, 10(6), 1142-1154.

- Hadid, A., Pietikainen, M., & Ahonen, T. (2004). (2004). A discriminative feature space for detecting and recognizing faces. Paper presented at the *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference On*, , 2 II-II.
- Haralick, R. M., & Shanmugam, K. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, (6), 610-621.
- Harris, C., & Stephens, M. (1988). (1988). A combined corner and edge detector. Paper presented at the *Alvey Vision Conference*, , 15(50) 10.5244.
- Heikell, J. (2005). *Electronic warfare self-protection of battlefield helicopters: A holistic view* Helsinki University of Technology.
- Hossain, E., Rasti, M., Tabassum, H., & Abdelnasser, A. (2014). Evolution toward 5G multi-tier cellular wireless networks: An interference management perspective. *IEEE Wireless Communications*, 21(3), 118-127.
- Howse, J., Puttemans, S., Hua, Q., & Sinha, U. (2015). *OpenCV 3 blueprints* Packt Publishing Ltd.
- Huang, T. (1996). *Computer vision: Evolution and promise*.
- ITU-R. (2008). (2008). Requirements related to technical performance for IMT-advanced radio interface(s). Paper presented at the , *Rep. ITU-R M.2134*
- Jack, K. (2011). *Video demystified: A handbook for the digital engineer* Elsevier.
- Jain, R., Kasturi, R., & Schunck, B. G. (1995). *Machine vision* McGraw-Hill New York.
- Klette, R. (2014). *Concise computer vision* Springer.
- Kohavi, R., & Provost, F. (1998). Confusion matrix. *Machine Learning*, 30(2-3), 271-274.
- Kokaram, A. Introduction to digital video processing image and video processing. Retrieved from http://www.mee.tcd.ie/~sigmedia/pmwiki/uploads/Teaching.4S1b/handout8_4s1.pdf
- Krig, S. (2016). Ground truth data, content, metrics, and analysis. *Computer vision metrics* (pp. 247-271) Springer.
- Kuliah, M. (2010). Lecture 04 : Computer vision - edge detection lecture T0283. Retrieved from <http://slideplayer.com/slide/9843289/>

- Kumar, R., Sawhney, H., Samarasekera, S., Hsu, S., Tao, H., Guo, Y., . . . Hirvonen, D. (2001). Aerial video surveillance and exploitation. *Proceedings of the IEEE*, 89(10), 1518-1539.
- Kurose, J.F : Ross, K.W. (2012). University of calgary: Courses/computer science/CPSC 441.W2014/chapter 7: Multimedia networking. Retrieved from [http://wiki.ucalgary.ca/page/Courses/Computer Science/CPSC 441.W2014/Chapter 7: Multimedia Networking.html](http://wiki.ucalgary.ca/page/Courses/Computer%20Science/CPSC%20441.W2014/Chapter%207%20Multimedia%20Networking.html)
- Kyrkou, C. (2017). Object detection using local binary patterns. Retrieved from <file:///C:/Users/user/Desktop/Local%20Binary%20Patterns/Object%20Detection%20Using%20Local%20Binary%20Patterns%20%E2%80%93%20Chris-tos%20Kyrkou%20%E2%80%93%20Medium.html>
- Lempiäinen, J., & Manninen, M. (2002). *Radio interface system planning for GSM/GPRS/UMTS* Springer Science & Business Media.
- Liu, K., & Mattyus, G. (2015). Fast multiclass vehicle detection on aerial images. *IEEE Geoscience and Remote Sensing Letters*, 12(9), 1938-1942.
- Liu, L., Zhao, L., Long, Y., Kuang, G., & Fieguth, P. (2012). Extended local binary patterns for texture classification. *Image and Vision Computing*, 30(2), 86-99.
- Luigi De Russis, Alberto Sacco. (2016). Face detection and tracking. Retrieved from <http://opencv-java-tutorials.readthedocs.io/en/latest/06-face-detection-and-tracking.html>
- Maggiori, E., Tarabalka, Y., Charpiat, G., & Alliez, P. (2017). (2017). Can semantic labeling methods generalize to any city? the INRIA aerial image labeling benchmark. Paper presented at the *IEEE International Symposium on Geoscience and Remote Sensing (IGARSS)*,
- Maier, M., Blakemore, C., & Koivisto, M. (2000). The health hazards of mobile phones. *BMJ (Clinical Research Ed.)*, 320(7245), 1288-1289.
- Mallick, S. (2017). Object tracking using OpenCV (C++/python). Retrieved from <https://www.learnopencv.com/object-tracking-using-opencv-cpp-python/>
- Miniwatts Marketing Group. (2017). Internet growth statistics. Retrieved from <http://www.internetworldstats.com/emarketing.htm>

- Moses, Y., Adini, Y., & Ullman, S. (1994). Face recognition: The problem of compensating for changes in illumination direction. *Computer Vision—ECCV'94*, , 286-296.
- Northstar Gallery. (1998). Aviation and aerial photography : A brief history of aerial photography. Retrieved from <http://northstargallery.com/aerialphotography/history%20aerial%20photography/history.htm>
- Ohm, J.:Sullivan G.:Tan T.: Wiegand T. (2012). Comparison of the coding efficiency of video coding standards - including high efficiency video coding (HECV). *IEEE Trans. Circuits and Systems for Video Tech*, 22(12), 1669-1694.
- Ojala, T., Pietikainen, M., & Harwood, D. (1994). (1994). Performance evaluation of texture measures with classification based on kullback discrimination of distributions. Paper presented at the *Pattern Recognition, 1994. Vol. 1-Conference A: Computer Vision & Image Processing., Proceedings of the 12th IAPR International Conference On*, , 1 582-585.
- Ojala, T., Pietikäinen, M., & Harwood, D. (1996). A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition*, 29(1), 51-59.
- Ojala, T., Pietikainen, M., & Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7), 971-987.
- Oliveira, M., & Santos, V. (2008). Automatic detection of cars in real roads using haar-like features. *Department of Mechanical Engineering, University of Aveiro, 3810*
- OpenCV by Doxygen. (2018). Cascade classifier training. Retrieved from https://docs.opencv.org/3.4/dc/d88/tutorial_traincascade.html
- Oxford University Press. (2017). OED online. Retrieved from <http://www.oed.com/viewdictionaryentry/Entry/11125>
- Ponce, J., Hebert, M., Schmid, C., & Zisserman, A. (2007). *Toward category-level object recognition* Springer.
- Puri, A. (2005). A survey of unmanned aerial vehicles (UAV) for traffic surveillance. *Department of Computer Science and Engineering, University of South Florida*, , 1-29.
- Razakarivony, S., & Jurie, F. (2016). Vehicle detection in aerial imagery: A small target detection benchmark. *Journal of Visual Communication and Image Representation*, 34, 187-203.

- Riiser, H. (2013). Adaptive bitrate video streaming over HTTP in mobile wireless networks.
- Rosebrock, A. (2016). Intersection over union (IoU) for object detection. Retrieved from <https://www.pyimagesearch.com/2016/11/07/intersection-over-union-iou-for-object-detection/>
- Rysavy Research, L. (2013). In 4G Americas (Ed.), *Mobile broadband explosion*
- Schapiro, R. E. (1990). The strength of weak learnability. *Machine Learning*, 5(2), 197-227.
- Seitz, S., & Szeliski, R. (1999). Applications of computer vision to computer graphics. *Computer Graphics*, 33(4), 35-37.
- Shan, C., Gong, S., & McOwan, P. W. (2009). Facial expression recognition based on local binary patterns: A comprehensive study. *Image and Vision Computing*, 27(6), 803-816.
- Shapiro, L., & Stockman, G. C. (2001). Computer vision. 2001. Ed: Prentice Hall,
- Simon, S., & Alejandro, A. (2007). Antennas and propagation for wireless communication systems. *John Wiley & Sons*,
- Sivaraman, S., & Trivedi, M. M. (2010). A general active-learning framework for on-road vehicle recognition and tracking. *IEEE Transactions on Intelligent Transportation Systems*, 11(2), 267-276.
- Skrzypietz, B. (2012). Unmanned aircraft systems for civilian missions.
- Soomro, K., Zamir, A. R., & Shah, M. (2012). UCF101: A dataset of 101 human actions classes from videos in the wild. *arXiv Preprint arXiv:1212.0402, CRCV-TR-12-01, November, 2012*
- Srinivasan, S., Latchman, H., Shea, J., Wong, T., & McNair, J. (2004). (2004). Airborne traffic surveillance systems: Video surveillance of highway traffic. Paper presented at the *Proceedings of the ACM 2nd International Workshop on Video Surveillance & Sensor Networks*, 131-135.
- Staelens, N., Moens, S., Van den Broeck, W., Marien, I., Vermeulen, B., Lambert, P., . . . Demeester, P. (2010). Assessing quality of experience of IPTV and video on demand services in real-life environments. *IEEE Transactions on Broadcasting*, 56(4), 458-466.

- Stan, B. (2007). CBCL StreetScenes challenge framework. Retrieved from <http://cbcl.mit.edu/software-datasets/streetscenes/>
- Statista. (2018). **Number of mobile (cellular) subscriptions worldwide from 1993 to 2017 (in millions)**. Retrieved from <https://www.statista.com/statistics/262950/global-mobile-subscriptions-since-1993/>
- Szeliski, R. (2010). *Computer vision: Algorithms and applications* Springer Science & Business Media.
- Tan, P. (2016). Steinbach, michael και kumar, vipin. *Εισαγωγή Στην Εξόρυξη Δεδομένων*,
- Tanner, F., Colder, B., Pullen, C., Heagy, D., Eppolito, M., Carlan, V., . . . Sallee, P. (2009). (2009). Overhead imagery research data set—an annotated data library & tools to aid in the development of computer vision algorithms. Paper presented at the *Applied Imagery Pattern Recognition Workshop (AIPRW), 2009 IEEE*, 1-8.
- Tektronix. (2000). A guide to standard and high-definition digital video measurements. Retrieved from <http://www.appliedelectronics.com/documents/Guide%20to%20Standard%20HD%20Digital%20Video%20Measurements.pdf>
- Thuyen, T. (2017). A dive into histogram of oriented gradients (HOG). Retrieved from <https://trinngocthuyen.github.io/2017-07-13-a-dive-into-hog.html>
- US Joint Chiefs of Staff (Ed.). (15 February 2016). *Dictionary of military and associated terms*. Washington DC: Department of Defense.
- Vali, R., Berber, S., & Nguang, S. (2009). Planning and optimisation of 3G and 4G wireless networks.
- Van Blyenburgh, P. (1999). UAVs: An overview. *Air & Space Europe*, 1(5-6), 43-47.
- Vergouw, B., Nagel, H., Bondt, G., & Custers, B. (2016). Drone technology: Types, payloads, applications, frequency spectrum issues and future developments. *The future of drone use* (pp. 21-45) Springer.
- Viola, P., & Jones, M. (2001). (2001). Rapid object detection using a boosted cascade of simple features. Paper presented at the *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference On*, , 1 I-I.

- Visser, A. E., Noordam, N. M., Epema, H. K., Verschoor, C. R., Kop, R. & Dokoupil, A. (2014). DutchUASDataset001: 577 annotated aerial images of rhinos, zebras, rangers and cars. Retrieved from <http://www.dutchuas.nl/dataset/>
- Vodafone, G. (2018). Χάρτες κάλυψης. Retrieved from <http://www.vodafone.gr/portal/client/cms/viewCmsPage.action?pageId=11123>
- W3C. (2018). Total number of web pages. Retrieved from <http://www.internetlivestats.com/total-number-of-websites/>
- Want, R., Bardram, J., Friday, A., Langheinrich, M., Brush, A., Taylor, A., . . . Dey, A. (2010). Ubiquitous computing fundamentals. *Chapman & Hall/CRC*,
- Want, R., & Pering, T. (2005). (2005). System challenges for ubiquitous & pervasive computing. Paper presented at the *Proceedings of the 27th International Conference on Software Engineering*, 9-14.
- Wasseem, I. Image processing. Retrieved from [https://www.uotechnology.edu.iq/ce/lecture%202013n/4th%20Image%20Processing%20Lectures/DIP Lecture2.pdf](https://www.uotechnology.edu.iq/ce/lecture%202013n/4th%20Image%20Processing%20Lectures/DIP%20Lecture2.pdf)
- William Stallings. (2007). *Data and computers communications* (8th ed.). Upper Saddle River, NJ 07458: Pearson Education, Inc.
- WordPress. (2018). History of aerial photography. Retrieved from <http://www.copterdrone.co.uk/history-of-aerial-photography/>
- Xiao, F. L. (2013). A brief history of intel CPU microarchitecture. Retrieved from [https://people.apache.org/~xli/presentations/history Intel CPU.pdf](https://people.apache.org/~xli/presentations/history%20Intel%20CPU.pdf)
- Xiao, J., Cheng, H., Sawhney, H., & Han, F. (2010). (2010). Vehicle detection and tracking in wide field-of-view aerial video. Paper presented at the *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference On*, 679-684.
- Yang, Y., Liu, F., Wang, P., Luo, P., & Liu, X. (2012). (2012). Vehicle detection methods from an unmanned aerial vehicle platform. Paper presented at the *Vehicular Electronics and Safety (ICVES), 2012 IEEE International Conference On*, 411-415.
- Zisserman, A. (2011). Lecture 2: The SVM classifier. *C19 Machine Learning*, 12

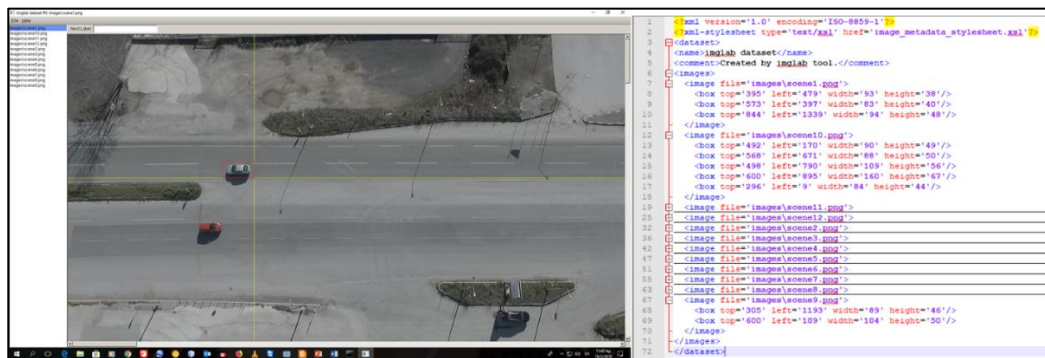
Παράρτημα Α

Annotation Systems

A.1 Imglab

Το Imglab¹ είναι ένα γραφιστικό εργαλείο για την επισήμανση των εικόνων με ένα ορθογώνιο παραλληλόγραμμο. Αποτελεί μέρος της βιβλιοθήκης dlib και για τη μεταγλώττιση (compilation) του προγράμματος χρησιμοποιούνται οι εντολές (King, 2018):

- `cd dlib/tools/implab`
- `mkdir build`
- `cd build`
- `cmake ..`
- `cmake --build . --config Release`



Εικόνα Α-1. Στο αριστερό μέρος φαίνεται το γραφικό περιβάλλον του imglab εν μέσω της διαδικασίας επισήμανσης (annotation) ενός οχήματος. Στο δεξιό μέρος φαίνεται μέρος του αρχείου xml που δημιουργείται στο οποίο περιέχονται το σύνολο των εικόνων που περιλαμβάνει ο αντίστοιχος φάκελος και οι συντεταγμένες (x,y) /διαστάσεις (w,h) των οχημάτων που επισημαίνονται.

Μετά τη μεταγλώττιση του προγράμματος με την εντολή «./implab -c mydataset.xml /tmp/images» δημιουργείται ένα xml αρχείο, το mydataset, στο οποίο έχει καταχωρηθεί μια λίστα με το σύνολο των εικόνων που περιλαμβάνονται στο φάκελο images. Στη

¹ Χρησιμοποιήθηκε η έκδοση v1.13

συνέχεια με την εντολή «./imglab mydataset.xml» ανοίγει το παράθυρο του προγράμματος imglab, το οποίο φαίνεται στην εικόνα A-1. Σε αυτό η επισήμανση των εικόνων γίνεται κρατώντας πατημένο το πλήκτρο shift και σέρνοντας το ποντίκι στο διαγώνιο άξονα του αντικειμένου, έχοντας πατημένο το αριστερό κομβίο αυτού. Στο πεδίο "Next Label" αναγράφεται ο χαρακτηρισμός του επόμενου αντικειμένου που θα επισημανθεί. Επίσης το πρόγραμμα δίνει τη δυνατότητα ονομασίας τμημάτων ενός αντικειμένου, zoom in/out και ομοιομορφοποίησης ιστογράμματος (histogram equalization).

A.2 Object Marker

Το objectmarker είναι ένα γραφιστικό εργαλείο για την επισήμανση των θετικών εικόνων. Είναι γραμμένο σε Microsoft Visual Basic, χρησιμοποιώντας βιβλιοθήκες C++ και OpenCV. Το πρόγραμμα καθορίζει την εισαγωγή των φωτογραφιών που περιέχουν το αντικείμενο ενδιαφέροντος σε ένα φάκελο που ονομάζεται raw data και ο οποίος περιέχεται στην ίδια διεύθυνση με το objectmarker.exe. Οι εικόνες που τοποθετούνται στο raw data θα πρέπει να είναι της μορφής *.bmp (Gunawan, 2006).



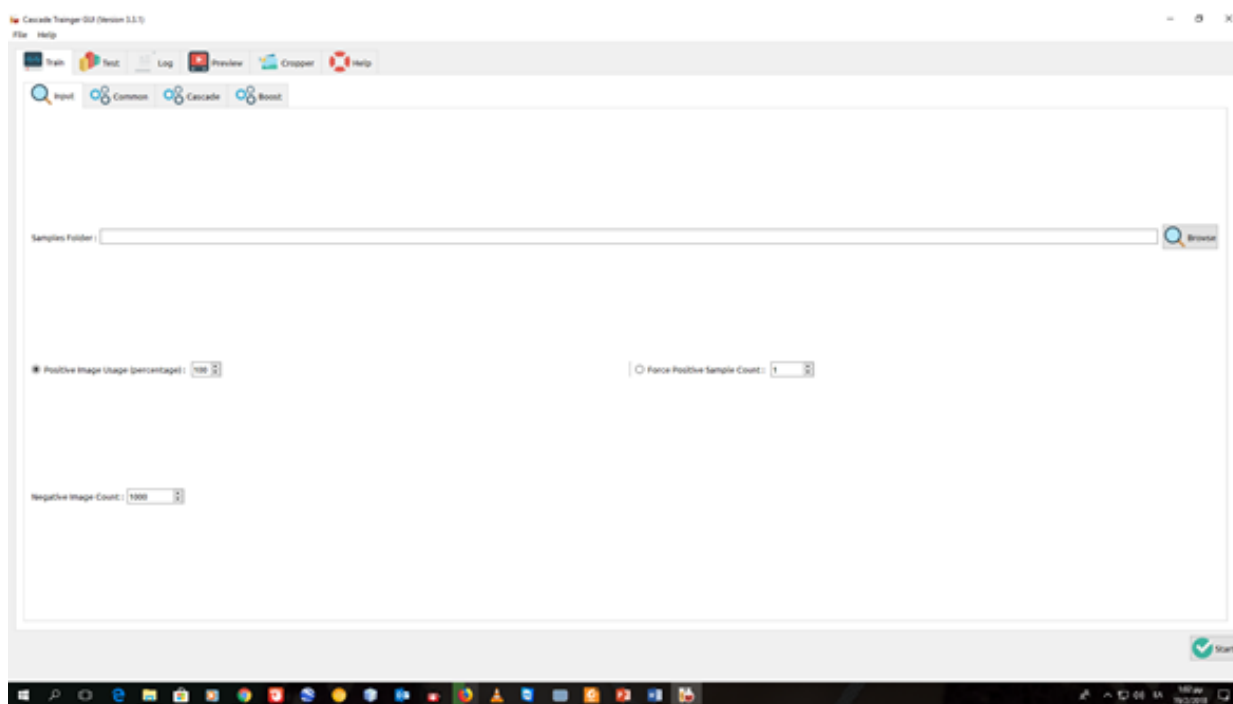
Εικόνα A-2. Γραφικό περιβάλλον του ObjectMarker. Στο πάνω μέρος της εικόνας παρουσιάζεται το γραφικό περιβάλλον του objectmarker και στο κάτω δίνεται ένα απόσπασμα από το txt αρχείο που δημιουργείται.

Μετά την επισήμανση των εικόνων και την έξοδο από το πρόγραμμα, δημιουργείται ένα αρχείο info.txt στο οποίο περιλαμβάνεται λίστα με τις εικόνες (file name), ο αριθμός των αντικειμένων ενδιαφέροντος που επισημάνθηκαν σε κάθε εικόνα, οι συντεταγμένες (x,y)

της θέσης τους και οι διαστάσεις(w,h) του κάθε ενός από αυτά. Το objectmarker, αφού κληθεί μέσω της κονσόλας του Command Prompt, ανοίγει το παράθυρο της εικόνας ... για την επισήμανση των αντικειμένων. Υπάρχουν δύο μέθοδοι επισήμανσης, μία ελευθέρων διαστάσεων (free-size) και μία σταθερής κλίμακας (fixed-scale). Η πρώτη περίπτωση χρησιμοποιείται όταν το ορθογώνιο επισήμανσης δεν επιβάλλεται να έχει συγκεκριμένη αναλογία μήκους – πλάτους (ibid)

A.3 Cascade Trainer GUI

Σύμφωνα με την οικεία ιστοσελίδα (Ahmadl, 2017), το Cascade Trainer GUI² είναι ένα πρόγραμμα που χρησιμοποιείται για την εκπαίδευση, έλεγχο και βελτίωση των cascade classifiers. Χρησιμοποιεί ένα γραφικό περιβάλλον το οποίο καθιστά εύκολη την εισαγωγή των αντίστοιχων παραμέτρων για την εκπαίδευση του κατηγοριοποιητή. Απαιτεί τη χρήση λειτουργικού Windows 7 (ή νεότερου).



Εικόνα A-3. Γραφικό περιβάλλον του Cascade Trainer GUI. Τα παράθυρα Input/Common/Cascade/Boost διαθέτουν τα αντίστοιχα πεδία για την εισαγωγή των παραμέτρων.

Για την εκπαίδευση του κατηγοριοποιητή δέχεται τις παρακάτω παραμέτρους : (1)

² Χρησιμοποιήθηκε η έκδοση v3.3.1 (για την εκπαίδευση σε HOG απαιτεί τη χρήση OpenCV v3.1 ή νεότερου).

αριθμών σταδίων της εκπαίδευσης (2) τιμές και δείκτες για τη μνήμη (buffer size) (3) αριθμό threads (4) διαστάσεις των δειγμάτων (5) επιλογή cascade μεταξύ LBP, HAAR και HOG (6) επιλογή κατηγοριοποιητή ενίσχυσης μεταξύ των Gentle/Discrete και Real AdaBoost (7) ελάχιστο ρυθμό επιτυχούς ανίχνευσης (min hit rate) (8) μέγιστο ρυθμό σφαλμάτων (max false alarm rate) (9) weight trim rate και (10) max weak trees.

Μία σημαντική δυνατότητα του προγράμματος είναι ότι διαθέτει εργαλείο για την περικοπή εκτός από εικόνων και video. Το video εμφανίζεται σε frames στα οποία δίνεται η δυνατότητα περικοπής (crop) του αντικειμένου ενδιαφέροντος. Στη συνέχεια οι θετικές και αρνητικές εικόνες αποθηκεύονται σε ξεχωριστούς φακέλους για την εκπαίδευση του κατηγοριοποιητή.

Παράρτημα Β

Μετρήσεις Ταχύτητας Δικτύου

B.1 Μετρήσεις της Ταχύτητας Δικτύου (UAS)

Συνθήκες – στοιχεία πραγματοποίησης των συγκεκριμένων μετρήσεων:

- Δίκτυο Κινητής Τηλεφωνίας : Cosmote
- Router : HUAWEI 4G USB Stick E3372h-153
- Θέση Μέτρησης : Βόλος (39°21'57''N, 22°55'26''E)
- Ύψος μέτρησης:12μ
- UAS:DJI Phantom pro
- Υπολογιστική συσκευή : Raspberry Pi3 B+ τοποθετημένη στη βάση του DJI Phantom

ΚΑΤΑΣΤΑΣΗ ΛΕΙΤΟΥΡΓΙΑΣ ΤΟΥ UAV	ΤΑΧΥΤΗΤΑ DOWNLOAD	ΤΑΧΥΤΗΤΑ UPLOAD
OFF	3,37	2,56
	4,01	2,11
	3,38	2,30
	3,54	2,57
	3,48	2,20
IDLE	4,03	2,51
	4,21	2,54
	3,86	2,36
	3,24	2,47
	3,89	2,50
FLY	4,06	2,66
	4,13	2,22
	3,83	2,58
	3,66	2,51
	3,70	2,26

Πίνακας B-1. Ρυθμός ροής δεδομένων (Upload/Download) δικτύου κινητής τηλεφωνίας βάσει της κατάστασης λειτουργίας του UAS (ταχύτητα σε Mbps)

B.2 Μετρήσεις της Ταχύτητας Δικτύου (E/Π)

Συνθήκες – στοιχεία πραγματοποίησης των συγκεκριμένων μετρήσεων:

- Δίκτυο Κινητής Τηλεφωνίας : Cosmote
- Router : HUAWEI 4G USB Stick E3372h-153
- Θέσεις Μέτρησης : (1) Α/Δ Στεφανοβικείου (2) Περιοχή Δήμου Κιλελέρ - Αρμενίου (για κάθε περιοχή οι συντεταγμένες προηγούνται του αντίστοιχου πίνακα) (3) Περιοχή Βόλου (4) Διαδρομή Βόλος – Α/Δ Στεφανοβικείου
- Ε/Π:UH-1H
- Υπολογιστική συσκευή : Laptop T510
- Ύψος πτήσης : 500ft, 1000ft και 3000ft
- Ταχύτητα πτήσης : 50Knots και 100 Knots (για κάθε ένα από τ παραπάνω ύψη)

(1) Α/Δ Στεφανοβικείου

ΚΑΤΑΣΤΑΣΗ Ε/Π →	OFF	IDLE	FLY	ΤΡΟΔΟΔΡΟΜΗΣΗ		ΑΙΩΡΗΣΗ ΓΙΑ Α/Γ
ΤΑΧΥΤΗΤΑ ΔΙΚΤΥΟΥ (Download σε Mbps)	13.46	9.97	10.85	7.70	4.41	8.26
ΤΑΧΥΤΗΤΑ ΔΙΚΤΥΟΥ (Upload σε Mbps)	4,01	2,90	3,92	2,73	2,83	3,53

Πίνακας Β-2. Ταχύτητα ροής δεδομένων (Download/Upload) εντός του Α/Δ

(2) Περιοχή Δήμου Κιλελέρ - Αρμενίου

Πραγματοποιήθηκε ελειψοειδές ίχνος πτήσης (orbit) το κέντρο του οποίου είχε συντεταγμένες: 39°30'03''N, 22°41'46''E

500ft/50Kn	500ft/100Kn	1000ft/50Kn	1000ft/100Kn	3000ft/50Kn	3000ft/100Kn
6,25	4,78	4,49	3,96	3,15	0,94
4,75	2,85	6,26	5,15	1,88	1,02
3,10	4,05	4,93	7,02	2,84	1,53
1,14	5,38	3,96	1,34	3,33	2,91
8,64	4,98	3,47	3,34	2,74	1,02
6,32	4,15	4,26	3,80	2,52	1,37
5,41	3,10	4,86	3,12	2,53	1,71
2,35	5,14	4,42	4,81	2,16	2,86
3,29	2,90	5,29	4,20	2,04	1,31
3,52	3,41	4,82	4,40	3,28	2,42

Κατά τη μετάβαση από τα 1000 στα 3000 ft με ταχύτητα 75 Knots : 4,13 /// 3,23///2,44///3,17

Πίνακας Β-3. Ταχύτητα ροής δεδομένων (Download) στην περιοχή Κιλελέρ – Αρμενίου

500ft/50Kn	500ft/100Kn	1000ft/50Kn	1000ft/100Kn	3000ft/50Kn	3000ft/100Kn
2,29	3,17	2,71	1,58	0,67	2,63
2,29	2,20	3,20	2,60	0,81	2,23
2,80	2,29	2,44	1,31	0,65	1,59
3,99	1,05	2,64	0,89	0,63	1,61
2,64	2,21	1,89	0,86	0,78	2,97
2,32	1,65	2,45	2,67	0,59	0,68
2,46	1,79	2,36	1,52	0,66	0,79
1,98	2,05	2,86	1,16	0,87	0,88
2,54	2,10	1,96	0,95	0,86	0,93
2,88	1,29	2,41	0,92	0,72	0,80

Κατά τη μετάβαση από τα 1000 στα 3000 ft με ταχύτητα 75 Knots : 1.06 /// 1.03///0.88///1.77

Πίνακας Β-4. Ταχύτητα ροής δεδομένων (Upload) στην περιοχή Κιλελέρ - Αρμενίου

(3) Περιοχή Βόλου

Πραγματοποιήθηκε ελειψοειδές ίχνος πτήσης (orbit) το κέντρο του οποίου είχε συντεταγμένες: 39°21'43''N, 22°55'32''E

500ft/50Kn	500ft/100Kn	1000ft/50Kn	1000ft/100Kn	3000ft/50Kn	3000ft/100Kn
2,07	1,94	2,01	1,58	2,30	0,68
2,82	1,17	1,10	0,64	1,05	2,48
2,97	1,59	1,33	2,66	1,24	1,77
4,05	1,63	0,84	0,38	1,39	1,52
3,14	1,16	1,25	0,93	0,80	0,88
2,87	1,24	1,37	0,81	0,91	1,21
2,56	1,60	1,55	1,19	1,45	0,76

Κατά την πτήση επίστροφής από Βόλο προς Α/Δ Στεφανοβικείου (2000 to 500ft/75Kn):
1,33///2,23///6,19///6,09///2,29///6,83///6,85///1,71///3,74///6,22

Πίνακας Β-5. Ταχύτητα ροής δεδομένων (Download) στην περιοχή Βόλου

500ft/50Kn	500ft/100Kn	1000ft/50Kn	1000ft/100Kn	3000ft/50Kn	3000ft/100Kn
1,71	1,14	1,32	0,45	0,64	0,88
0,72	1,04	1,45	0,58	1,51	0,44
1,46	0,88	0,82	0,58	1,03	0,87
2,18	1,23	0,72	0,65	0,46	0,95
1,25	0,97	0,56	0,48	0,58	0,52
0,96	1,01	1,12	0,56	0,97	0,64
1,03	0,93	0,86	0,52	1,01	0,81

Κατά την πτήση επίστροφής από Βόλο προς Α/ΔΣτεφανοβικείου (2000 to 500ft/75Kn):
1.25///0.79///1.22///2.88///1.60///6.65///1.28///2.91///3.30///0.77

Πίνακας Β-6. Ταχύτητα ροής δεδομένων (Upload) στην περιοχή Βόλου