



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

**«Ανίχνευση Ανθρώπων σε Εικόνες με Χρήση της
Τεχνικής των Ιστογραμμάτων Κατά Κατεύθυνση
Παραγώγου»**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΤΟΥ

Μακρυμάλλη Αντωνίου

Επιβλέπων :Στέφανος Κόλλιας
Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούλιος 2007



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

«Ανίχνευση Ανθρώπων σε Εικόνες με Χρήση της Τεχνικής των Ιστογραμμάτων Κατά Κατεύθυνση Παραγώγου»

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΤΟΥ

Μακρυμάλλη Αντωνίου

Επιβλέπων : Στέφανος Κόλλιας
Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την Χ^η 2007.

.....
Στ. Κόλλιας
Καθηγητής Ε.Μ.Π

.....
Α. Σταφυλοπάτης
Καθηγητής Ε.Μ.Π

.....
Π. Τσανάκας
Καθηγητής Ε.Μ.Π

Αθήνα, Ιούλιος 2007

.....

Αντώνιος Σ. Μακρυμάλλης

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

Copyright © Αντώνιος Μακρυμάλλης, 2007

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

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

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

ΠΕΡΙΛΗΨΗ

Η ανίχνευση αντικειμένων σε εικόνες ή ακολουθίες εικόνων, αποτελεί ένα σημαντικό πεδίο έρευνας τα τελευταία χρόνια. Σε αυτή την αναφορά παρουσιάζεται και μελετάται μια μέθοδος για την ανίχνευση ανθρώπων σε σταθερές εικόνες, με απροσδιόριστες σκηνικές συνθήκες, όπως είναι η ύπαρξη πολύπλοκου υπόβαθρου και μη ελεγχόμενου φωτισμού. Για την περιγραφή της ανθρώπινης παρουσίας χρησιμοποιείται διάνυσμα χαρακτηριστικών, τα οποία εξάγονται υιοθετώντας την τεχνική των ιστογραμμάτων κατά κατεύθυνση παραγώγου (HOGs) σε ένα πυκνό πλέγμα πάνω στην εικόνα. Η ταξινόμηση της εικόνας επιτυγχάνεται μέσω μιας γραμμικής μηχανής SVM. Η μελέτη και αξιολόγηση της μεθόδου πραγματοποιείται μέσω δύο υλοποιήσεων της, που διαφοροποιούνται στην επί μέρους επιλογή των χαρακτηριστικών των HOGs. Επιπλέον, αξιολογείται και παρατηρείται η ευαισθησία κάθε υλοποίησης σε βασικές παραμέτρους της μεθόδου, μεταβάλλοντας κατάλληλα τις τιμές τους. Από την εφαρμογή της μεθόδου γενικότερα και ειδικότερα από τις τελικές αξιολογήσεις και παρατηρήσεις, συμπεραίνεται ότι η χρήση πληροφορίας από τα HOGs αιχμαλωτίζει ιδανικά τα χαρακτηριστικά της ανθρώπινης μορφής και μας δίνει ιδιαίτερος ικανοποιητικά και αξιόπιστα αποτελέσματα για την ανίχνευση ανθρώπων σε σταθερές εικόνες.

ABSTRACT

The object detection in images or sequences of images constitutes an important field of research in the past few years. This report presents and studies a method for the detection of people in still images, with undefined scenic conditions, such as the existence of complicated background and unpredictable lighting. For the description of human presence, a feature vector is used, which is exported adopting the technique of Histograms of Oriented Gradients (HOGs) in a dense grid on the image. The classification of image is achieved via a linear SVM. The study and evaluation of the method are being achieved through two implementations of itself, which are differentiated in the part choice of the features of HOGs. Moreover, the sensitivity of each implementation in basic parameters of the method is evaluated and observed, altering suitably the values of these parameters. From the application of the method generally and specifically from the final evaluations and observations, is concluded that the use of information from HOGs captures ideally the characteristics of human form and give us significantly satisfactory and reliable results for the detection of people in still images.

ΠΡΟΛΟΓΟΣ

Η παρούσα διπλωματική εργασία εκπονήθηκε το ακαδημαϊκό έτος 2006-2007 στο Εθνικό Μετσόβιο Πολυτεχνείο και συγκεκριμένα στο εργαστήριο Ψηφιακής Επεξεργασίας Εικόνας, Βίντεο και Πολυμέσων, υπό την επίβλεψη του καθηγητή κ. Στέφανου Κόλλια. Θα ήθελα να ευχαριστήσω τους Ιωάννη Αβρίθη και Γεώργιο Ακρίβα για την γόνιμη συνεργασία που είχα μαζί τους καθ' όλη την διάρκεια εκπόνησης της εργασίας αυτής και για την ουσιαστική βοήθειά τους στο να ξεπεράσω πολλές από τις δυσκολίες που προέκυψαν, καθώς και τον επιβλέποντα καθηγητή κύριο Στέφανο Κόλλια για την εμπιστοσύνη που μου έδειξε και την ευκαιρία που μου έδωσε. Επίσης, θα ήθελα να ευχαριστήσω όλα τα μέλη της ερευνητικής ομάδας του εργαστηρίου Εικόνας, Βίντεο και Πολυμέσων και ιδιαιτέρως τον Πέτρο Καψάλα για την πολύτιμη βοήθεια που μου προσέφεραν, όποτε τη χρειαζόμουν. Τέλος, θα ήθελα να ευχαριστήσω τους γονείς μου καθώς και όλους όσους με υποστήριξαν τον τελευταίο χρόνο.

Αθήνα, Ιούλιος 2007
Μακρυμάλλης Αντώνιος

ΠΕΡΙΕΧΟΜΕΝΑ

1	Εισαγωγή	10
1.1	Στόχος	11
1.2	Δυσκολίες	13
1.3	Διάρθρωση και Πορεία της Αναφοράς	15
2	Βασικές Έννοιες-Απαραίτητες Γνώσεις	16
2.1	Όραση Υπολογιστών	16
2.2	Αναγνώριση Προτύπων	18
2.3	Χαρακτηριστικά Εικόνων	21
2.4	Ταξινομητές	22
3	Μέθοδοι και Τεχνικές Για Ανίχνευση Αντικειμένων	24
3.1	Ανίχνευση Αντικειμένων στη Βιβλιογραφία	24
3.2	Υπάρχουσες Μέθοδοι και Τεχνικές για την Ανίχνευση Ανθρώπων	25
3.3	Η Μέθοδος που Επιλέξαμε	26
4	Παρουσίαση της Μεθόδου	27
4.1	Τα Βήματα Της Μεθόδου	28
5	Υλοποίηση και Αξιολόγηση της Μεθόδου	33
5.1	Υλοποίηση με Μία Επικρατούσα Κατεύθυνση	34
5.1.1	Αλγόριθμος αναφοράς	34
5.1.2	Κανονικοποίηση της εικόνας πριν την εφαρμογή του αλγόριθμου	36
5.1.2.1	Γάμμα κανονικοποίηση	36
5.1.2.2	Εξισορρόπηση Ιστογράμματος	37
5.1.3	Μεταβολή στο μέγεθος του κελιού	38
5.1.4	Μεταβολή πλήθους των Bins	39
5.1.5	Κανονικοποίηση σε τοπικό επίπεδο μπλοκ	41
5.1.5.1	Κανονικοποίηση με χρήση της νόρμας	42
5.1.5.2	Κανονικοποίηση με χρήση φίλτρου Gauss	42

5.1.6	Μεταβολή στο πλήθος των εικόνων προς εκπαίδευση	44
5.2	Υλοποίηση Με Περισσότερες Κατευθύνσεις	46
5.2.1	Αλγόριθμος αναφοράς	47
5.2.2	Κανονικοποίηση της εικόνας πριν την εφαρμογή του αλγόριθμου	48
5.2.2.1	Γάμμα κανονικοποίηση με χρήση της τετραγωνικής ρίζας	48
5.2.2.2	Γάμμα κανονικοποίηση με χρήση λογάριθμου	48
5.2.3	Μεταβολή στο μέγεθος του κελιού	49
5.2.4	Μεταβολή πλήθους Bins	51
5.2.5	Κανονικοποίηση σε τοπικό επίπεδο μπλοκ	52
5.3	Συμπεράσματα υλοποιήσεων	55
6	Σύνοψη και προτάσεις επέκτασης	57
6.1	Επέκταση και βελτίωση της μεθόδου	57
6.2	Σύνοψη –Γενικά συμπεράσματα	60
7	Βιβλιογραφία	61

ΚΕΦΑΛΑΙΟ 1

1 Εισαγωγή

Οι υπολογιστές είναι πλέον μέρος της καθημερινότητας μας. Εκτελούν ποικίλες, υπολογιστικές και μη εργασίες, περισσότερο αποτελεσματικά, γρήγορα και με μεγαλύτερη ακρίβεια από ό,τι ο άνθρωπος. Είναι λοιπόν φυσικό να επιζητείται η διεύρυνση και πλήρης εκμετάλλευση των δυνατοτήτων τους, ούτως ώστε να είναι σε θέση να φέρνουν σε πέρας διεργασίες που απαιτούν “νοημοσύνη”. Τέτοιες διεργασίες, όπως η ικανότητα να βγάζουν λογικά συμπεράσματα ή να κάνουν λογικούς συλλογισμούς, να αναλύουν εικόνες, οπτικές σκηνές ή ομιλίες και τόσα άλλα, εκτελούνται από έναν άνθρωπο υποσυνείδητα, εκατοντάδες φορές κάθε μέρα με τέτοια άνεση και ευκολία που τις περισσότερες φορές δεν γίνονται καν αντιληπτές από τον ίδιο τον άνθρωπο. Για να φτάσουμε λοιπόν, σε έναν υπολογιστή με “νοημοσύνη”, θα πρέπει αφ’ ενός να του μεταβιβάσουμε όλα αυτά τα οπτικά, ακουστικά ή οποιασδήποτε φύσης ερεθίσματα που χρησιμοποιεί ένας άνθρωπος ως πληροφορίες για την περάτωση μιας νοήμονος διεργασίας και αφ’ ετέρου, με κατάλληλες υπολογιστικές διεργασίες να διαχειριστούμε και να εκμεταλλευτούμε αυτές τις πληροφορίες με τέτοιο τρόπο, ώστε να φτάσουμε σε ένα αποτέλεσμα αντίστοιχο με αυτό που θα έφτανε ένας άνθρωπος. Όσον αφορά το πρώτο μέρος, τη λύση έδωσε η ψηφιοποίηση της πληροφορίας. Με την ψηφιοποίηση της πληροφορίας έχουμε πλέον τη δυνατότητα ενιαίου τρόπου αναπαράστασης και διαχείρισης σχεδόν κάθε μορφής πληροφορίας. Έτσι, μπορούμε να τροφοδοτούμε τον υπολογιστή με πληροφορίες σε μορφή κειμένων, σχεδίων, στατικής ή κινούμενης εικόνας και φωνής. Μπορούμε επίσης να τον τροφοδοτούμε με πληροφορίες οι οποίες έχουν τη μορφή ηλεκτρικών σημάτων που αντιστοιχούν σε φυσικά μεγέθη όπως η θερμοκρασία, η πίεση, η υγρασία, το ύψος, η ταχύτητα κίνησης και πολλά άλλα. Έτσι, με τον τρόπο της, η ψηφιοποίηση της πληροφορίας άνοιξε το δρόμο στην επιστημονική κοινότητα και το μόνο που έμενε ήταν η προσπάθεια για

αξιοποίηση αυτής της ψηφιακής πληροφορίας, με τη βοήθεια της πληροφορικής για την κατασκευή «έξυπνων» μηχανών που δεν θα υστερούσαν σε τίποτα έναντι της ανθρώπινης νοημοσύνης. Στην προσπάθεια αυτή ξεπήδησε ο κλάδος της Τεχνητής Νοημοσύνης. Τεχνητή Νοημοσύνη είναι ο κλάδος που ασχολείται με τον προγραμματισμό υπολογιστών, ώστε να κάνουν νοήμονα, ανθρώπινα πράγματα, αλλά όχι απαραίτητα με ανθρώπινο τρόπο. Έτσι, ένας από τους βασικούς στόχους της “Τεχνητής Νοημοσύνης ” και πιο συγκεκριμένα ενός κλάδου της, της “Όρασης Υπολογιστών,” είναι να προικίσουν τους υπολογιστές με τη δυνατότητα να “βλέπουν”, να κάνουν δηλαδή, οπτική ανάλυση και ερμηνεία εικόνων και βίντεο. Πρωταρχικό και βασικό στάδιο για την πραγματοποίηση αυτού του στόχου είναι η ανίχνευση και ο διαχωρισμός των διάφορων αντικειμένων σε μία εικόνα. Μια τέτοια δυνατότητα θα είχε πολλές εφαρμογές, όπως για παράδειγμα στην επικοινωνία ανθρώπου –μηχανής, στη ρομποτική, στην αυτόματη ανάλυση προσωπικών ή εμπορικών ψηφιακών δεδομένων, σε έξυπνα αυτόνομα οχήματα .Στην παρούσα εργασία γίνεται μια προσπάθεια μελέτης, ανάλυσης και αξιολόγησης μιας μεθόδου ανίχνευσης ανθρώπων σε σταθερές εικόνες. Στη συνέχεια του πρώτου αυτού κεφαλαίου θα γίνει μία εισαγωγή στο στόχο της παρούσας εργασίας, τα προβλήματα που θα κληθούμε να αντιμετωπίσουμε και τέλος θα παρουσιαστεί η διάρθρωσή της.

1.1 Στόχος

Η παρούσα αναφορά έχει σαν στόχο την ανίχνευση ανθρώπων σε σταθερές εικόνες. Η ανίχνευση, στοχεύουμε να γίνεται μέσω ενός παραθύρου συγκεκριμένης διάστασης το οποίο ονομάζουμε παράθυρο ανίχνευσης. Μέσω του παραθύρου αυτού θα μπορούμε να αποφανθούμε αν υπάρχει άνθρωπος, σε οποιαδήποτε περιοχή εικόνας, ίδιου μεγέθους με το παράθυρο ανίχνευσής μας. Ως ανθρώπινη παρουσία μέσα στο παράθυρο δε νοείται μέρος του ανθρώπου, ούτε ανθρώπινο σώμα μεγέθους κατά πολύ μικρότερο του παραθύρου. Το ανθρώπινο σώμα θα πρέπει να γεμίζει το παράθυρο ανίχνευσης ή σε διαφορετική περίπτωση, να αφήνει κάποιο μικρό περιθώριο μέχρι τα όρια του παραθύρου ή να εξέχει ελάχιστα από αυτό. Στο Σχήμα 1

απεικονίζεται η έννοια του παράθυρου ανίχνευσης και της ανθρώπινης παρουσίας μέσα σε αυτό. Έτσι λοιπόν, το βασικό μας ενδιαφέρον προς το στόχο αυτό επικεντρώνεται στην εύρεση κατάλληλων χαρακτηριστικών που "περιγράφουν" με επαρκή και ισχυρό τρόπο το αντικείμενο "άνθρωπος", στην κατασκευή ενός αλγόριθμου εξόρυξης των χαρακτηριστικών αυτών και στην εκμετάλλευση του αλγόριθμου αυτού για την απόφαση ύπαρξης ή όχι ανθρώπου. Τα στάδια αυτά περιγράφουν επιγραμματικά, αλλά επαρκώς το στόχο της παρούσας αναφοράς και τα βήματα για την επίτευξή του. Παράλληλα με το βασικό μας στόχο και προσβλέποντας στην αποτελεσματικότερη επίτευξή του, θα πρέπει να ξεπεράσουμε αρκετές δυσκολίες και προκλήσεις που θα μας αντιταχθούν. Επιπλέον στόχος μας λοιπόν, είναι η αντιμετώπιση με τον καλύτερο δυνατό τρόπο οποιουδήποτε προβλήματος σταθεί εμπόδιο στην αποτελεσματικότητα του ανιχνευτή μας. Στην επόμενη παράγραφο παρουσιάζονται τα βασικότερα προβλήματα που πρέπει να ξεπεράσουμε με τον καλύτερο δυνατό τρόπο, ώστε να ολοκληρώσουμε το στόχο μας αποτελεσματικότερα.



Σχήμα 1. Ανίχνευση ανθρώπου σε εικόνα μέσω παραθύρου ανίχνευσης. Με κόκκινο χρώμα απεικονίζεται το παράθυρο ανίχνευσης.

1.2 Δυσκολίες

Η μεγαλύτερη δυσκολία στην προσπάθειά μας να υλοποιήσουμε έναν ανιχνευτή ανθρώπου είναι το τεράστιο πλήθος των παραλλαγών που συναντάμε σε ανθρώπους, σε εικόνες και ειδικότερα σε ανθρώπους μέσα σε εικόνες. Για να γίνουμε πιο συγκεκριμένοι οι βασικές δυσκολίες που πρέπει να ξεπεράσουμε έχουν ως εξής :

-Πρώτα από όλα, η δομή μίας εικόνας συμπιέζει την τρισδιάστατη πληροφορία και προκαλεί μεγάλη εξάρτηση από την οπτική γωνία της λήψης, τόσο μεγάλη, που έστω και μία μικρή μετακίνηση του υπό φωτογράφιση αντικειμένου ή της φωτογραφικής μηχανής θα μπορούσε να προκαλέσει σημαντικές αλλαγές στην εμφάνισή του. Ένα παραπλήσιο θέμα είναι η μεγάλη ποικιλία του μεγέθους του ανθρώπινου σώματος με βάση την κλίμακα κάτω από την οποία έχει φωτογραφηθεί. Ένας αξιόπιστος ανιχνευτής αντικειμένων θα πρέπει να διαχειρίζεται τέτοια θέματα αναλόγως και να μην επηρεάζεται.

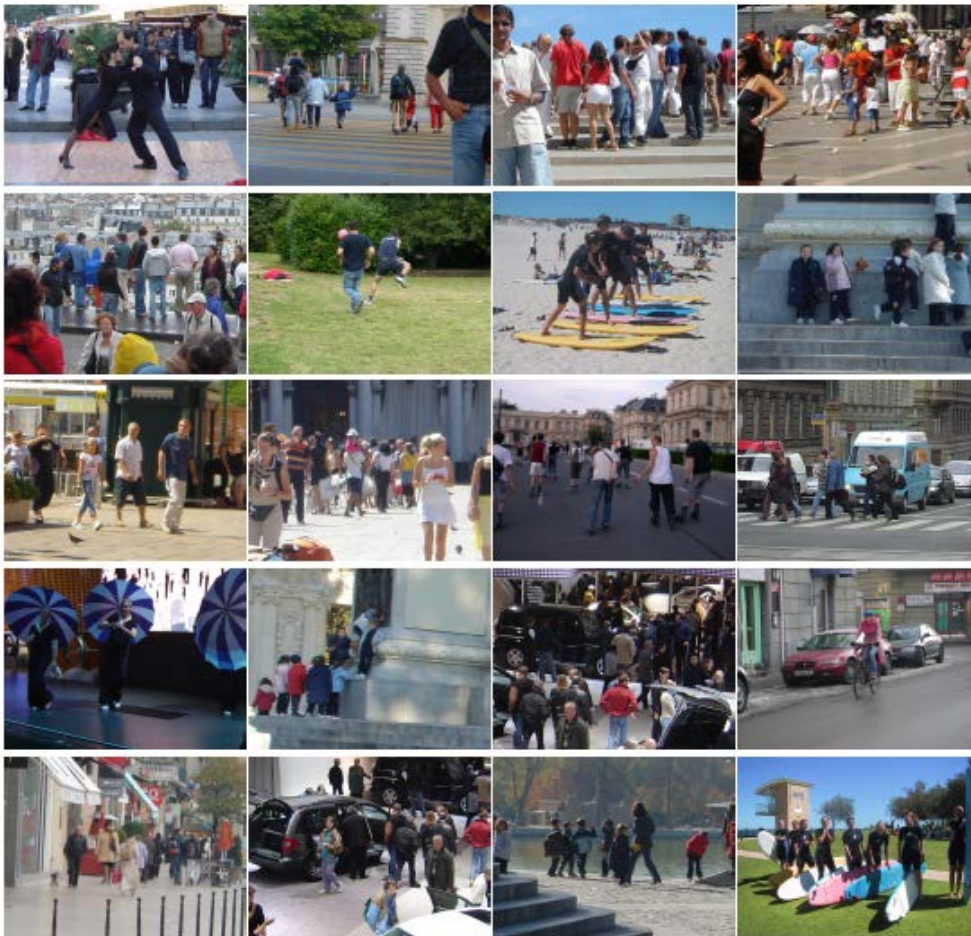
-Δεύτερον, όπως τα περισσότερα αντικείμενα έτσι και το αντικείμενο "άνθρωπος" εμφανίζει αρκετά μεγάλη ποικιλία. Γίνεται εύκολα αντιληπτό ότι οι αλλαγές στην εμφάνιση, στην πόζα ή ακόμα και στα ρούχα ενός ανθρώπου, το πλήθος των οποίων είναι πολύ μεγάλο από εικόνα σε εικόνα, όπως άλλωστε από άνθρωπο σε άνθρωπο, προκαλούν άπειρες διαφοροποιήσεις στο αντικείμενό μας. Ένας ισχυρός ανιχνευτής λοιπόν, θα πρέπει να λειτουργεί ανεξάρτητος από τέτοιες διαφοροποιήσεις.

-Τρίτον, η ακατάστατη ποικιλομορφία του φόντου είναι ένα συχνό και φυσιολογικό φαινόμενο που διαφέρει από εικόνα σε εικόνα, ανάλογα σε τι σκηνικό έχει τραβηχτεί. Ο ανιχνευτής μας θα πρέπει να διαθέτει τέτοια ικανότητα, ώστε να διαχωρίζει τις περιοχές μιας εικόνας στις οποίες ανήκει κάποιο αντικείμενο, από το περίπλοκο φόντο που τις περιβάλλει.

-Τέταρτον, το χρώμα του δέρματος και των ρούχων ή ο φωτισμός ποικίλουν από άνθρωπο σε άνθρωπο και από εικόνα σε εικόνα. Για να είναι αποτελεσματικός ο ανιχνευτής μας θα πρέπει να αντεπεξέρχεται σε χρωματικές αλλαγές και να παραμένει ανεξάρτητος σε ένα μεγάλο εύρος φωτισμού.

-Τέλος, πρόβλημα μπορούν να προκαλέσουν και εμπόδια που καλύπτουν μέρος του ανθρώπινου σώματος.

Στο Σχήμα 2, μπορούμε να δούμε διάφορα παραδείγματα εικόνων που παρουσιάζουν τα βασικά προβλήματα που προαναφέραμε. Είναι χαρακτηριστικό το μέγεθος των διακυμάνσεων από εικόνα σε εικόνα και πιο συγκεκριμένα οι αποκλίσεις στη στάση και την εμφάνιση κάθε ανθρώπου, στην φωτισμό και τη φωτεινότητα της εικόνας, την ενδυματολογική ποικιλία και τη διαφορετικότητα των σκηνών που έχουν τραβηχτεί οι εικόνες αυτές.



Σχήμα 2. Μερικές εικόνες από προσωπική συλλογή ψηφιακών φωτογραφιών .

1.3 Διάρθρωση και Πορεία της Αναφοράς

Σε αυτό το σημείο θα παρουσιάσουμε επιγραμματικά την πορεία της δουλειάς μας σε αυτήν την εργασία. Υιοθετήσαμε την τεχνική των ιστογραμμάτων κατά κατεύθυνση παραγώγου όπως αυτή παρουσιάζεται από τους Navneet Dalal & Bill Triggs [11]. Δημιουργήσαμε έναν αλγόριθμο εξόρυξης χαρακτηριστικών βασιζόμενοι στα βήματα που περιγράφουν και ελέγξαμε την αποτελεσματικότητα του χρησιμοποιώντας μία μηχανή SVM για ταξινομητή και σύνολο εικόνων από τη βάση δεδομένων της INRIA [<http://pascal.inrialpes.fr/data/human/>] για εκπαίδευση και δοκιμή. Στη συνέχεια αξιολογήσαμε την απόδοση του αλγόριθμου με την προσθήκη ή αλλαγή παραμέτρων σε κάποια στάδια του. Στο κεφάλαιο 2 γίνεται μία παρουσίαση και περιγραφή εννοιών και όρων που μας εισαγάγουν καλύτερα στο στόχο μας και ίσως φανούν χρήσιμα στη συνέχεια της εργασίας. Στο κεφάλαιο 3 κάνουμε μία βιβλιογραφική έρευνα πάνω σε θέματα, μεθόδους και τεχνικές που εμφανίζουν κοινά στοιχεία ή στόχους με το δικό μας, ενώ στο κεφάλαιο 4 παρουσιάζουμε θεωρητικά τη μέθοδο την οποία αποφασίσαμε να ακολουθήσουμε. Στο κεφάλαιο 5 υλοποιούμε τη μέθοδο μας και αξιολογούμε την απόδοση της σε συνάρτηση με συγκεκριμένες παραμέτρους. Τέλος, στο κεφάλαιο 6 καταγράφονται τα συμπεράσματα μας, καθώς και προτάσεις για επέκταση ή βελτίωση της μεθόδου μας.

ΚΕΦΑΛΑΙΟ 2

2 Βασικές Έννοιες-Απαραίτητες Γνώσεις

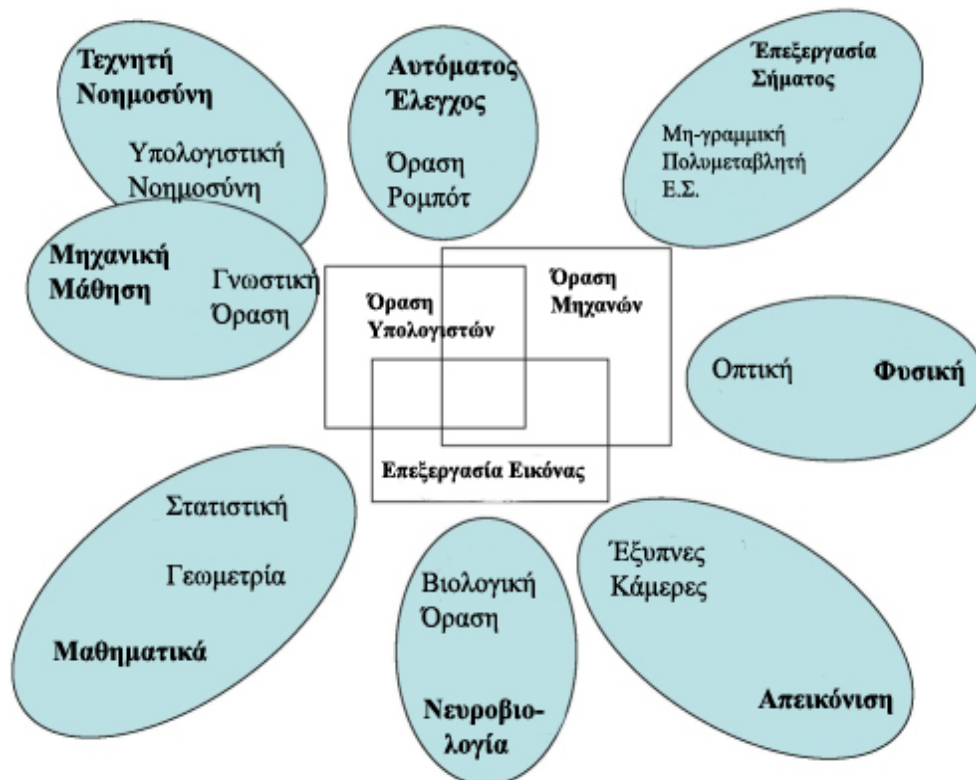
Πριν την λεπτομερή ανάλυση του θέματος της παρούσας αναφοράς, θεωρούμε σκόπιμη την επεξήγηση κάποιων βασικών εννοιών και όρων τους οποίους είτε θα συναντήσουμε είτε θα μας εισαγάγουν με καλύτερο τρόπο στο θέμα μας.

2.1 Όραση Υπολογιστών

Όραση είναι η αισθητηριακή αντίληψη των αντικειμένων του κόσμου και των ιδιοτήτων τους με βάση τη φωτεινή ενέργεια που εκπέμπεται από αυτά και συλλέγεται από οπτικούς αισθητήρες. Η σημασία της όρασης για τον άνθρωπο και για όσα έμβια όντα διαθέτουν αυτή την αίσθηση επαυξάνεται από το γεγονός ότι το μεγαλύτερο ποσό της πληροφορίας που εισέρχεται στο ανθρώπινο σύστημα αντίληψης μέσω των αισθήσεων είναι οπτικό. Η είσοδος στο ανθρώπινο (ή κάθε άλλο βιολογικό) σύστημα όρασης είναι οι εικόνες. Για να δούμε όμως πρέπει να συνδέσουμε τα χωριστά συμπεράσματα για την κίνηση, το χρώμα, το σχήμα, την υφή και το βάθος σε μια ενοποιημένη εξήγηση. Αυτή η ενοποιημένη πληροφορία χρησιμοποιείται για την εκτέλεση διάφορων σύνθετων εργασιών, όπως για παράδειγμα για την αναγνώριση και παρακολούθηση ενός κινούμενου αντικειμένου σε μια ακολουθία εικόνων.

Σε αντίθεση όμως με τα βιολογικά συστήματα, όπου αυτή η δυνατότητα της ενοποίησης και επεξεργασίας των οπτικών πληροφοριών είναι αυτονόητη, αυτόματη και έμφυτη, στα μηχανικά συστήματα είναι απαραίτητο να προηγηθεί επεξεργασία μιας πλειάδας χαρακτηριστικών και πληροφοριών, που προέρχονται από την απεικόνιση του πραγματικού κόσμου σε μορφή ακολουθίας εικόνων, ώστε να καταλήξουμε σε δυνατότητες αντίστοιχες αυτών των βιολογικών συστημάτων. Η επεξεργασία αυτή γίνεται με μεθόδους που υποδεικνύει η Όραση Υπολογιστών, της οποίας βασικός στόχος είναι,

δεδομένης μιας ή περισσότερων δισδιάστατων (2Δ) εικόνων, η εύρεση μιας όσο το δυνατόν πληρέστερης συμβολικής περιγραφής των αντικειμένων του τρισδιάστατου (3Δ) κόσμου που περιέχονται στη σκηνή αυτή.



Σχήμα 3. Η Όραση Υπολογιστών στον Επιστημονικό Χώρο.

Η Όραση Υπολογιστών άρχισε ως επιστημονικό πεδίο τη δεκαετία του 1960 από προσπάθειες ερευνητών της Τεχνητής Νοημοσύνης να δημιουργήσουν υπολογιστές που βλέπουν. Τα επιστημονικά πεδία που συνέβαλαν στη δημιουργία και ανάπτυξη της Όρασης Υπολογιστών είναι η Τεχνητή Νοημοσύνη, η Επεξεργασία Σημάτων και η Αναγνώριση Προτύπων. Στις μέρες μας, όμως η Όραση υπολογιστών είναι άρρηκτα συνδεδεμένη με πάρα πολλούς χώρους της πληροφορικής και όχι μόνο. Η σχέση του τομέα αυτού με διάφορα επιστημονικά πεδία παρουσιάζεται επαρκώς στην εικόνα που συναντήσαμε στο wiki [http://en.wikipedia.org/wiki/Computer_vision] και παρουσιάζουμε στο

Σχήμα 3. Τα συστήματα όρασης υπολογιστή βασίζονται συνήθως σε μια προσέγγιση η οποία ξεκινάει από την επεξεργασία της εικόνας σε χαμηλό επίπεδο και σταδιακά την επεξεργάζεται σε ολοένα και πιο υψηλό επίπεδο. Η εικόνα αντιμετωπίζεται σαν ένα σύνολο από pixels, τα οποία αναλύονται σταδιακά για να εξαχθεί η σημασία τους. Τυπικά παραδείγματα λειτουργιών επεξεργασίας χαμηλού επιπέδου είναι η αφαίρεση του θορύβου από την εικόνα και η προσαρμογή των επιπέδων φωτεινότητας. Παράδειγμα επεξεργασίας του αμέσως υψηλότερου επιπέδου είναι η ανίχνευση και λέπτυνση των περιγραμμάτων των αντικειμένων που τυχόν περιέχει η εικόνα. Κατόπιν, σε ακόμα υψηλότερο επίπεδο επεξεργασίας, οι παραγόμενες γραμμές θα μπορούσαν να ενώνονται για την δημιουργία ολόκληρων σχημάτων και σ' ένα επόμενο στάδιο θα μπορούσαν να χρησιμοποιούνται υψηλότερου επιπέδου πληροφορίες, τριών διαστάσεων, για την εξαγωγή περιγραφών των σχημάτων από τα περιγράμματα που υπάρχουν στην εικόνα. Η υφή και το χρώμα, μπορούν επίσης να χρησιμοποιηθούν σαν βοηθητικά στοιχεία (όπως π.χ. σε μια εικόνα στην οποία θέλουμε να απομονώσουμε το δέρμα των χεριών και του προσώπου ενός ανθρώπου) μέχρι να φτάσουμε στο υψηλότερο δυνατό επίπεδο περιγραφής της εικόνας, βάσει των αντικειμένων που περιλαμβάνει και των σχετικών θέσεων τους. Ωστόσο, τα προβλήματα που πρέπει να επιλυθούν κατά την σχεδίαση ενός συστήματος όρασης είναι πολύπλοκα και για το λόγο αυτό η πρόοδος σε αυτό τον τομέα γίνεται με σχετικά μικρούς ρυθμούς, αν και αποτελεί αντικείμενο με έντονο ερευνητικό ενδιαφέρον.

2.2 Αναγνώριση Προτύπων

Οι άνθρωποι καθημερινά στη ζωή μας σκεφτόμαστε, αντιλαμβανόμαστε, αποφασίζουμε, ενεργούμε κάνοντας αναρίθμητες κατηγοριοποιήσεις. Οι κατηγοριοποιήσεις που κάνουμε, γίνονται αυτόματα και ασυνείδητα και περιλαμβάνουν τόσο αντικείμενα του πραγματικού κόσμου (άνθρωποι, ζώα, φυσικά και τεχνητά αντικείμενα) όσο και αφηρημένες έννοιες (γεγονότα, ενέργειες, συναισθήματα, χωρικές και κοινωνικές σχέσεις). Μπορούμε, δηλαδή

εύκολα να ξεχωρίσουμε και να αναγνωρίσουμε μέσω της γεύσης αν το φρούτο που τρώμε είναι ροδάκινο, καρπούζι, κεράσι ή κάτι άλλο, μέσω της ακοής ότι η φωνή που ακούμε είναι της μητέρας, του φίλου, του αδερφού μας ή κάποιου άλλου και το ίδιο εύκολα αν κάποιος ζώο που βλέπουμε είναι σκύλος, γάτα ή οτιδήποτε άλλο. Άπειρα είναι τα ανάλογα παραδείγματα στη ζωή μας, όπου καλούμαστε να κατηγοριοποιήσουμε μία αφηρημένη ή συγκεκριμένη οντότητα του κόσμου που μας περιβάλλει. Αυτό που κάνουμε σε κάθε τέτοια περίπτωση είναι να αναγνωρίζουμε τα ιδιαίτερα χαρακτηριστικά της οντότητας ή του αντικείμενου που το κάνουν να εντάσσεται σε μια κατηγορία και όχι σε άλλη. Τις περισσότερες μάλιστα φορές, αυτό γίνεται ασυνείδητα. Δεν καθόμαστε να σκεφτούμε ότι ο σκύλος έχει αυτό και το άλλο χαρακτηριστικό που δεν έχει η γάτα, οπότε αυτό που βλέπουμε είναι σκύλος. Απλά το ξέρουμε σχεδόν στιγμιαία από τότε που θα το δούμε, με βάση το εννοιολογικό μας οπλοστάσιο. Ιδιαίτερα εντυπωσιακή είναι η ικανότητα του ανθρώπου να αναγνωρίζει οντότητες ή αντικείμενα χωρίς κόπο, ακόμα και αν κάποια χαρακτηριστικά αυτών έχουν αλλάξει έως ένα σημείο. Για παράδειγμα, αν ένας φίλος μας κουρευτεί και φορέσει γυαλιά, το πιθανότερο είναι ότι και πάλι θα τον αναγνωρίσουμε. Τα παραδείγματα από την ανθρώπινη εμπειρία που αναφέραμε εντάσσονται στον γενικό κλάδο της Αναγνώρισης Προτύπων. Αν θέλαμε να ορίσουμε την Αναγνώριση Προτύπων, τότε θα λέγαμε ότι είναι η διαδικασία της κατάταξης ενός προτύπου σε μία κατηγορία. Με τον όρο πρότυπο εννοούμε γενικά μία αφηρημένη ή συγκεκριμένη οντότητα. Η αναγνώριση προτύπων από τους ανθρώπους είναι μια διαδικασία που συντελείται συνέχεια χωρίς κόπο. Η πολύπλοκη και εντυπωσιακή ανάπτυξη του ανθρώπινου εγκεφάλου περιλαμβάνει ασυνείδητες διαδικασίες που καθιστούν την αναγνώριση αυτή ιδιαίτερα εύκολη ακόμα και σε πιο δύσκολες και μη τετριμμένες περιπτώσεις. Σε πιθανή προσπάθειά μας να δημιουργήσουμε μια μηχανή που να προσομοιώνει την ικανότητα και την ευκολία του ανθρώπου να αναγνωρίζει πρότυπα, θα διαπιστώσουμε ότι δεν είναι τόσο απλό όσο για τους ανθρώπους. Ωστόσο κρίνεται χρήσιμο να δούμε ποιες είναι οι βασικές λειτουργίες που περιλαμβάνει η αναγνώριση ενός προτύπου. Αρχικά, οι

«αισθήσεις» της μηχανής συλλαμβάνουν το πρότυπο και το αποθηκεύουν υπό μορφή δεδομένων διάφορων τύπων. Οι «αισθήσεις» αυτές μπορεί να είναι πραγματικές, δηλαδή η μηχανή να μετράει κάποιες ποσότητες με τη βοήθεια ειδικών αισθητήρων, ή έμμεσες, δηλαδή κάποιος άνθρωπος να συλλέγει τα δεδομένα για να τα τροφοδοτήσει στη μηχανή. Τα δεδομένα με τη σειρά τους μπορεί να είναι αριθμοί, αρχεία αριθμών, εικόνες, σήματα και άλλα. Ακολουθεί η προ-επεξεργασία των δεδομένων. Αυτή περιλαμβάνει την ανάλυση και τον μετασχηματισμό τους σε μορφή που ενδείκνυται για την περαιτέρω επεξεργασία και κατάταξη τους. Το πρώτο κομμάτι της προ-επεξεργασίας είναι ο διαχωρισμός (segmentation). Ο διαχωρισμός αφορά την αποσυσχέτιση των προτύπων εισόδου, ώστε να επιτευχθεί ο χειρισμός του κάθε δεδομένου ξεχωριστά. Η διαδικασία αυτή είναι ιδιαίτερα σημαντική. Ο συσχετισμός των προτύπων πρέπει να αποφεύγεται προκειμένου να ελπίζουμε σε επιτυχή αναγνώριση. Παραδείγματα συσχετισμού είναι η εισαγωγή μιας εικόνας στην οποία, το αντικείμενο που μας ενδιαφέρει επικαλύπτεται από ένα άλλο ή ενός σήματος φωνής, όπου δύο φθόγγοι δε διαχωρίζονται επαρκώς. Το δεύτερο κομμάτι της προ-επεξεργασίας είναι η λεγόμενη εξαγωγή χαρακτηριστικών (feature extraction). Κύρια λειτουργία του είναι ο εντοπισμός εκείνων των χαρακτηριστικών που παίζουν πρωταρχικό ρόλο στον διαχωρισμό προτύπων. Ειδικότερα, αναζητά χαρακτηριστικά που είναι παρόμοια για τα πρότυπα που ανήκουν στην ίδια τάξη, αλλά πολύ διαφορετικά για πρότυπα ξεχωριστών τάξεων και επιπλέον είναι ανεξάρτητα του μετασχηματισμού του χώρου εισόδων. Θα δούμε παρακάτω τι εννοούμε, όταν λέμε χαρακτηριστικά μιας εικόνας, τα οποία είναι και τα χαρακτηριστικά που θα χρησιμοποιήσουμε. Το τρίτο και τελευταίο μέρος της προ-επεξεργασίας είναι η επιλογή χαρακτηριστικών (feature selection). Μέσω αυτής ο αριθμός των χαρακτηριστικών που θα ληφθούν υπόψη για την ταξινόμηση του προτύπου μειώνεται. Ειδικότερα, με χρήση μαθηματικών τρόπων κρατούνται μόνο εκείνα τα χαρακτηριστικά που κρίνονται ως πιο σημαντικά. Με αυτόν τον τρόπο μειώνεται η διάσταση του χώρου εισόδων και κατ' επέκταση η διάσταση του προβλήματος. Τελικό κέρδος αποτελεί η απλούστευση του συστήματος. Η

κεντρική φάση της αναγνώρισης προτύπων είναι η ταξινόμηση (classification). Τα στάδια που προηγούνται αυτής επωμίζονται ουσιαστικά τη διαδικασία απλοποίησης των δεδομένων, ώστε να διευκολυνθεί και να γίνει πιο αποδοτική η δουλειά της φάσης αυτής. Θα αναφερθούμε σε αυτή περαιτέρω παρακάτω. Τελευταίο στάδιο είναι η τελική επεξεργασία των αποτελεσμάτων της ταξινόμησης (post-processing). Οι λειτουργίες αυτού του βήματος ποικίλλουν ανάλογα με τη φύση του προβλήματος. Μπορεί να είναι η εξαγωγή μίας ένδειξης βεβαιότητας για την απόφαση ή, με άλλα λόγια, με τι ποσοστό ρίσκου έχει ταξινομηθεί το πρότυπο. Ακόμα μπορεί να είναι ένας συνδυασμός αποφάσεων πολλών ταξινομητών με σκοπό τη μείωση του ρίσκου στην ταξινόμηση, δηλαδή τη μεγιστοποίηση της πιθανότητας σωστής ταξινόμησης.

2.3 Χαρακτηριστικά Εικόνων

Με τη λέξη εικόνα δεν νοείται απλά η απεικόνιση μίας σκηνής. Η εικόνα είναι ένα μέσο, ένας τρόπος, με τον οποίο μπορούμε να αποτυπώσουμε πληροφορίες διαφόρων ειδών. Έτσι με την ψηφιοποίηση της εικόνας η πληροφορία αυτή γίνεται μετρήσιμη και επεξεργάσιμη.

Ως χαρακτηριστικό μπορεί να θεωρηθεί οποιοδήποτε μετρήσιμο μέγεθος που εξάγεται από μία εικόνα. Στην παρούσα αναφορά μας ενδιαφέρουν ιδιαίτερος σύνολα χαρακτηριστικών τα οποία χρησιμοποιούνται στην αναγνώριση αντικειμένων ή ακόμα πιο συγκεκριμένα χρησιμοποιούνται για να περιγράψουν τα αντικείμενα που περιέχονται σε εικόνες. Τέτοια σύνολα χαρακτηριστικών μπορούμε να τα φανταστούμε σαν μία μαθηματική απεικόνιση των αντικειμένων. Ουσιαστικά η περιγραφή των αντικειμένων με χαρακτηριστικά απεικονίζει τα αντικείμενα στο χώρο των χαρακτηριστικών, με αποτέλεσμα, η αναγνώριση τους να ισοδυναμεί με τη μέτρηση της ομοιότητας μεταξύ των χαρακτηριστικών των αντικειμένων. Συνεπώς, τα χαρακτηριστικά που επιλέγονται πρέπει να είναι σε θέση να διαχωρίζουν και να περιγράφουν κατά το δυνατόν, μονοσήμαντα τα αντικείμενα. Έτσι είναι λογικό πως για ανίχνευση αντικειμένων, αντί να χρησιμοποιούμε ακατέργαστες εικόνες των αντικειμένων,

είναι προτιμότερο να επιλέγουμε χαρακτηριστικά της εικόνας ,δηλαδή μία μορφή πιο εξειδικευμένων τοπικών στοιχείων περιγραφής της εικόνας. Τέτοια χαρακτηριστικά μπορούν να κατηγοριοποιηθούν με βάση το είδος της πληροφορίας που προσφέρουν, αν βασίζονται σε σημεία της εικόνας ,σε εντάσεις της εικόνας, στην κλίση, στο χρώμα ,στην υφή ,σε συνδυασμό όλων αυτών, αλλά και σε άλλα στοιχεία.

2.4 Ταξινομητές

Η ευκολία με την οποία αντιλαμβανόμαστε ένα πρόσωπο, διακρίνουμε τους χαρακτήρες, ξεχωρίζουμε τα κλειδιά μας ακόμα και μόνο με την αφή και αποφασίζουμε αν ένα φρούτο είναι ώριμο από τη μυρωδιά του, είναι διαδικασίες αναγνώρισης προτύπων. Ο άνθρωπος επιλύει αυτά τα προβλήματα με την εμπειρία και την εκπαίδευση που έχουν υποστεί τα δισεκατομμύρια κύτταρα του εγκεφάλου του. Στην περιοχή της Ψηφιακής επεξεργασίας εικόνων, το πρόβλημα της αναγνώρισης προτύπων ανάγεται στην ταξινόμηση ενός συνόλου κατάλληλων και "ισχυρών" χαρακτηριστικών σε κλάσεις με τη βοήθεια κατάλληλων Ταξινομητών. Το διάνυσμα χαρακτηριστικών που χρησιμοποιείται σε ένα σύστημα ταξινόμησης καλείται "πρότυπο". Οι ταξινομητές είναι ουσιαστικά αλγόριθμοι που χρησιμοποιούνται για να επιτύχουμε την ταξινόμηση των προτύπων σε κατά το δυνατόν ομοιογενείς, από άποψη χαρακτηριστικών, κλάσεις. Η ταξινόμηση γίνεται συνήθως είτε με επίβλεψη είτε χωρίς. Στην πρώτη περίπτωση οι κλάσεις δημιουργούνται καθορίζοντας την επιθυμητή απόκριση του ταξινομητή για κάθε διάνυσμα χαρακτηριστικών(πρότυπο). Στη δεύτερη περίπτωση, οι κλάσεις δημιουργούνται αυτόματα, χωρίς να παρέχεται στον ταξινομητή καμία πληροφορία για το που ανήκει κάθε διάνυσμα χαρακτηριστικών. Οι ταξινομητές χωρίζονται σε διάφορες κατηγορίες ανάλογα με τον τρόπο λειτουργίας τους όπως στατιστικοί ταξινομητές, νευρωνικοί ταξινομητές κ.α. Η επιλογή του κατάλληλου σχετίζεται άμεσα με το είδος των χαρακτηριστικών και το είδος της εφαρμογής στην οποία

θα χρησιμοποιηθεί. Στην παρούσα αναφορά ο ταξινομητής που χρησιμοποιήθηκε ήταν μία μηχανή SVM.

Οι μηχανές SVM πραγματοποιούν ταξινόμηση με επίβλεψη και έχουν χρησιμοποιηθεί ευρέως την τελευταία δεκαετία για αναγνώριση αντικειμένων. Η λειτουργία και η αποτελεσματικότητα των μηχανών SVM βασίζεται στη δυνατότητά τους να βρίσκουν το διαχωριστικό υπερεπίπεδο που μεγιστοποιεί το διάστημα μεταξύ της κλάσης αντικείμενο και της κλάσης μη-αντικείμενο. Περισσότερες λεπτομέρειες και γνώσεις για τις μηχανές SVM και την λειτουργικότητά τους για αναγνώριση προτύπων μπορούν να αποκτηθούν διαβάζοντας την αναφορά του Christofer Burges [10].

ΚΕΦΑΛΑΙΟ 3

3 Μέθοδοι και Τεχνικές για Ανίχνευση Αντικειμένων

Στο κεφάλαιο αυτό γίνεται μία αναφορά σε προηγούμενες προσεγγίσεις στο θέμα της παρούσας αναφοράς, είτε ως προς το στόχο ,είτε ως προς τη μέθοδο.

3.1 Ανίχνευση Αντικειμένων στη Βιβλιογραφία

Ανατρέχοντας στη βιβλιογραφία, διαπιστώνουμε την ύπαρξη πολλών και διαφορετικών προσεγγίσεων για την ανίχνευση αντικειμένων και πιο συγκεκριμένα ανθρώπων. Οι Pope & Lowe [13] προσπάθησαν να ανιχνεύσουν αντικείμενα μοντελοποιώντας την εμφάνιση τους κάτω από διάφορες οπτικές γωνίες. Οι Gavrilu and Philomin [1] παρουσίασαν μία μέθοδο βασιζόμενη στο σχήμα, για την ανίχνευση σε πραγματικό χρόνο αντικειμένων, όπως ανίχνευση πεζών, για χρήση σε έξυπνα οχήματα. Οι Mohan, Papageorgiou & Poggio [4] επιχείρησαν ανίχνευση ανθρώπων βασιζόμενοι στα κυρίως μέρη που αποτελούν το ανθρώπινο σώμα.

Αν το επιχειρούσαμε, θα μπορούσαμε να χωρίσουμε αυτές τις προσεγγίσεις σε δύο γενικότερες κατηγορίες ανάλογα με τη μέθοδο που χρησιμοποιούν για την ανίχνευση κάποιου αντικειμένου.

-Η πρώτη κατηγορία μπορεί να χαρακτηριστεί ως μέθοδος που βασίζεται στα μέρη που αποτελούν το αντικείμενο. Σύμφωνα με τη μέθοδο αυτή η αναγνώριση περιοχών της εικόνας, ως μέρη που απαρτίζουν το αντικείμενο, προσδιορίζει την ύπαρξή του. Έτσι η παρουσία ενός ζεύγους από ρόδες, ενός τιμονιού, μίας σέλας και ενός άξονα που ενώνει τις ρόδες, σε συνδυασμό με την απουσία μηχανής, μας υποδηλώνει την ύπαρξη ενός ποδηλάτου. Αν τώρα το αντικείμενο μας είναι άνθρωπος, τα μέρη που θα μπορούσαμε να πούμε πως τον απαρτίζουν είναι το κεφάλι, τα πόδια, τα χέρια και ο κορμός.

-Η δεύτερη κατηγορία περιλαμβάνει μία γενικά πιο απλή μεθοδολογία, χωρίς αυτό να σημαίνει ότι είναι λιγότερο αποτελεσματική. Συγκεκριμένα, συλλέγει και κωδικοποιεί πληροφορία που χαρακτηρίζει το αντικείμενό μας. Τέτοιες πληροφορίες σε μία εικόνα μπορούν να σχετίζονται με το σχήμα, την υφή, αλλά και με οποιοδήποτε άλλο χαρακτηριστικό που μπορεί να περιγράψει το αντικείμενό μας.

3.2 Υπάρχουσες Μέθοδοι και Τεχνικές για την Ανίχνευση Ανθρώπων

Συγκεκριμένα, για την ανίχνευση ανθρώπων, έχουν χρησιμοποιηθεί μέθοδοι και τεχνικές που μπορούν να ενταχθούν και στις δύο κατηγορίες που αναφέραμε στην προηγούμενη ενότητα. Όσον αφορά την πρώτη κατηγορία, είναι πολύ συνηθισμένες οι περιπτώσεις χρησιμοποίησης της. Οι S. Ioffe & D.A. Forsyth [2] πρότειναν ανίχνευση ανθρώπων εντοπίζοντας πιθανά ανθρώπινα μέρη μέσω τμηματοποίησης της εικόνας. Με συνδυασμό των “κομματιών” αυτών ελέγχεται η “κατασκευή” ή μη ανθρώπινου σώματος και αντίστοιχα η ύπαρξη του ή όχι. Βασικά προβλήματα αυτής της προσπάθειας ήταν η έλλειψη ιδανικής τμηματοποίησης της εικόνας και η ενδεχόμενη και πολύ πιθανή απουσία βασικών “κομματιών” του ανθρώπινου σώματος λόγω της ποικιλότητας του ρουχισμού ή της ύπαρξης κάποιου εμποδίου. Τα προβλήματα αυτά προσπάθησαν να αντιμετωπίσουν οι Sprague & Luo [5] εισάγοντας τα mixtures of trees και οι Mikolajczyk et al [6] χρησιμοποιώντας ένα συνδυασμό ιστογραμμάτων κατεύθυνσης με δυαδικής κατωφλίωσης τιμές της κλίσης. ([binary-thresholded gradient magnitudes..](#))

Όσον αφορά τη δεύτερη κατηγορία, σε αρκετές περιπτώσεις για την ανίχνευση ανθρώπων, χρησιμοποιείται ως βασικό χαρακτηριστικό αυτό της κίνησης, συνδυαζόμενο κατάλληλα με πληροφορία από την ένταση της εικόνας. Για παράδειγμα οι Viola et al [7] κατασκεύασαν έναν ανιχνευτή κινούμενου ανθρώπου χρησιμοποιώντας πληροφορία από την ανθρώπινη κίνηση και συνδυάζοντας την με την ανθρώπινη εμφάνιση. Στην περίπτωση μας δουλεύουμε με σταθερές εικόνες και χρησιμοποιούμε την τεχνική των

ιστογραμμάτων κατά κατεύθυνση παραγωγού, την οποία θα αναλύσουμε ενδελεχώς στη συνέχεια της αναφοράς αυτής. Σχετικά με την τεχνική αυτή, μπορούμε να πούμε ότι έχει χρησιμοποιηθεί κατά κόρον στον τομέα της όρασης υπολογιστών. [McConnell 1986, Bichsel 1991, Freeman and Roth 1995, Freeman et al. 1996, Lowe 1999, 2004, Belongie et al. 2002]. Ο Mc Connell [8] πρότεινε ιστογράμματα κατά την κλίση των ακμών σε γειτονικές χωρικές περιοχές για αναγνώριση προτύπων, μία τεχνική που αργότερα επεκτάθηκε σε ιστογράμματα κατά κατεύθυνση παραγωγού εικόνας από τους Bichsel [14] και Freeman and Roth [3]. Οι Freeman and Roth [3] χρησιμοποίησαν την τεχνική αυτή σε ένα σύστημα αναγνώρισης χειριών. Υπολόγισαν πολύ αποτελεσματικά το ιστόγραμμα 36 διαφορετικών κατευθύνσεων της παραγωγού εικόνας και στη συνέχεια προκάλεσαν θόλωση κατά την φορά της κύριας κατεύθυνσης, ώστε να παρέχουν ανεξαρτησία σε μικρές μεταβολές της κατεύθυνσης. Παρ' όλα αυτά όλες αυτές οι προσεγγίσεις κατάφεραν μόνο πρόσφατα να αγγίξουν την ωριμότητα όταν συνδυάστηκαν στη προσέγγιση του Lowe [9] με τοπικά χωρικά ιστογράμματα και κανονικοποιήσεις.

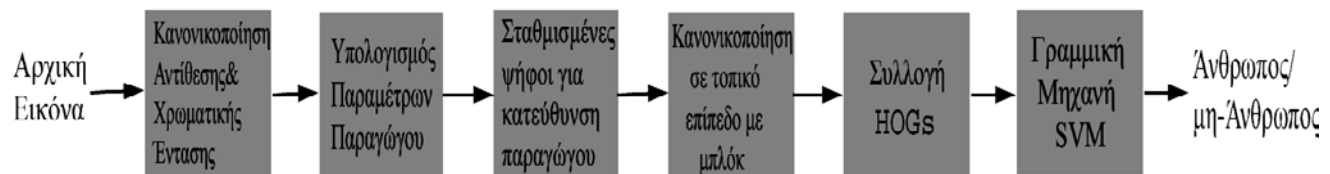
3.3 Η Μέθοδος που Επιλέξαμε

Στην εργασία μας βασιστήκαμε στη μέθοδο που προτείνουν οι Dalal & Triggs [11]. Για την εξαγωγή των χαρακτηριστικών της εικόνας για την περιγραφή του ανθρώπου χρησιμοποιήσαμε τα ιστογράμματα κατά κατεύθυνση παραγωγού. Ως βασικό χαρακτηριστικό επιλέξαμε την κυρίαρχη κατεύθυνση του ιστογράμματος. Με αυτά τα χαρακτηριστικά κωδικοποιήσαμε τις πληροφορίες από εικόνες που περιέχουν άνθρωπο, αλλά και από αυτές που δεν περιέχουν, εκπαιδύοντας τον ταξινομητή μας θετικά και αρνητικά αντίστοιχα. Η επιλογή μας αυτή οφείλεται στην πολύ απλή "αρχιτεκτονική" της μεθόδου σε συνδυασμό με τα εξαιρετικά αποτελέσματα της. Η μέθοδος αυτή καταφέρνει με πολύ απλό τρόπο να εκπληρώσει με πολύ μεγάλη επιτυχία τους στόχους της, ξεπερνώντας τα περισσότερα από τα προβλήματα που συναντούν άλλες μέθοδοι.

ΚΕΦΑΛΑΙΟ 4

4 Παρουσίαση της Μεθόδου

Στο κεφάλαιο αυτό δίνονται σε γενικές γραμμές τα βήματα που θα πρέπει να ακολουθήσουμε για να πετύχουμε το στόχο μας. Ολοκληρώνοντας όλα τα βήματα θα πρέπει να είμαστε σε θέση να αποφανθούμε μέσω του ταξινομητή μας αν ένα παράθυρο εικόνας προκαθορισμένου μεγέθους περικλείει ανθρώπινη μορφή ή όχι.



Σχήμα 4. Μία παρουσίαση της αλυσίδας των βημάτων που θα ακολουθήσουμε.

Για να φτάσουμε σε αυτό το σημείο, αρχικά πρέπει να εξάγουμε και να χρησιμοποιήσουμε ένα αποτελεσματικό σύνολο χαρακτηριστικών που θα μας δίνουν τη δυνατότητα να δημιουργήσουμε ένα ισχυρό διάνυμα χαρακτηριστικών που θα χαρακτηρίζει αποτελεσματικά την ανθρώπινη μορφή ξεπερνώντας τα όποια προβλήματα προαναφέραμε. Η βασική ιδέα είναι ότι η εμφάνιση και το σχήμα ενός αντικειμένου μπορούν να περιγραφούν ή μάλλον να χαρακτηριστούν με πολύ καλά αποτελέσματα από τη διακύμανση της τοπικής έντασης ή της κατεύθυνσης της παραγώγου του. Δηλαδή, ένα αντικείμενο και στην προκειμένη περίπτωση ο άνθρωπος και ειδικότερα η εμφάνιση και η φιγούρα του, χαρακτηρίζονται επαρκώς από την τοπική μέτρηση της έντασης και της κατεύθυνσης της παραγώγου της εικόνας του. Βασιζόμενοι σε αυτή την ιδέα, η μέθοδος μας χρησιμοποιεί κανονικοποιημένα τοπικά ιστογράμματα κατεύθυνσης παραγώγου σε ένα πυκνό πλέγμα πάνω στην

εικόνα. Τα ιστογράμματα αυτά ονομάζονται Ιστογράμματα κατά Κατεύθυνση Παραγώγου (Histogrammes of Oriented Gradients) και πολύ συχνά θα αναφερόμαστε σε αυτά από εδώ και στο εξής ως HOGs. Η υλοποίηση αφορά πάντα παράθυρο ή εικόνα σταθερού και καθορισμένου μεγέθους έτσι ώστε τα διανύσματα χαρακτηριστικών να έχουν πάντα το ίδιο μέγεθος. Η μέθοδος μας συνοψίζεται στην Σχήμα 4.

4.1 Τα Βήματα Της Μεθόδου

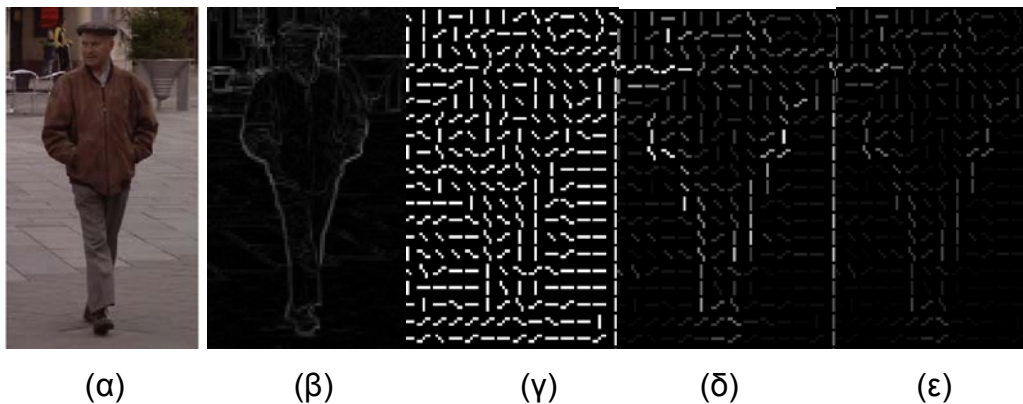
Για να επιτύχουμε το στόχο μας μέσω της μεθόδου που επιλέξαμε θα πρέπει να ακολουθήσουμε άλλες φορές πιστά και άλλες φορές όχι, τα παρακάτω βήματα:

- Για μεγαλύτερη ανεξαρτησία σε θέματα φωτισμού, σκίασης και σε άλλα τέτοια φαινόμενα που συνήθως επηρεάζουν τα αποτελέσματα κατά την επεξεργασία εικόνων ίσως φανεί χρήσιμο να κανονικοποιήσουμε το contrast του παραθύρου χωρικά και πιθανόν τοπικά πριν προβούμε στην εφαρμογή της μεθόδου εξαγωγής χαρακτηριστικών. Αυτό μπορεί να γίνει με πολλούς τρόπους όπως με γάμμα συμπίεση, με εξισορρόπηση του ιστογράμματος ή υπολογίζοντας ένα τοπικό μέγεθος σε μία μεγαλύτερη χωρικά περιοχή που θα ονομάσουμε μπλοκ και χρησιμοποιώντας το για να κανονικοποιήσουμε όλα τα κελιά που ανήκουν στο μπλοκ αυτό.

- Στη συνέχεια υπολογίζουμε τις παραμέτρους της πρώτης παραγώγου (πλάτος και κατεύθυνση) για κάθε pixel. Με αυτόν τον τρόπο συλλαμβάνουμε το περίγραμμα, τη σιλουέτα και μερικές πληροφορίες για την υφή, ενώ προσφέρουμε ακόμα μεγαλύτερη αντοχή σε αποκλίσεις του φωτισμού. Αν η εικόνα μας είναι έγχρωμη, υπολογίζουμε τις παραμέτρους της πρώτης παραγώγου για το τοπικά ισχυρότερο χρωματικό κανάλι, το οποίο παρέχει ανεξαρτησία σε χρωματικές αλλαγές σε αρκετά μεγάλο βαθμό.

- Το επόμενο στάδιο στοχεύει σε μία τέτοια κωδικοποίηση της πληροφορίας, ώστε να πετύχουμε ευαισθησία στο τοπικό περιεχόμενο της εικόνας και ανεξαρτησία σε μικρές αλλαγές της πόζας ή της εμφάνισης. Έτσι χρησιμοποιούμε την πληροφορία από την κατεύθυνση της παραγώγου τοπικά με παρόμοιο τρόπο με αυτόν των χαρακτηριστικών του αλγόριθμου SIFT (Lowe

2004) [9]. Χωρίζουμε την εικόνα σε μικρές χωρικές περιοχές ίδιου μεγέθους που τις ονομάζουμε κελιά. Για καθένα από αυτά τα κελιά φτιάχνουμε ένα τοπικό ιστόγραμμα κατευθύνσεων της παραγώγου, από κάθε pixel του. Το ιστόγραμμα αυτό δέχεται ως είσοδο τις διαφορετικές τιμές κατεύθυνσης από 0 έως 180 μοίρες κάθε pixel. Είναι πολύ πιθανό, η χρησιμοποίηση όλου του συνεχούς φάσματος κατεύθυνσης να είναι περιττή, οπότε για το λόγο αυτό μπορούμε να χρησιμοποιήσουμε ένα πεπερασμένο και προκαθορισμένο από εμάς αριθμό διαφορετικών τιμών κατεύθυνσης που ονομάζουμε Bins. Δηλαδή, το πλήθος των Bins αντιπροσωπεύει με περισσότερο ή λιγότερο πυκνό τρόπο το συνεχές φάσμα των κατευθύνσεων από 0 ως 180 μοίρες. Το κάθε pixel συνεισφέρει στο ιστόγραμμα του κελιού στο οποίο ανήκει ποσότητα ανάλογη με το πλάτος της παραγώγου του. Ως χαρακτηριστικό κάθε κελιού επιλέγουμε είτε ολόκληρο το ιστόγραμμα είτε την επικρατούσα γωνία του ιστογράμματος του. Στη δεύτερη περίπτωση το τελικό διάνυσμα χαρακτηριστικών διαμορφώνεται περιλαμβάνοντας ως δεύτερο χαρακτηριστικό κάθε κελιού το μέσο όρο της έντασης της παραγώγου αποκλειστικά και μόνο από τα pixel που έχουν κατεύθυνση ίδια με την επικρατούσα. Το χαρακτηριστικό αυτό, το ονομάζουμε βαρύτητα της επικρατούσας γωνίας.

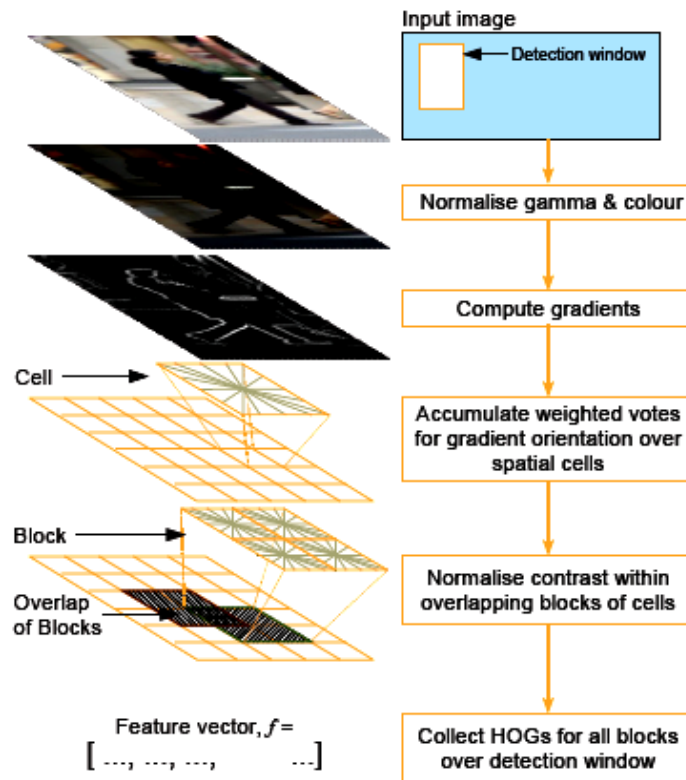


Σχήμα 5. Ο HOG ανιχνευτής μας στοχεύει κυρίως στο περίγραμμα της σιλουέτας(κυρίως το κεφάλι, τους ώμους και τα πόδια). (α) Η αρχική εικόνα προς δοκιμή. (β) Το πλάτος της παραγώγου της. (γ) Απεικόνιση των χαρακτηριστικών μας λαμβάνοντας υπ' όψιν μόνο την επικρατούσα γωνία κάθε κελιού. (δ) Απεικόνιση των ίδιων χαρακτηριστικών, πολλαπλασιάζοντας με τη βαρύτητα της επικρατούσας γωνίας κάθε κελιού. (ε) Η ίδια εικόνα μετά από κανονικοποίηση με φίλτρο Gauss.

- Πριν περάσουμε στη συλλογή ή κατασκευή των διανυσμάτων χαρακτηριστικών υπάρχει η δυνατότητα για κανονικοποίηση σε τοπικό επίπεδο. Αυτό μπορεί να επιτευχθεί με διάφορους τρόπους, η βασική ιδέα των οποίων είναι η ομαδοποίηση γειτονικών κελιών σε μεγαλύτερες χωρικές περιοχές που ονομάζουμε μπλοκ. Πιο συγκεκριμένα κατατάσσουμε κάθε κελί σε μία ομάδα 2x2 ή περισσότερων γειτονικών κελιών τα οποία αποτελούν ένα μπλοκ. Τα μπλοκ μπορούν να είναι επικαλυπτόμενα, δηλαδή κάθε κελί είναι δυνατόν να ανήκει σε περισσότερα από ένα μπλοκ. Στη συνέχεια πραγματοποιείται η κανονικοποίηση τοπικά ανά μπλοκ. Αυτό μπορεί να γίνει κανονικοποιώντας είτε τα χαρακτηριστικά των κελιών μετά την εξαγωγή τους είτε τις εντάσεις των pixels της παραγωγού των κελιών που ανήκουν σε κοινό μπλοκ. Στην πρώτη περίπτωση η κανονικοποίηση μπορεί να γίνει χρησιμοποιώντας ένα μέγεθος μέτρησης της τοπικής “ενέργειας” του ιστογράμματος, όπως η νόρμα, ενώ στη δεύτερη περίπτωση χρησιμοποιώντας ένα γκαουσιανό φίλτράρισμα τοπικά σε κάθε μπλοκ. Με την κανονικοποίηση σε τοπικό επίπεδο προσδοκούμε να επιτύχουμε μεγαλύτερη ανεξαρτησία του ανιχνευτή μας στη διακύμανση του φωτισμού, της φωτεινότητας, της αντίθεσης, αλλά και περιορισμό της άχρηστης πληροφορίας όπως ακμές στο φόντο ή θόρυβος που μπορεί να επηρεάσει αρνητικά τον ανιχνευτή μας.

- Στο τελικό στάδιο, συλλέγονται όλα τα HOGs από όλα τα μπλοκ και δημιουργούμε ένα διάνυσμα χαρακτηριστικών που θα χρησιμοποιήσουμε στον ταξινομητή μας.

Τα βήματα αυτά παρουσιάζονται με πολύ ωραίο τρόπο εικονικά στη διατριβή του Navneet Dalal [12] όπως φαίνεται στο Σχήμα 6, ενώ απεικόνιση των αποτελεσμάτων υλοποίησης της μεθόδου σε μία από τις εικόνες μας παρουσιάζεται στο Σχήμα 5 .



Σχήμα 6. Γραφική παρουσίαση της μεθόδου μας βήμα-βήμα από τον Navneet Dalal

Στην πράξη η υλοποίηση της μεθόδου μπορεί να διαφέρει ελαφρώς σε κάποια σημεία καθώς βήματα έχουν παραπονηθεί, προστεθεί ή παραληφθεί για λόγους βελτιστοποίησης. Η βασική ιδέα όμως της αναπαράστασης αυτής αξιοποιείται εξ ολοκλήρου και έχει αρκετά πλεονεκτήματα. Καταγράφει τη δομή της παραγώγου στην εικόνα και επομένως και τη δομή των ακμών η οποία χαρακτηρίζει έντονα το σχήμα του ζητούμενου αντικειμένου. Επίσης, αποκλίσεις στο φωτισμό και την αντίθεση της εικόνας, καθώς και περιστροφές ή μετακινήσεις των μελών, αν είναι μικρότερες του κελιού ή του πλήθους των Bins του ιστογράμματος, δεν αλλάζουν πολύ το αποτέλεσμα. Έτσι έχουμε σχετική αντοχή σε τοπικές γεωμετρικές και φωτομετρικές αλλαγές.



Σχήμα 7. Παράδειγμα εικόνων που χρησιμοποιήθηκαν στην εκπαίδευση της μηχανής SVM για την κλάση ‘άνθρωπος’

Στη συνέχεια, θα χρησιμοποιήσουμε τον αλγόριθμο αυτό εξαγωγής των διανυσμάτων χαρακτηριστικών για να εκπαιδεύσουμε τη μηχανή SVM θετικά με διανύσματα από παράθυρα εικόνας που περικλείουν άνθρωπο για την κλάση ‘άνθρωπος’ και αρνητικά, με διανύσματα από παράθυρα εικόνας που δεν περιέχουν άνθρωπο για την κλάση ‘μη-άνθρωπος’. Κάποιες ενδεικτικές εικόνες που χρησιμοποιήθηκαν για την εκπαίδευση της κλάσης άνθρωπος παρουσιάζονται στο Σχήμα 7. Μετά το πέρας της εκπαίδευσης του ταξινομητή, θα μπορούμε εξάγοντας διάνυσμα χαρακτηριστικών από οποιοδήποτε παράθυρο εικόνας του καθορισμένου μεγέθους να αποφανθούμε μέσω του ταξινομητή αν περικλείει άνθρωπο ή όχι.

ΚΕΦΑΛΑΙΟ 5

5 Υλοποίηση και Αξιολόγηση της Μεθόδου

Στο κεφάλαιο αυτό περιγράφουμε με περισσότερες λεπτομέρειες την υλοποίηση της μεθόδου μας και αναλύουμε την ευαισθησία του ανιχνευτή μας σε σημαντικές παραμέτρους της μεθόδου. Συγκεκριμένα παρουσιάζονται δύο υλοποιήσεις, η διαφορετικότητα των οποίων έγκειται στα χαρακτηριστικά που εξάγουμε από τα HOGs. Στην πρώτη υλοποίηση ως χαρακτηριστικά για κάθε κελί επιλέγουμε την επικρατούσα γωνία του αντίστοιχου HOG και το μέσο όρο της έντασης της παραγώγου, αποκλειστικά και μόνο από τα pixels που έχουν κατεύθυνση ίδια με την επικρατούσα. Στην υλοποίηση αυτή θα αναφερόμαστε ως Υλοποίηση με Μια Επικρατούσα Κατεύθυνση. Αντίθετα με την πρώτη, στη δεύτερη υλοποίηση δεν απορρίπτουμε πληροφορία από το ιστόγραμμά μας, αλλά το αξιοποιούμε ολόκληρο. Έτσι ως χαρακτηριστικό για κάθε κελί επιλέγουμε όλες τις τιμές του αντίστοιχου HOG. Στην υλοποίηση αυτή θα αναφερόμαστε ως Υλοποίηση με Περισσότερες Κατευθύνσεις. Για κάθε μία από τις υλοποιήσεις εξετάζουμε την απόδοσή τους και αξιολογούμε κάποιες προσθήκες ή τροποποιήσεις στη μέθοδο. Όσον αφορά την αξιολόγηση κάθε υλοποίησης και σε κάθε πειραματισμό στις παραμέτρους της μεθόδου μας, πρέπει να αναφέρουμε ότι ο εκάστοτε ανιχνευτής μας ελέγχθηκε σε ένα σύνολο 300 εικόνων παρουσία ανθρώπου και σε ένα σύνολο 300 εικόνων χωρίς άνθρωπο. Οι εικόνες αυτές επιλέχθηκαν από τη βάση δεδομένων του INRIA και από το διαδίκτυο μετά από σχετική αναζήτηση στο Google Image και περιέχουν ανθρώπους σε απροσδιόριστες σκηνικές συνθήκες. Από το πρώτο σύνολο εικόνων υπολογίσαμε την απόδοση του ανιχνευτή μας στην επιτυχημένη ανίχνευση ανθρώπων, δηλαδή αξιολογήσαμε την ικανότητα και αξιοπιστία του στην ταξινόμηση εικόνων, όπου υπάρχει άνθρωπος στην κλάση άνθρωπος. Από το δεύτερο σύνολο εικόνων εκτιμήσαμε την απόδοση του ανιχνευτή μας, υπολογίζοντας τη λανθασμένη ανίχνευση ανθρώπων σε εικόνες που δεν

υπάρχουν. Εκτιμήσαμε δηλαδή την αποτελεσματικότητα και αξιοπιστία του στην ταξινόμηση εικόνων απουσία ανθρώπου στην κλάση μη-άνθρωπος, υπολογίζοντας το αντίστοιχο σφάλμα.

5.1 Υλοποίηση με Μία Επικρατούσα Κατεύθυνση

Στο σημείο αυτό περιγράφουμε την πρώτη μας υλοποίηση παρουσιάζοντας στην πράξη τα βήματα της μεθόδου μας. Στην πρώτη ενότητα δίνεται λεπτομερώς η υλοποίηση του αρχικού μας αλγόριθμου, με τις προεπιλεγμένες παραμέτρους, στον οποίο θα αναφερόμαστε ως αλγόριθμο αναφοράς. Στις επόμενες ενότητες επεμβαίνουμε στα βασικά στάδια υλοποίησης του αλγόριθμού μας και αξιολογούμε την επίδραση των επεμβάσεων αυτών, συγκριτικά με τον αλγόριθμο αναφοράς. Συγκεκριμένα, στη δεύτερη ενότητα επεμβαίνουμε στην αρχική μας εικόνα, κανονικοποιώντας πριν την εφαρμογή του αλγόριθμου. Στην τρίτη, στο μέγεθος των κελιών και στην τέταρτη στο πλήθος των bins. Στην πέμπτη ενότητα επιχειρούμε νέα κανονικοποίηση, αλλά αυτή τη φορά σε τοπικό επίπεδο και τέλος στην έκτη ενότητα επεμβαίνουμε στο μέγεθος του συνόλου των εικόνων προς εκπαίδευση.

5.1.1 Αλγόριθμος αναφοράς

-Κανονικοποίηση αντίθεσης / γάμμα: Ο αρχικός μας αλγόριθμος εφαρμόζεται πάνω στην εικόνα μας, αυτούσια, χωρίς να επέμβουμε στην αντίθεση ή να επιδιορθώσουμε το γάμμα. Αν η εικόνα μας είναι έγχρωμη χρησιμοποιούμε το μοντέλο RGB και επιλέγουμε το χρωματικό κανάλι με τη μεγαλύτερη νόρμα.

-Υπολογισμός παραμέτρων παραγώγου: Το πλάτος και η κατεύθυνση της παραγώγου της εικόνας υπολογίζονται με τις διαφορές των pixel, αντίστοιχα ως

$$\begin{aligned} \text{εξής : } \quad M_{ij} &= \sqrt{(A_{i,j} - A_{i+1,j})^2 + (A_{i,j} - A_{i,j+1})^2} \\ R_{ij} &= a \tan 2(A_{i,j} - A_{i+1,j}, A_{i,j+1} - A_{i,j}) \end{aligned}$$

-Εξαγωγή διανύσματος χαρακτηριστικών: Έχουμε ήδη χωρίσει την εικόνα σε κελιά μεγέθους 6x6. Στη συνέχεια χρησιμοποιούμε τις παραμέτρους της

παραγώγου για να δημιουργήσουμε τα ιστογράμματα κατά κατεύθυνση παραγώγου για κάθε κελί. Το ιστόγραμμα αυτό δέχεται 9 διαφορετικές τιμές κατεύθυνσης, δηλαδή 9 τιμές που κυμαίνονται από 0 ως 180 μοίρες. Το κάθε ρixel συνεισφέρει στο ιστόγραμμα του κελιού στο οποίο ανήκει και συγκεκριμένα στην κατεύθυνση του, ποσότητα ανάλογη με το πλάτος της παραγώγου του.

Αφού έχουμε δημιουργήσει από ένα ιστόγραμμα κατά κατεύθυνση παραγώγου για κάθε κελί, επιλέγουμε ως χαρακτηριστικό κάθε κελιού την επικρατούσα γωνία του ιστογράμματος. Με αυτό τον τρόπο συγκεντρώνοντας τα χαρακτηριστικά από κάθε κελί δημιουργούμε το διάνυσμα χαρακτηριστικών της εικόνας μας. Επανερχόμαστε σε κάθε κελί και επιλέγουμε σαν δεύτερο χαρακτηριστικό του κελιού το μέσο όρο της έντασης της παραγώγου αποκλειστικά και μόνο από τα ρixel που έχουν κατεύθυνση ίδια με την επικρατούσα. Συγκεντρώνοντας από κάθε κελί και τα δεύτερα χαρακτηριστικά και προσαρμόζοντας τα στο ήδη υπάρχον διάνυσμα δημιουργούμε το τελικό διάνυσμα χαρακτηριστικών .

-Κανονικοποίηση σε μπλοκ: Στον αρχικό μας αλγόριθμο δεν χρησιμοποιήθηκε καμία τεχνική κανονικοποίησης σε τοπικό επίπεδο ή πιο συγκεκριμένα σε επίπεδο μπλοκ. Προτιμήθηκε αυτή η προσέγγιση έτσι ώστε να ισχυροποιηθεί ο αρχικός αλγόριθμος ως αλγόριθμος αναφοράς και να αξιολογήσουμε καλύτερα την απόδοση του, όταν δεν υπάρχει καμία αλλοίωση της αρχικής πληροφορίας που μας δίνει μία εικόνα.

-Παράθυρο ανίχνευσης: Όλα τα παραπάνω βήματα θα πρέπει να εφαρμοστούν σε μία εικόνα ή ένα παράθυρο εικόνας συγκεκριμένου μεγέθους, το οποίο ονομάζουμε παράθυρο ανίχνευσης. Έχοντας στο μυαλό μας ότι το παράθυρο μας προορίζεται να ανιχνεύει κατά βάση ανθρώπους σε όρθια θέση, επιλέγουμε σαν μέγεθος του 132x84 ρixel. Ένας άνθρωπος θα μπορεί να ανιχνευτεί από το παράθυρο μας, αν περιβάλλεται από αυτό, αφήνοντας ένα μικρό περιθώριο γύρω από τον άνθρωπο και στις τέσσερις πλευρές του παραθύρου.

-Dataset & Ταξινομητής: Ο ταξινομητής που χρησιμοποιήσαμε για το διαχωρισμό των κλάσεων "άνθρωπος" και "μη-άνθρωπος" είναι μία γραμμική μηχανή SVM. Για την εκπαίδευση του ταξινομητή χρησιμοποιήσαμε 854,

132x84 θετικές εικόνες από τη βάση δεδομένων του INRIA και 400 εικόνες χωρίς παρουσία ανθρώπου για κάθε μία από τις οποίες χρησιμοποιήσαμε 5 τυχαία παράθυρα 132x84, δηλαδή σύνολο 2000 αρνητικές εικόνες. Οι θετικές μας εικόνες περιλαμβάνουν ανθρώπους σε όρθια στάση, από μπροστά και από πίσω, σε διάφορες θέσεις και πόζες, διάφορους φωτισμούς και με πολλά και διαφορετικά φόντα.

-Αξιολόγηση: Ο ανιχνευτής μας ελέγχθηκε και με χρήση του αλγόριθμου αναφοράς κατάφερε με επιτυχία να ανιχνεύσει την ανθρώπινη παρουσία σε ποσοστό 80,14% ενώ ανίχνευσε λανθασμένα άνθρωπο, ενώ δεν υπήρχε σε ποσοστό 15,1%.

5.1.2 Κανονικοποίηση της εικόνας πριν την εφαρμογή του αλγόριθμου

Αξιολογήσαμε 2 διαφορετικές μεθόδους κανονικοποίησης της φωτεινότητας και της χρωματικής αντίθεσης της εικόνας, ώστε να επιτύχουμε μεγαλύτερη αντοχή του ανιχνευτή σε μεταβολές τους: την κανονικοποίηση γάμμα και την εξισορρόπηση του ιστογράμματος. Με τον τρόπο αυτό προσδοκούμε να βγάλουμε κάποιο συμπέρασμα για την ιδανικότερη “προετοιμασία” της αρχικής εικόνας, ώστε να ενισχύσουμε την βέλτιστη εξαγωγή των χαρακτηριστικών μας.

5.1.2.1 Γάμμα κανονικοποίηση

Επιτύχαμε γάμμα συμπίεση χρησιμοποιώντας την τετραγωνική ρίζα κάθε χρωματικού καναλιού του μοντέλου RGB. Με αυτόν τον τρόπο προσπαθήσαμε να βελτιώσουμε την αντοχή του ανιχνευτή μας στις απροσδιόριστες χρωματικές αντιθέσεις. Στην επιλογή της τετραγωνικής ρίζας μας ώθησε το γεγονός ότι ο θόρυβος λόγω του φωτισμού στον ανιχνευτή CCD (Charge-Coupled Device) είναι ανάλογος της τετραγωνικής ρίζας της έντασης. Συμπεριλαμβάνοντας την προσθήκη αυτή στον αρχικό ανιχνευτή μας παρατηρήσαμε μείωση της

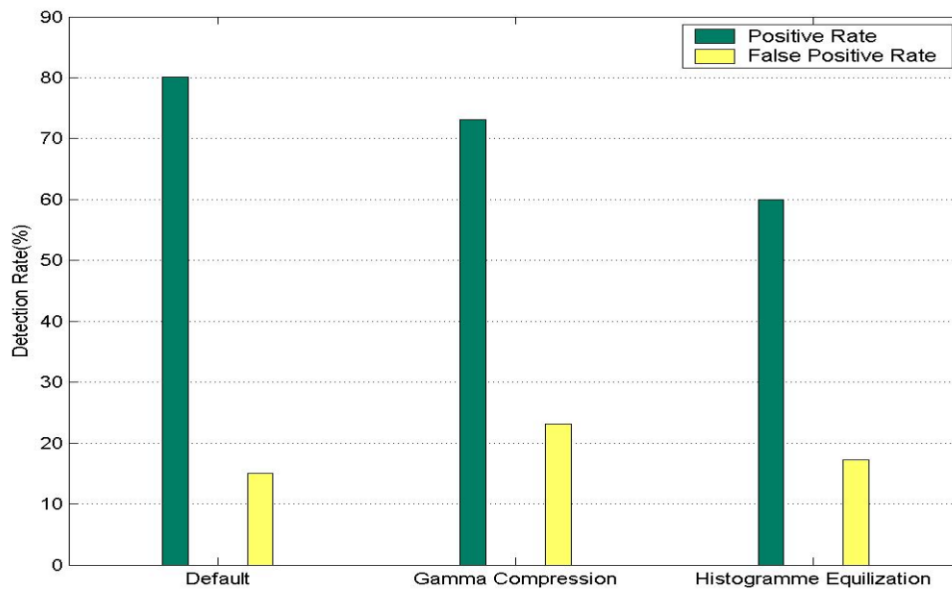
απόδοσης του της τάξης του 7.07%. Η λανθασμένη αναγνώριση ανθρώπου κυμάνθηκε γύρω στο 23,2%, που αποτελεί μία αρκετά αισθητή αύξηση.

5.1.2.2 Εξισορρόπηση Ιστογράμματος

Εφαρμόσαμε τον αλγόριθμο μας για εξαγωγή των διανυσμάτων χαρακτηριστικών πάνω στην εικόνα, αφού πρώτα εξισορροπήσαμε τις τιμές του ιστογράμματος έντασης της. Με τον τρόπο αυτό ελπίσαμε να επιτύχουμε μεγαλύτερη όξυνση των ακμών της εικόνας. Παρατηρήσαμε ότι η σκέψη μας αυτή μείωσε την απόδοση του ανιχνευτή κατά 20,14%, ενώ η λανθασμένη ανίχνευση επίσης αυξήθηκε στο 17,2%.

Αξιολόγηση

Το Σχήμα 8 συνοψίζει τις επιδράσεις των προσεγγίσεων μας για κανονικοποίηση στην αρχική εικόνα. Η κανονικοποίηση αυτή πραγματοποιείται σε γενικό και όχι σε τοπικό επίπεδο και αφορά την εικόνα που θα χρησιμοποιήσουμε ως είσοδο στον αλγόριθμο μας. Παρατηρώντας το Σχήμα 8 καταλήγουμε στο συμπέρασμα ότι οποιαδήποτε προσπάθεια βελτιστοποίησης της αρχικής εικόνας για ισχυρότερη εξαγωγή των χαρακτηριστικών της δεν έχει τα αναμενόμενα αποτελέσματα. Μπορούμε να αποδώσουμε τα αρνητικά αποτελέσματα σε αλλοίωση σημαντικής πληροφορίας κατά τη λανθασμένη προσπάθεια μας να κανονικοποιήσουμε χρωματικά την εικόνα μας.



Σχήμα 8. Η κανονικοποίηση στην αρχική εικόνα δε φαίνεται να μας δίνει τα επιθυμητά αποτελέσματα. Με πράσινο θα απεικονίζεται η επιτυχής ανίχνευση ανθρώπων (%) και με κίτρινο η λανθασμένη (%).

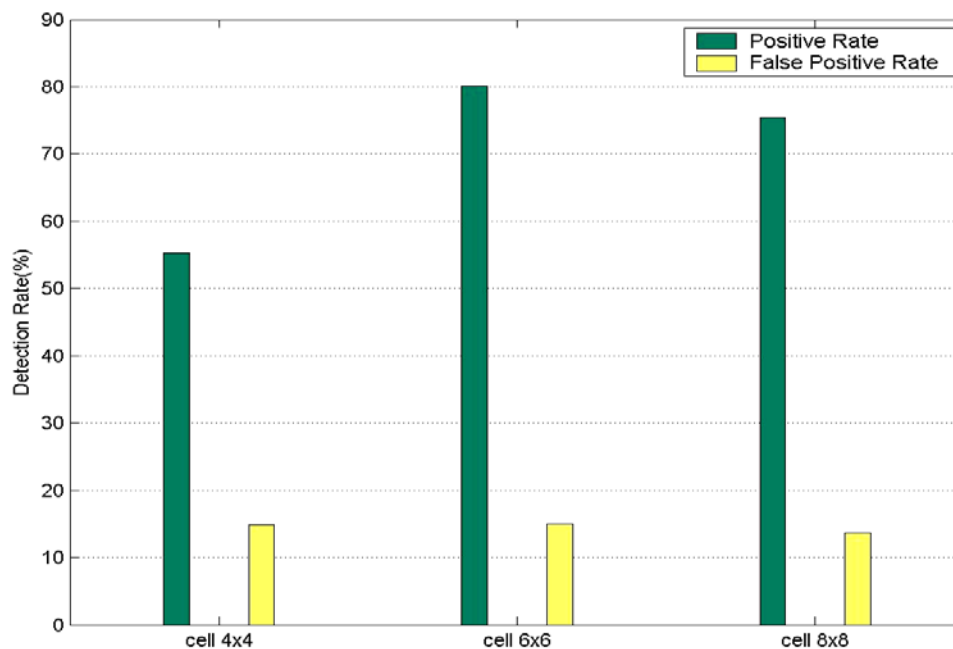
5.1.3 Μεταβολή στο μέγεθος του κελιού

Η μέθοδος μας βασίζεται στην εξαγωγή χαρακτηριστικών από το τοπικό περιεχόμενο της εικόνας. Είναι λογικό λοιπόν το μέγεθος του κελιού που επιλέγουμε να αποτελεί κρίσιμη παράμετρο στην απόδοση του ανιχνευτή μας. Βασιζόμενοι στο μέγεθος του παράθυρου ανίχνευσης, άρα και του ανθρώπου που περιβάλλει χρησιμοποιήσαμε μέγεθος κελιού 6x6 για τον αρχικό μας αλγόριθμο και για να ελέγξουμε την ευαισθησία του στο μέγεθος αυτό, αυξήσαμε και μειώσαμε το μέγεθος του σε 4 και 8 pixels αντίστοιχα. Δεν επιχειρήσαμε περαιτέρω μείωση ή αύξηση του μεγέθους του κελιού καθώς μια τέτοια κίνηση θα περιόριζε την τοπικότητα της πληροφορίας που επιθυμούμε να "αιχμαλωτίσουμε". Η χρήση κελιού μεγέθους 4x4 pixels μείωσε την απόδοση του ανιχνευτή μας κατά 24,8%, επηρεάζοντας ελάχιστα την λανθασμένη ανίχνευση ανθρώπων (14,87%). Αντίθετα η χρήση κελιού μεγέθους 8x8 διατήρησε την

απόδοση του ανιχνευτή μας σε σχετικά καλά επίπεδα της τάξης του 75,43% υποδεικνύοντας παρά όλα αυτά πως το βέλτιστο μέγεθος κελιού είναι αυτό που χρησιμοποιήθηκε στον αλγόριθμο αναφοράς. Αξίζει να σημειωθεί πως το μέγεθος κελιού 8x8 μείωσε τη λανθασμένη ανίχνευση ανθρώπων κατά 1,37%.

Αξιολόγηση

Τα αποτελέσματα στην απόδοση του ανιχνευτή μας από τη χρήση κελιών διαφορετικού μεγέθους επικεντρώνονται στο Σχήμα 9. Είναι ξεκάθαρο πως τα μεγέθη 6x6 και 8x8 έχουν σαφές προβάδισμα έναντι του 4x4, που φαίνεται να αποτελεί πολύ μικρό μέγεθος κελιού για να είναι αντιπροσωπευτικό της επικρατούσας κατεύθυνσης της παραγώγου ανθρώπινης σιλουέτας. Το μέγεθος 6x6 δείχνει να υπερτερεί στην αποτελεσματική ανίχνευση των ανθρώπων αλλά να υστερεί στο θέμα της λανθασμένης ανίχνευσης τους συγκριτικά με το 8x8 κατά ένα μικρό ποσοστό



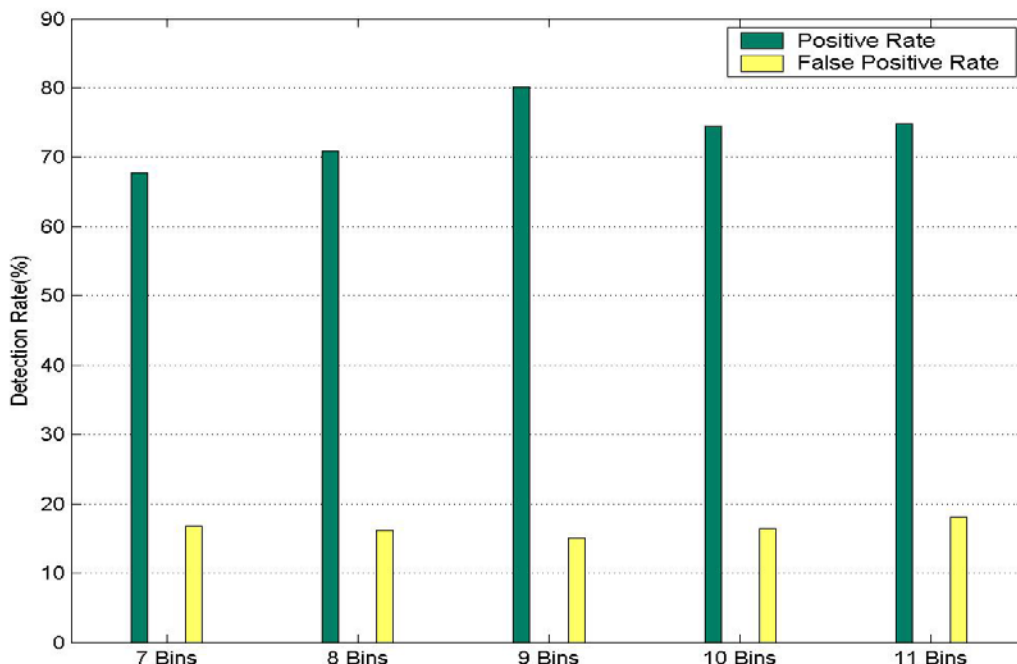
Σχήμα 9. Μεταβολή της απόδοσης του ανιχνευτή μας καθώς αυξάνουμε το μέγεθος του κελιού.

5.1.4 Μεταβολή πλήθους των Bins

Εφόσον ο αποτελεσματικός υπολογισμός της κατεύθυνση της παραγωγού είναι κρίσιμος για την επιτυχία του ανιχνευτή μας, στο σημείο αυτό έχουμε σαν στόχο να αξιολογήσουμε την ευαισθησία του στον αριθμό των διαφορετικών τιμών της κατεύθυνσης που ισοκατανέμονται μεταξύ 0-180 μοίρες. Στον αρχικό μας ανιχνευτή χρησιμοποιήσαμε 9 διαφορετικές τιμές Bins τις οποίες μειώσαμε και αυξήσαμε σε 7, 8 και 10,11 αντίστοιχα για το σκοπό αυτό. Στην περίπτωση μείωσης των Bins σε 7 η απόδοση του ανιχνευτή μας μειώθηκε στο 67,74%, ενώ όταν τα αυξήσαμε σε 8 το ποσοστό μας βελτιώθηκε ελάχιστα σε 70,86%. Αυξάνοντας τα Bins σε 10 ή 11 η απόδοση μας μεταβλήθηκε σε 74,41 και 74,79 αντίστοιχα. Η λανθασμένη ανίχνευση ανθρώπων αυξάνοντας τον αριθμό των bins από 7 σε 11 παρατηρήσαμε ότι μεταβλήθηκε από 16,74% σε 16,13% και στη συνέχεια από 16,4% σε 18,1% στην περίπτωση των 11 bins.

Αξιολόγηση

Στο Σχήμα 10 συνοψίζονται τα αποτελέσματα που πήραμε μεταβάλλοντας το νούμερο των bins. Είναι εύκολο να παρατηρήσει κανείς την αύξηση της απόδοσης του ανιχνευτή μας καθώς αυξάνουμε το νούμερο των bins από 7 μέχρι 9 και την σταθεροποίηση της ή πιο συγκεκριμένα ελαφριά μείωση της αυξάνοντας περαιτέρω το νούμερο τους. Επίσης χαρακτηριστική είναι η αύξηση του ποσοστού της λανθασμένης ανίχνευσης ανθρώπου καθώς αυξάνουμε το νούμερο των bins πάνω από 9.



Σχήμα 10. Το πλήθος των Bins φαίνεται να επηρεάζει την απόδοση σημαντικά

5.1.5 Κανονικοποίηση σε τοπικό επίπεδο μπλοκ

Η ένταση της παραγωγού παίρνει τιμές που κυμαίνονται σε ένα μεγάλο εύρος εξαιτίας των τοπικών διακυμάνσεων του φωτισμού και του φόντου. Είναι λοιπόν λογικό πως μία τοπική κανονικοποίηση θα είναι απαραίτητη για την αποτελεσματικότητα του ανιχνευτή μας. Υπάρχουν πολλοί τρόποι για να κανονικοποιήσουμε τοπικά μία εικόνα. Σχεδόν όλοι βασίζονται στη συγκέντρωση κελιών δημιουργώντας μία μεγαλύτερη χωρικά περιοχή στην εικόνα, τα μπλοκ και εφαρμογή της κανονικοποίησης σε αυτά. Ουσιαστικά τα μπλοκ αυτά μπορούν να είναι επικαλυπτόμενα έτσι ώστε να επιτύχουμε ισχυρότερη κανονικοποίηση. Στη συγκεκριμένη αναφορά αξιολογήσαμε την αποτελεσματικότητα 2 μεθόδων κανονικοποίησης σε τοπικό επίπεδο, οι οποίες παρατίθενται στη συνέχεια:

5.1.5.1 Κανονικοποίηση με χρήση της νόρμας

Η μέθοδος αυτή εφαρμόζεται μετά την εξαγωγή των χαρακτηριστικών στα κελιά μίας εικόνας και κανονικοποιεί τις τιμές των χαρακτηριστικών κάθε κελιού χρησιμοποιώντας σαν μέγεθος κανονικοποίησης την L2-νόρμα του εκάστοτε μπλοκ. Αφού έχουμε εξαγάγει τα χαρακτηριστικά κάθε κελιού, ομαδοποιούμε τα κελιά σε μπλοκ μεγέθους 2x2 κελιών και διαιρούμε κάθε στοιχείο του με τη νόρμα του μπλοκ στο οποίο ανήκει. Τα μπλοκ είναι επικαλυπτόμενα. Ακολουθεί αριθμητική παρουσίαση της L2-νορμας και της εφαρμογής της:

$$L2 - norm: v \leftarrow v / \sqrt{\|v\|_2^2 + \varepsilon^2}$$

όπου v το μπλοκ πριν την κανονικοποίηση, $\|v\|_k$ η k -νόρμα και ε μία πολύ μικρή σταθερά κανονικοποίησης.

Προσθέτοντας την κανονικοποίηση αυτή στον αρχικό μας αλγόριθμο η απόδοση του μεταβλήθηκε ελάχιστα κατεβαίνοντας στο 78,7% ενώ η λανθασμένη ανίχνευση ανθρώπων παρέμεινε στα ίδια επίπεδα (15,6%).

5.1.5.2 Κανονικοποίηση με χρήση φίλτρου Gauss

Η μέθοδος αυτή εφαρμόζεται στην παράγωγο της εικόνας, τοπικά, σε μπλοκ μεγέθους 2x2 κελιών. Κάθε μπλοκ φιλτράρεται από ένα Γκαουσιανό φίλτρο με τυπική απόκλιση ίση με το μισό ης ελάχιστης τιμής του ιστογράμματος του κάθε μπλοκ. Με το Γκαουσιανό φίλτρο επιτυγχάνουμε ένα επιλεκτικό φιλτράρισμα δίνοντας βαρύτητα στις ακμές του ανθρώπινου σώματος και περιορίζοντας το θόρυβο και ακμές από το φόντο. Στη φάση αυτή δοκιμάσαμε δύο μερικώς διαφορετικές προσεγγίσεις εφαρμογής του φίλτρου.

Κανονικοποίηση Πρώτη

Στην προσέγγιση αυτή το φιλτράρισμα έγινε εξαρχής, πριν δηλαδή την δημιουργία των ιστογραμμάτων κατά κατεύθυνση παραγωγού και με αυτόν τον τρόπο διαμόρφωσε ολόκληρη την πορεία εξαγωγής και των δύο χαρακτηριστικών ανά κελί. Το αποτέλεσμα δεν ήταν το αναμενόμενο και η

απόδοση του ανιχνευτή μας έπεσε στο 75,47%, ενώ η λανθασμένη ανίχνευση ανθρώπων ανέβηκε στο 18,05%.

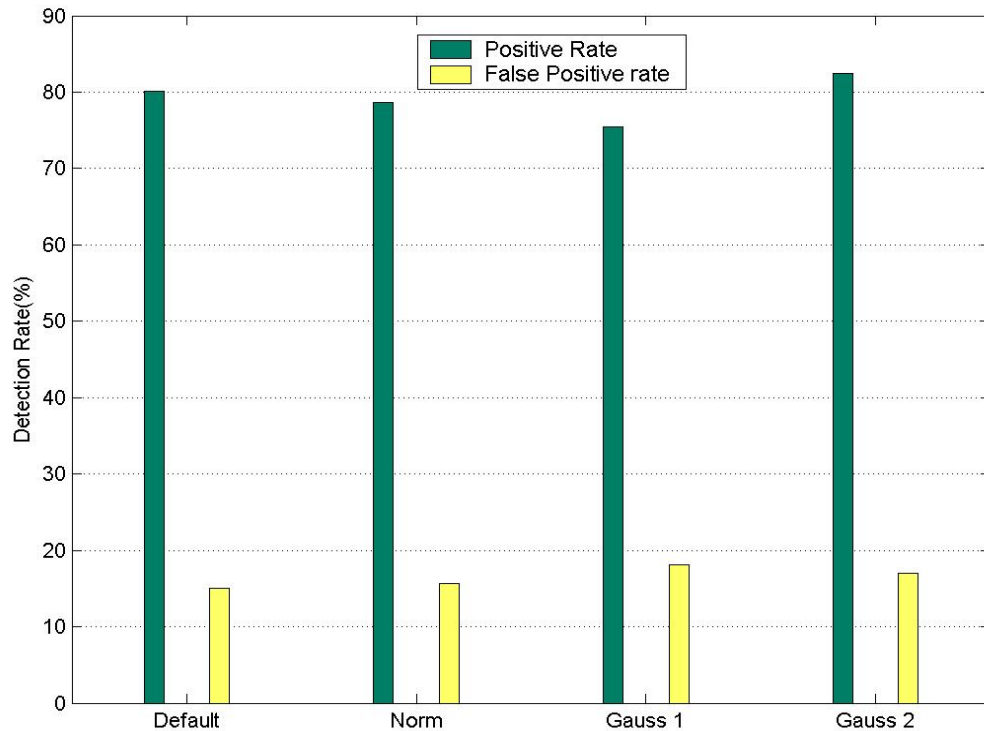
Κανονικοποίηση Δεύτερη

Στην προσέγγιση αυτή το φιλτράρισμα έγινε μετά τη δημιουργία των ιστογραμμάτων κατά κατεύθυνση παραγώγου και πριν την εξαγωγή του δεύτερου χαρακτηριστικού από κάθε κελί. Ως γνωστόν το δεύτερο χαρακτηριστικό του κελιού αποτελεί ο μέσος όρος της έντασης της παραγώγου αποκλειστικά και μόνο από τα ρixel που έχουν κατεύθυνση ίδια με την επικρατούσα. Το φιλτράρισμα, δηλαδή διαμορφώνει αποκλειστικά και καθοριστικά την ένταση της παραγώγου των ρixel με ίδια κατεύθυνση με την επικρατούσα του κελιού στο οποίο ανήκει, ενισχύοντας τα ρixel αυτά που καθορίζουν την ανθρώπινη φιγούρα και αποδυναμώνοντας αυτά που προκύπτουν από θόρυβο ή ακμές του φόντου. Με την προσέγγιση αυτή παρατηρήσαμε βελτίωση του ποσοστού απόδοσης της τάξης του 2,22%, ενώ το ποσοστό λανθασμένης ανίχνευσης ανθρώπων ανέβηκε στο 17%.

Αξιολόγηση

Παρατηρώντας τα αποτελέσματα από τη χρήση ή τη μη χρήση της κανονικοποίησης στο τοπικό επίπεδο των μπλοκ που παρατίθενται στο Σχήμα 11, αλλά και το αντίστοιχο Σχήμα 8 μπορούμε να συμπεράνουμε πως οποιαδήποτε κανονικοποίηση ή φιλτράρισμα της εικόνας πριν την εξαγωγή των ιστογραμμάτων κατά κατεύθυνση παραγώγου με αποτέλεσμα τη θόλωση της, είναι επισφαλής. Αυτό επιβεβαιώνει το γεγονός ότι αρκετά χρήσιμη πληροφορία για την ανθρώπινη φιγούρα και συγκεκριμένα για την κατεύθυνση της παραγώγου της περιέχεται στην αρχική εικόνα μέσα από τις έντονες και απότομες ακμές. Άρα η δημιουργία των ιστογραμμάτων κατά κατεύθυνση παραγώγου πρέπει να γίνεται με την παράγωγο της αρχικής εικόνας και όχι με κάποια κανονικοποίηση της ή της παραγώγου της. Όσον αφορά την

κανονικοποίηση χρησιμοποιώντας τη νόρμα κάθε μπλοκ, δεν υπήρχαν ορατές και ξεκάθαρες διαφοροποιήσεις από την παράλειψή της.



Σχήμα 11. Από όλες τις μεθόδους κανονικοποίησης σε τοπικό επίπεδο μόνο η δεύτερη μέθοδος χρησιμοποίησης Γκαουσιανού Φίλτρου δείχνει να αυξάνει την απόδοση μας

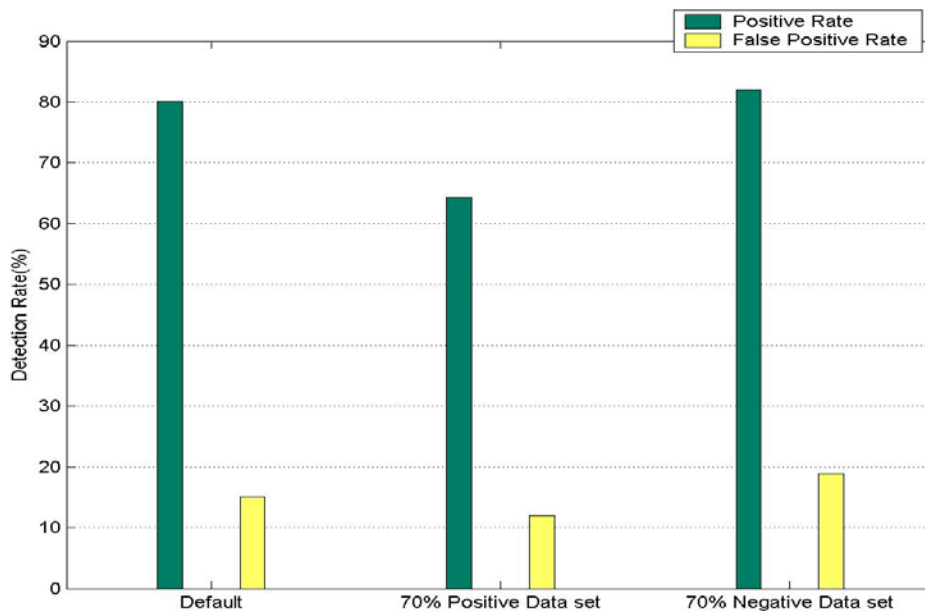
5.1.6 Μεταβολή στο πλήθος των εικόνων προς εκπαίδευση

Αν εξετάσουμε ως προς το πλήθος τα αντικείμενα που ανήκουν στην κλάση “άνθρωπος” με αυτά που ανήκουν στην κλάση “μη-άνθρωπος” είναι αναμενόμενο να καταλήξουμε σε μία χαώδη διαφορά υπέρ της δεύτερης. Η σκέψη αυτή μας οδηγεί στο συμπέρασμα ότι η εκπαίδευση του ταξινομητή μας για την κλάση “μη-άνθρωπος” σίγουρα πρέπει να γίνει με κατά πολύ περισσότερες τον αριθμό εικόνες, έναντι των αντίστοιχων για την εκπαίδευση

της κλάσης “άνθρωπος”, χωρίς όμως να ξέρουμε ποια είναι η ιδανική αναλογία και αν αποτελεί σημαντική παράμετρο για την απόδοση του ανιχνευτή μας. Για το λόγο αυτό εκπαιδεύσαμε εκ νέου τη μηχανή SVM μεταβάλλοντας την αναλογία αυτή. Αρχικά κρατήσαμε το 70% των θετικών εικόνων προς εκπαίδευση, διατηρώντας το πλήθος των αρνητικών και παρατηρήσαμε βελτίωση στην απόδοση του ανιχνευτή μας στο 12,01% όσο αφορά τη λανθασμένη ανίχνευση ανθρώπων, που συνοδεύονταν όμως από μείωση της επιτυχούς ανίχνευσης στο 64,25%. Στη συνέχεια κινηθήκαμε αναλόγως, κρατώντας το 70% των αρνητικών προς εκπαίδευση εικόνων και επαναφέραμε τις θετικές στο αρχικό τους πλήθος. Η απόδοση του ανιχνευτή μας βελτιώθηκε στο 81,94%, ανεβάζοντας όμως τη λανθασμένη ανίχνευση ανθρώπων στο 18,82%.

Αξιολόγηση

Είναι εύκολο να διαπιστώσει κανείς κοιτάζοντας το Σχήμα 12 ότι αυξάνοντας το μερίδιο των αρνητικών εικόνων στην εκπαίδευση του SVM επιτυγχάνεται η ζητούμενη μείωση του ποσοστού λανθασμένης ανίχνευσης ανθρώπων, θυσιάζοντας όμως σημαντικό ποσοστό από την απόδοση του ανιχνευτή μας στην επιτυχημένη ανίχνευση ανθρώπων. Ομοίως αυξάνοντας το μερίδιο των θετικών εικόνων στην εκπαίδευση του SVM, πέραν της ζητούμενης βελτίωσης του ποσοστού επιτυχούς ανίχνευσης, έχουμε και τη μη επιθυμητή αύξηση της λανθασμένης ανίχνευσης ανθρώπων. Συμπεραίνουμε λοιπόν πως η αναλογία των δεδομένων προς εκπαίδευση αποτελεί σημαντική παράμετρο για την απόδοση του ανιχνευτή μας.



Σχήμα 12. Η αναλογία των θετικών και αρνητικών εικόνων προς εκπαίδευση έχει σημαντική επίδραση στην απόδοση του ανιχνευτή μας.

5.2 Υλοποίηση με Περισσότερες Κατευθύνσεις

Στο σημείο αυτό θα παρουσιάσουμε τη δεύτερη μας υλοποίηση. Όπως έχουμε προαναφέρει η βασική διαφοροποίηση της υλοποίησης αυτής είναι στην επιλογή των χαρακτηριστικών της εικόνας μέσα από τα HOGs. Ακολουθώντας πορεία ανάλογη της υλοποίησης με μία επικρατούσα κατεύθυνση, παρουσιάζουμε στην πρώτη ενότητα τον αλγόριθμο αναφοράς μας και στις επόμενες ενότητες επεμβαίνουμε στα βασικά στάδια υλοποίησης του αλγόριθμου μας και αξιολογούμε την επίδραση των επεμβάσεων αυτών, συγκριτικά με τον αλγόριθμο αναφοράς.

5.2.1 Αλγόριθμος αναφοράς

Ο αρχικός μας αλγόριθμος επιλέγεται να είναι κοινός με τον αλγόριθμο αναφοράς της υλοποίησης με μία επικρατούσα κατεύθυνση, οπότε για εκτενείς λεπτομέρειες και συγκεκριμένες αναφορές σε κάθε βήμα της μεθόδου μπορούμε να ανατρέξουμε στην αντίστοιχη παράγραφο της προηγούμενης υλοποίησης. Στην παράγραφο αυτή αναφέρουμε επιγραμματικά, για διευκόλυνση, τα στάδια και τις τιμές των παραμέτρων που χρησιμοποιούμε και αναλυτικά το μοναδικό κομμάτι που διαφοροποιείται, τη συλλογή των χαρακτηριστικών της εικόνας δηλαδή.

Αρχικά, υπολογίζουμε τις παραμέτρους της παραγώγου της εικόνας μας. Στη συνέχεια η εικόνα μας χωρίζεται σε κελιά μεγέθους 6x6 pixel και για κάθε κελί εξαγάγουμε το αντίστοιχο HOG επιλέγοντας ως αριθμό Bins τα 9. Τέλος, για την εξαγωγή του διανύσματος χαρακτηριστικών μας, επιλέγουμε να αξιοποιήσουμε όλη την πληροφορία του ιστογράμματος κάθε κελιού και έτσι χρησιμοποιούμε όλες τις τιμές του ως χαρακτηριστικά του. Κάθε ιστόγραμμα αποτελείται από τόσες τιμές όσα τα Bins που έχουμε καθορίσει. Άρα στον αλγόριθμο αναφοράς μας επιλέγουμε τις εννέα τιμές του ιστογράμματος κάθε κελιού ως χαρακτηριστικά του. Σημειώνεται, ότι ως ταξινομητή χρησιμοποιούμε μία γραμμική μηχανή SVM, η οποία εκπαιδεύεται με 854 θετικές και 2000 αρνητικές εικόνες, ενώ αξίζει να υπενθυμίσουμε ότι οι εικόνες μας ή μάλλον πιο συγκεκριμένα το παράθυρο ανίχνευσης μας είναι μεγέθους 132x84 pixel.

Αξιολόγηση

Ο ανιχνευτής μας ελέγχθηκε και με χρήση του αλγόριθμου αναφοράς κατάφερε με επιτυχία να ανιχνεύσει την ανθρώπινη παρουσία σε ποσοστό 92,26% ενώ ανίχνευσε λανθασμένα άνθρωπο ενώ δεν υπήρχε σε ποσοστό 5,6%.

5.2.2 Κανονικοποίηση της εικόνας πριν την εφαρμογή του αλγόριθμου

Όσον αφορά πιθανή κανονικοποίηση της εικόνας μας πριν την εφαρμογή του αλγόριθμου, στην υλοποίηση αυτή δοκιμάσαμε μόνο γάμμα κανονικοποίηση. Ξεκινήσαμε, ακολουθώντας τον ίδιο τρόπο γάμμα συμπίεσης όπως στην πρώτη υλοποίηση, δηλαδή με χρήση της τετραγωνικής ρίζας της έντασης και στη συνέχεια δοκιμάσαμε μία διαφορετική γάμμα συμπίεση με χρήση του λογάριθμου.

5.2.2.1 Γάμμα κανονικοποίηση με χρήση της τετραγωνικής ρίζας

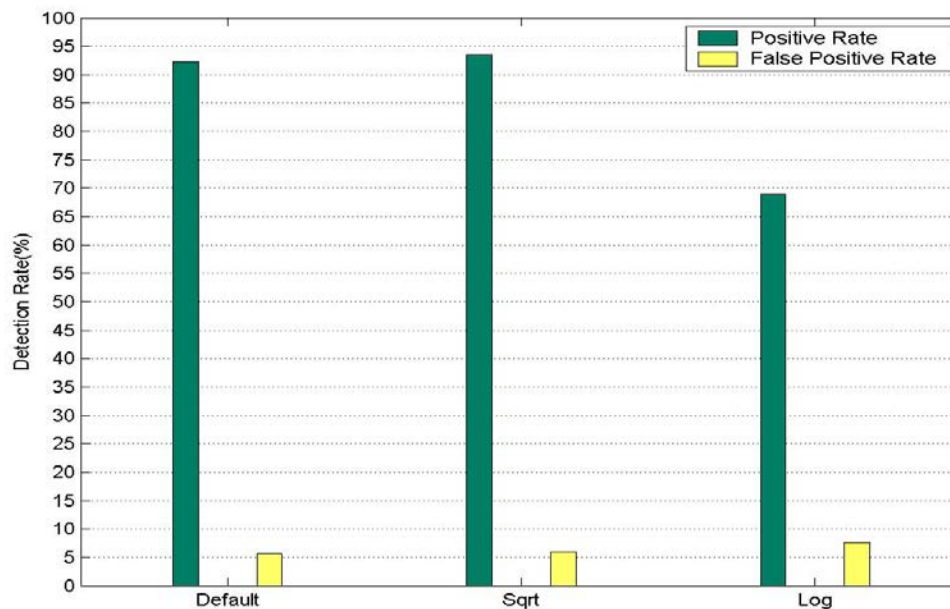
Συγκεκριμένα, επιτύχαμε γάμμα συμπίεση της αρχικής μας εικόνας χρησιμοποιώντας στον ανιχνευτή μας την τετραγωνική ρίζα κάθε χρωματικού καναλιού του μοντέλου RGB. Ο ανιχνευτής μας ανταποκρίθηκε θετικά αυξάνοντας την απόδοση του κατά 1,99%, αλλά παράλληλα και την λανθασμένη αναγνώριση ανθρώπων με το αμελητέο ποσοστό του 0,4%.

5.2.2.2 Γάμμα κανονικοποίηση με χρήση λογάριθμου

Αντίστοιχα, επιτύχαμε γάμμα συμπίεση της αρχικής μας εικόνας χρησιμοποιώντας στον ανιχνευτή μας το λογάριθμο της έντασης κάθε χρωματικού καναλιού του μοντέλου RGB. Η ανταπόκριση του ανιχνευτή μας ήταν αρνητική, καθώς έπεσε στο 69,03%, ενώ η λανθασμένη ανίχνευση αυξήθηκε στο 7,6%.

Αξιολόγηση

Η απόδοση του αλγόριθμου μας με γάμμα κανονικοποίηση της εικόνας φαίνεται διαγραμματικά στο Σχήμα 13. Η γάμμα συμπίεση με χρήση του λογάριθμου είναι ξεκάθαρα πως είναι υπερβολικά ισχυρή και ελαττώνει σημαντικά την απόδοση. Αντίθετα με χρήση της τετραγωνικής ρίζας έχουμε μία αύξηση της απόδοσης που σε συνδυασμό με τη αμελητέα αύξηση της λανθασμένης ανίχνευσης μπορεί να θεωρηθεί αξιόλογη.



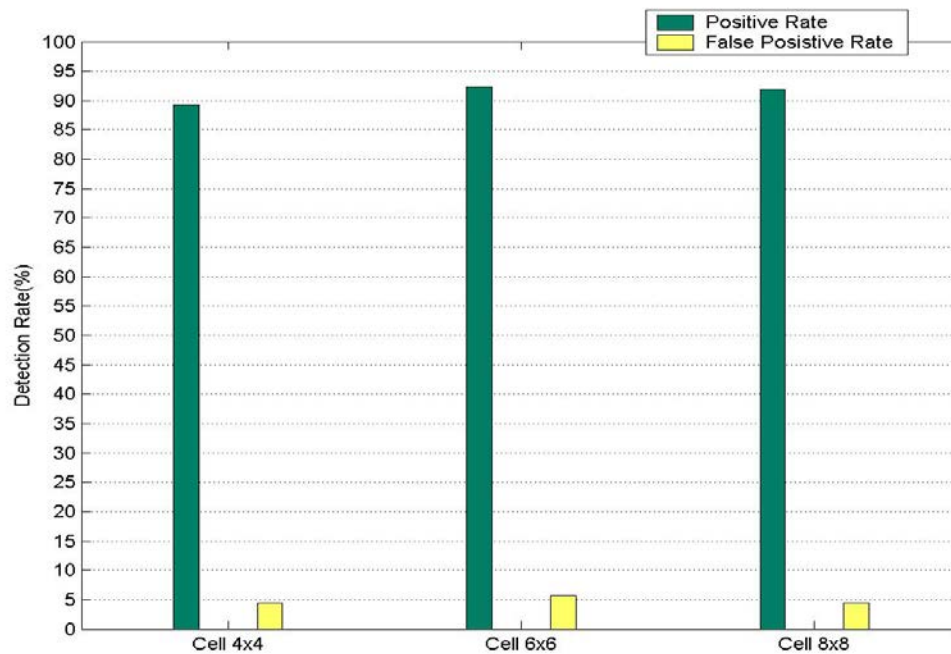
Σχήμα 13. Κανονικοποίηση της εικόνας πριν την εφαρμογή του αλγόριθμου. Με Sqrt συμβολίζεται η γάμμα κανονικοποίηση με χρήση τετραγωνικής ρίζας και με Log η αντίστοιχη με χρήση λογάριθμου

5.2.3 Μεταβολή στο μέγεθος του κελιού

Προσπαθήσαμε να επιβεβαιώσουμε την ορθότητα της επιλογής μας ως προς το μέγεθος του κελιού των 6x6 pixel, μεταβάλλοντας το και σε αυτή την υλοποίηση σε 4x4 και 8x8 pixel. Η συμπεριφορά του αλγόριθμου μας προσδοκούσαμε να είναι ανάλογη με αυτή της πρώτης υλοποίησης και εν μέρει επιβεβαιωθήκαμε. Μειώνοντας το μέγεθος του κελιού μας σε 4x4 pixel η απόδοση μειώθηκε, αλλά ελάχιστα, κατά 2,91%, ενώ αυξάνοντας το σε 8x8, επίσης μειώθηκε, αλλά λιγότερο και συγκεκριμένα κατά 0,32%. Και στις δύο περιπτώσεις η λανθασμένη ανίχνευση ανθρώπων έπεσε στο 4,4%.

Αξιολόγηση

Όπως μπορούμε να παρατηρήσουμε στο Σχήμα 14 και συγκρίνοντας το με το αντίστοιχο της πρώτης υλοποίησης, βλέπουμε πως τα μεγέθη των 6x6 ιδιαίτερως, αλλά και των 8x8 ρικελι υπερέχουν και σε αυτή την υλοποίηση, χωρίς όμως να παρατηρούμε σημαντικές διαφορές στην απόδοσή τους.



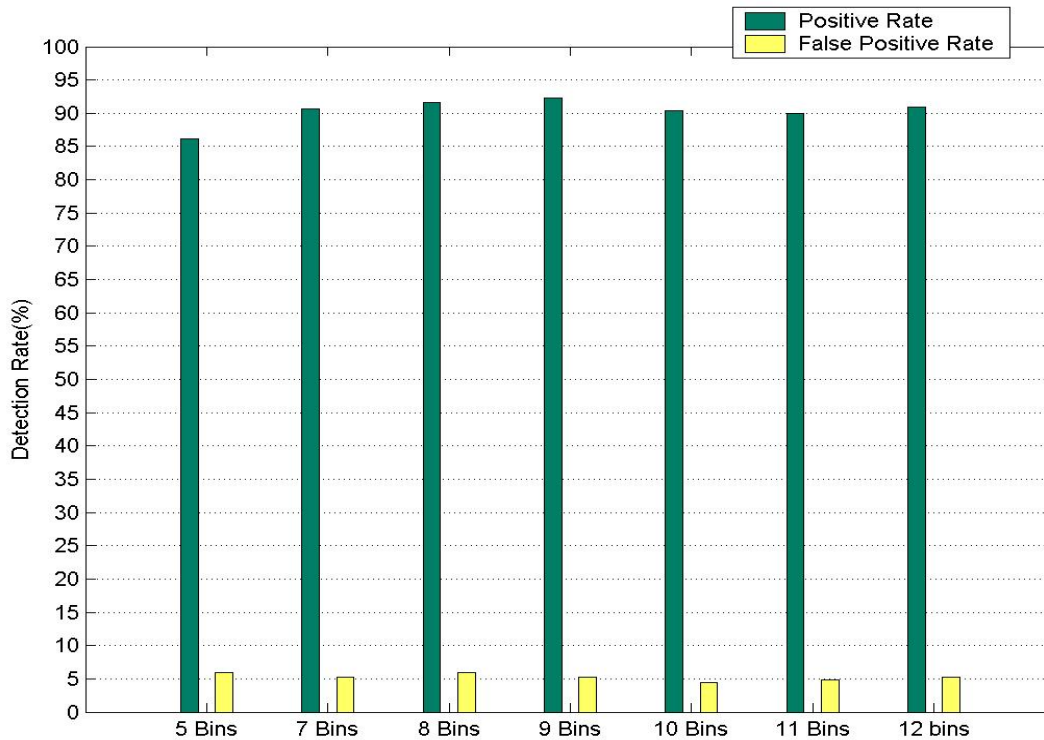
Σχήμα 14. Μεταβολή της απόδοσης του ανιχνευτή μας καθώς αυξάνουμε το μέγεθος του κελιού.

5.2.4 Μεταβολή πλήθους Bins

Όπως έχει γίνει ήδη φανερό το πλήθος των χαρακτηριστικών που εξαγάγουμε από κάθε κελί είναι ισάριθμο με το πλήθος των Bins. Για το λόγο αυτό έχει πολύ μεγάλο ενδιαφέρον να παρατηρήσουμε την απόκριση του αλγόριθμου μας καθώς αυξάνουμε ή μειώνουμε το πλήθος των Bins. Έτσι πέραν των μεταβολών που επιχειρήσαμε στην προηγούμενη υλοποίηση, στην περίπτωση μας δοκιμάσαμε και δύο επιπλέον μεταβολές που θα χαρακτηρίζαμε ως περισσότερο ακραίες. Για να γίνουμε πιο συγκεκριμένοι, μειώσαμε και αυξήσαμε επιπλέον το πλήθος των Bins σε 5 και 13 αντίστοιχα. Παρατηρήσαμε πως η απόδοση του ανιχνευτή μας καθώς αυξάνουμε τον αριθμό των Bins σταδιακά από 5 μέχρι 9 αυξάνεται από το ποσοστό του 86,13 για την περίπτωση των 5 Bins, σε 90,64 στην περίπτωση 7 Bins, σε 91,61 σε περίπτωση 8 Bins, μέχρι να φτάσουμε το ποσοστό του 92,26 για τα 9 Bins. Από εκεί και πέρα, περαιτέρω αύξηση των Bins σταθεροποιεί την απόδοση λίγο πάνω από το 90%. Όσον αφορά τη λανθασμένη ανίχνευση ανθρώπων, κυμαίνεται σε όλες τις περιπτώσεις γύρω από το 5%.

Αξιολόγηση

Όπως πολύ χαρακτηριστικά φαίνεται στο Σχήμα 15 το πλήθος των Bins παίζει πολύ σημαντικό ρόλο. Αύξηση του πλήθους των Bins μέχρι τον αριθμό των 9 αυξάνει σημαντικά την απόδοση του αλγόριθμου, αλλά περαιτέρω αύξηση του πλήθους τους απλά σταθεροποιεί την απόδοση λίγο κάτω από τη βέλτιστη. Σε ό,τι έχει να κάνει με την λανθασμένη ανίχνευση ανθρώπων, ο πειραματισμός μας στο πλήθος των Bins έδειξε να έχει τυχαίες και αμελητέες αυξομειώσεις γύρω από το ποσοστό του 5%.



Σχήμα 15. Απόκριση του ανιχνευτή μας καθώς μεταβάλλουμε το πλήθος των Bins

5.2.5 Κανονικοποίηση σε τοπικό επίπεδο μπλοκ

Δοκιμάσαμε και σε αυτή την υλοποίηση να κανονικοποιήσουμε την πληροφορία μας σε τοπικό εκτός από γενικό επίπεδο με την ελπίδα της βελτίωσης των αποτελεσμάτων. Συγκεκριμένα κάναμε ξανά χρήση της τεχνικής των μπλοκ όπως έχει ήδη περιγραφεί, αλλά αυτή τη φορά πειραματιστήκαμε λίγο με το μέγεθος τους. Σαν μέγεθος κανονικοποίησης χρησιμοποιήσαμε την ίδια L2-νόρμα της προηγούμενης υλοποίησης, την οποία υπενθυμίζουμε στο σημείο αυτό :

$$L2 - norm: v \leftarrow v / \sqrt{\|v\|_2^2 + \varepsilon^2}$$

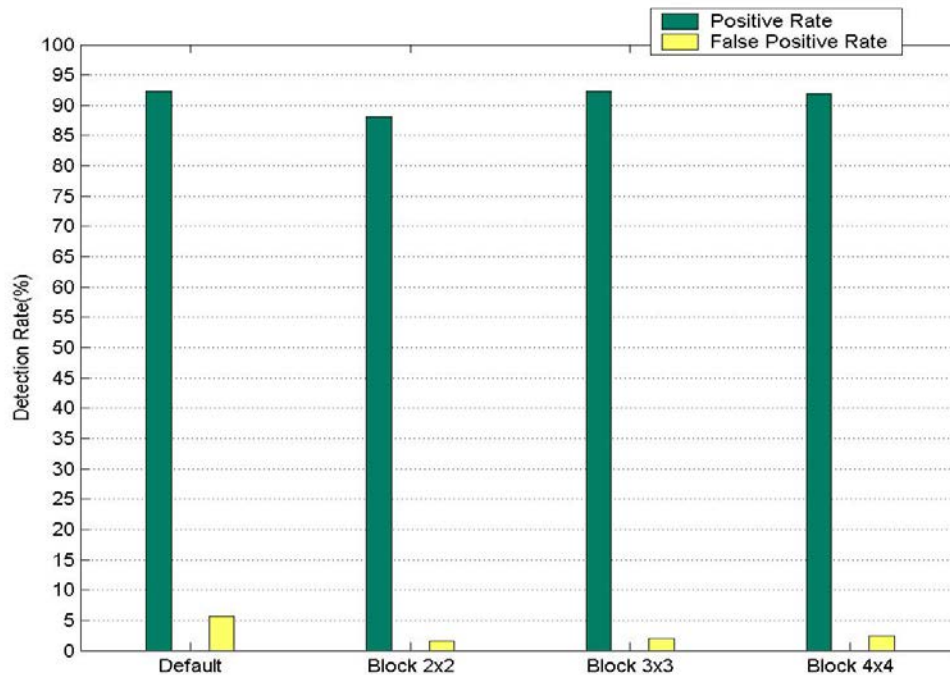
όπου v το μπλοκ πριν την κανονικοποίηση, $\|v\|_k$ η k -νόρμα και ε μία πολύ μικρή σταθερά κανονικοποίησης.

Χρησιμοποιώντας μέγεθος μπλοκ τα 2x2 κελιά παρατηρήσαμε μία μικρή μείωση στην απόδοση του ανιχνευτή μας της τάξης του 4,2%, αλλά μία εντυπωσιακή βελτίωση της λανθασμένης ανίχνευσης σε 1,6%. Με αύξηση του μεγέθους του

μπλοκ σε 3x3 κελιά η απόδοση μας ταυτίστηκε με αυτή του αλγόριθμου αναφοράς διατηρώντας όμως τη λανθασμένη ανίχνευση ανθρώπων χαμηλά στο 2%. Τέλος, αύξηση του μεγέθους του μπλοκ σε 4x4 κελιά μείωσε ξανά την απόδοση κάτω από την αντίστοιχη του αλγόριθμου αναφοράς (91,94%), διατηρώντας όμως το ποσοστό της λανθασμένης ανίχνευσης χαμηλά στο 2,4%.

Αξιολόγηση

Είναι χαρακτηριστική η βελτίωση της απόδοσης του αλγόριθμου μας, όχι τόσο στην επιτυχή ανίχνευση ανθρώπων, αλλά στη λανθασμένη ανίχνευση τους. Μετά την τοπική κανονικοποίηση μας και χρησιμοποιώντας οποιοδήποτε από τα μεγέθη μπλοκ που επιλέξαμε μπορούμε να παρατηρήσουμε στο Σχήμα 16 τη εντυπωσιακή μείωση της λανθασμένης ανίχνευσης ανθρώπων συγκριτικά με τον αλγόριθμο αναφοράς μας.



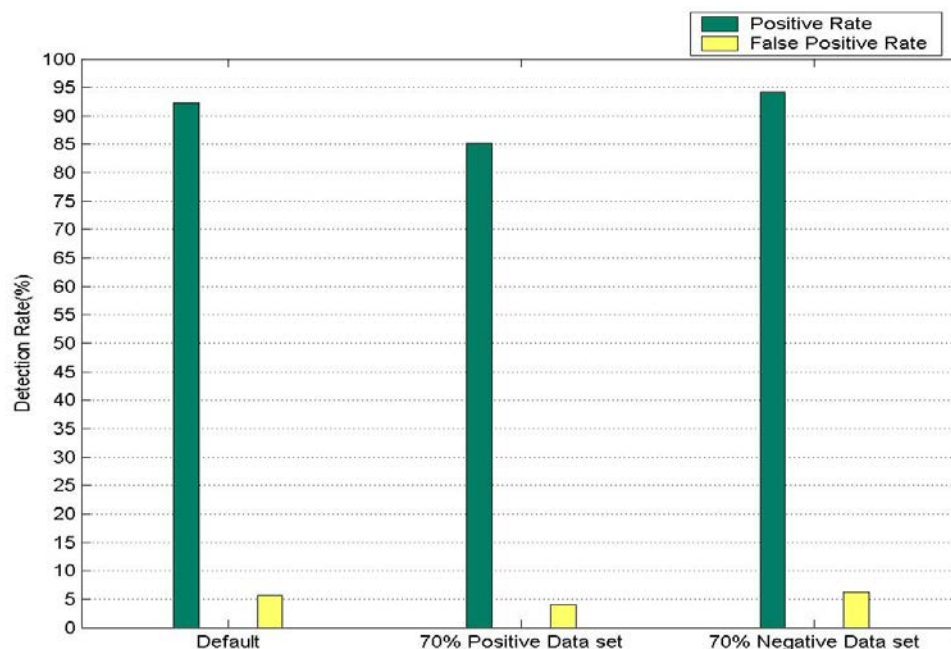
Σχήμα 16. Η κανονικοποίηση σε επίπεδο μπλοκ μειώνει εντυπωσιακά την λανθασμένη ανίχνευση ανθρώπων και διατηρεί την επιτυχημένη ανίχνευση σε υψηλά επίπεδα

5.2.6 Μεταβολή στο πλήθος των εικόνων προς εκπαίδευση

Στο σημείο αυτό μεταβάλαμε το πλήθος των εικόνων μας προς εκπαίδευση ακριβώς με τον ίδιο τρόπο της υλοποίησης με μία επικρατούσα κατεύθυνση. Κρατώντας το 70% των θετικών εικόνων προς εκπαίδευση καταγράψαμε απόδοση 85,16% για επιτυχή και 4% για λανθασμένη ανίχνευση ανθρώπων. Ομοίως, κρατώντας το 70% των αρνητικών εικόνων προς εκπαίδευση καταγράψαμε απόδοση 94,19% για επιτυχή και 6% για λανθασμένη ανίχνευση ανθρώπων.

Αξιολόγηση

Όπως ήταν αναμενόμενο, διαπιστώσαμε πως η μεταβολή της αναλογίας των θετικών και αρνητικών εικόνων προς εκπαίδευση, έχει πανομοιότυπη επίδραση στην απόδοση του ανιχνευτή μας με την αντίστοιχη της υλοποίησης με μία επικρατούσα κατεύθυνση. Τα αποτελέσματα της μεταβολής αυτής παρουσιάζονται γραφικά στο Σχήμα 17.

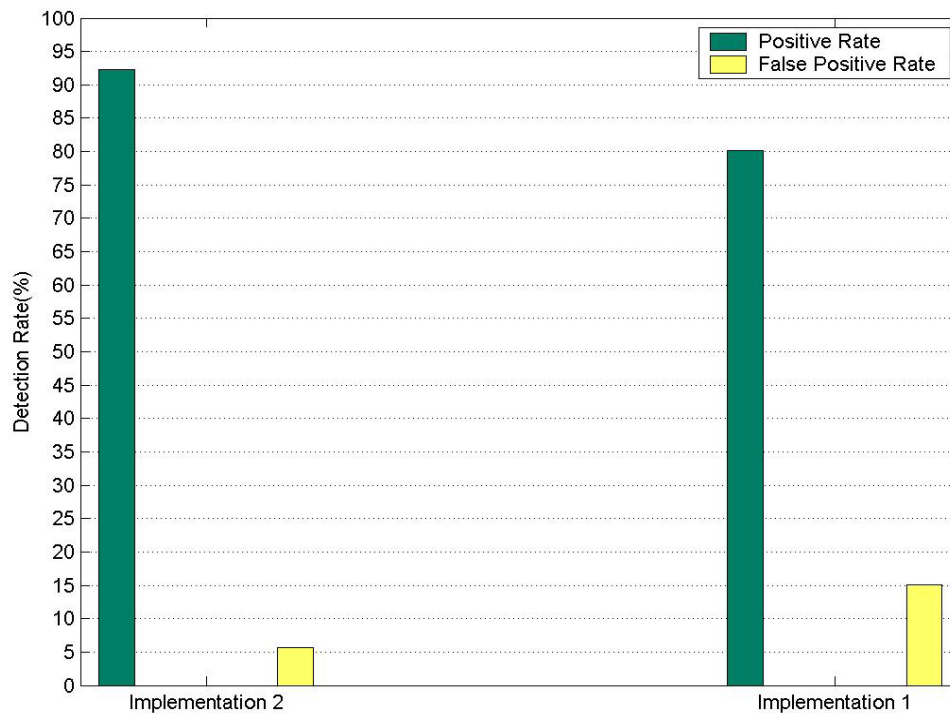


Σχήμα 17. Η αναλογία των θετικών και αρνητικών εικόνων προς εκπαίδευση έχει σημαντική επίδραση στην απόδοση του ανιχνευτή μας.

5.3 Συμπεράσματα πειραμάτων - Σύγκριση υλοποιήσεων

Με βάση τις διαφοροποιήσεις, αλλά και τα αποτελέσματα κάθε υλοποίησης μέσα από τους πειραματισμούς μας, μπορούμε να βγάλουμε κάποια χρήσιμα συμπεράσματα. Στην υλοποίηση με μία επικρατούσα κατεύθυνση αιχμαλωτίζουμε την κυρίαρχη κατεύθυνση κάθε κελιού, αλλά και τη βαρύτητα της, συγκριτικά με τις άλλες κατευθύνσεις του κελιού. Αντίστοιχα, στην υλοποίηση περισσότερων χαρακτηριστικών αιχμαλωτίζουμε την κατανομή όλων των κατευθύνσεων σε κάθε κελί. Είναι λοιπόν εξ αρχής φανερό πως στην πρώτη περίπτωση έχουμε μειωμένο αριθμό χαρακτηριστικών και κατ' επέκταση μείωση της προς επεξεργασία πληροφορίας. Καταλήγουμε δηλαδή, σε ένα περισσότερο απλό σύστημα ταξινόμησης. Αντίθετα στη δεύτερη περίπτωση καταλήγουμε μεν σε ένα περισσότερο πολύπλοκο σύστημα, αλλά διαθέτουμε μέσω της κατανομής των κατευθύνσεων ένα εμπειριστατωμένο, πληρέστερο και σαφώς μεγαλύτερο σύνολο χαρακτηριστικών, που μας "περιγράφει" με πιο ολοκληρωμένο και σφαιρικό τρόπο την τοπική πληροφορία κάθε κελιού. Ήταν λοιπόν απολύτως φυσιολογικό να περιμένουμε καλύτερα αποτελέσματα στην υλοποίηση περισσότερων χαρακτηριστικών. Στο Σχήμα 18 παρουσιάζονται παράλληλα οι γραφικές αναπαραστάσεις της απόδοσης των αλγόριθμων αναφοράς κάθε υλοποίησης. Αν επιχειρούσαμε μία συνολική αξιολόγηση και σύγκριση των δύο υλοποιήσεων θα μπορούσαμε να πούμε πως η υλοποίηση με μία επικρατούσα κατεύθυνση μας δίνει αρκετά ικανοποιητικά αποτελέσματα αλλά όχι στο βαθμό που θα επιθυμούσαμε. Η ανίχνευση ανθρώπων με ποσοστό επιτυχίας λίγο πάνω από 80%, αν και σίγουρα χωρά περιθώρια βελτίωσης, είναι ένα αρκετά καλό δείγμα γραφής. Όσον αφορά όμως τη λανθασμένη ανίχνευση ανθρώπων, μπορούμε να πούμε πως τα ποσοστά που μας δίνει, αρκούν για να χάσει την αξιοπιστία της. Επίσης, παρατηρούμε ότι παρουσιάζει μεγάλη ευαισθησία στις βασικές παραμέτρους της, όπως το μέγεθος των κελιών ή ο αριθμός των Bins, με συνέπεια τις υπερβολικά μεγάλες, σε ορισμένες περιπτώσεις, ποσοστιαίες αποκλίσεις της απόδοσης από

μεταβολή σε μεταβολή. Το φαινόμενο αυτό ίσως να δικαιολογείται από το γεγονός ότι το διάλυμα χαρακτηριστικών μας βασίζεται ουσιαστικά σε ένα μόνο μέγεθος, αυτό της κυρίαρχης γωνίας και όχι σε κατανομή περισσότερων. Από την άλλη μεριά η υλοποίηση περισσότερων χαρακτηριστικών θα λέγαμε ότι προσεγγίζει, στα πλαίσια των απαιτήσεων μας, το ιδανικό. Με πολύ μεγάλο ποσοστό επιτυχούς ανίχνευσης ανθρώπου και πολύ μικρό ποσοστό λανθασμένης, σίγουρα εμπνέει αξιοπιστία. Πειραματιζόμενοι με τις βασικές παραμέτρους της, παρατηρούμε λογικές και σχετικά μικρές μεταβολές στην απόδοση της και γενικότερα ανταποκρίνεται πολύ καλύτερα στις απαιτήσεις μας, αλλά και στις προκλήσεις του θέματος της παρούσας εργασίας.



Σχήμα 18. Συγκριτική παρουσίαση της απόδοσης των αλγόριθμων αναφοράς των δύο υλοποιήσεων.

ΚΕΦΑΛΑΙΟ 6

6 Σύνοψη και προτάσεις επέκτασης

Αυτό το κεφάλαιο συνοψίζει τα σημαντικά εμπειρικά αποτελέσματα και περιγράφει τις περαιτέρω επεκτάσεις και τις μελλοντικές βελτιώσεις στο υπάρχον σύστημα ανίχνευσης ανθρώπων.

6.1 Επέκταση και βελτίωση της μεθόδου

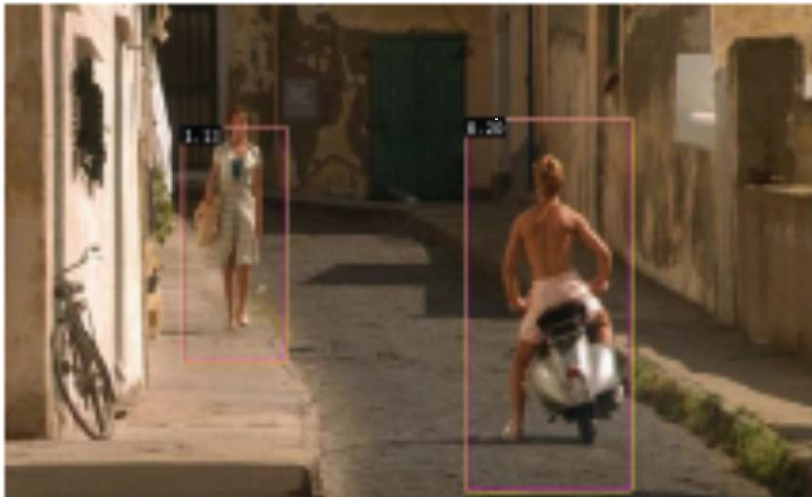
Αν και η μέθοδος μας λειτούργησε πάρα πολύ καλά και αποτελεσματικά στην ανίχνευση ανθρώπων σε όρθια κατά βάση στάση, υπάρχει αναμφίβολα χώρος για βελτίωση. Σίγουρα, θα μπορούσαμε να βελτιώσουμε την ταχύτητα λειτουργίας του αλγόριθμου μας, ώστε να είναι σε θέση να ανταποκρίνεται εξίσου καλά και γρήγορα και σε μεγαλύτερου μεγέθους παράθυρα ανίχνευσης. Επίσης, θα μπορούσαμε να βελτιστοποιήσουμε και την απόδοση αλλά και την ανεξαρτησία του αλγόριθμου μας, σε εμπόδια που κρύβουν μέρος του ανθρώπινου σώματος. Αυτό θα μπορούσε να επιτευχθεί προσαρμόζοντας τη μέθοδο μας σε μία τεχνική που θα βασίζεται στα μέρη που αποτελούν το ανθρώπινο σώμα, όπως κεφάλι, κορμός, χέρια και πόδια. Τέλος, ο ανιχνευτής μας θα μπορούσε να γίνει περισσότερο ισχυρός και αξιόπιστος προσθέτοντας στο διάνυσμα χαρακτηριστικών του επιπλέον χαρακτηριστικά, όπως χαρακτηριστικά που θα σχετίζονται με τη χρωματική πληροφορία των pixels.

Πέραν των βελτιώσεων, θα μπορούσαμε να επεκτείνουμε τη μέθοδο μας προς διάφορες κατευθύνσεις, ώστε να διευρυνθεί το πεδίο εφαρμογής της. Μέχρι τώρα παρουσιάσαμε έναν ανιχνευτή που είναι σε θέση να μας πληροφορήσει αν σε ένα συγκεκριμένου μεγέθους παράθυρο περικλείεται ανθρώπινη μορφή ή όχι. Στο σημείο αυτό θα περιγράψουμε μία πολύ απλή αλλά θεωρητική επέκταση της μεθόδου μας σε μία αντίστοιχη μέθοδο εντοπισμού της ύπαρξης αλλά και της θέσης ενός ανθρώπου σε μία εικόνα οποιουδήποτε μεγέθους. Θεωρητικά αν πραγματοποιήσουμε μία εξαντλητική σάρωση της εικόνας με το

παράθυρο ανίχνευσης μας, μπορούμε να αποφανθούμε με χρήση του ανιχνευτή μας, για κάθε δυνατό παράθυρο, άρα και για κάθε δυνατή θέση στην εικόνα αν υπάρχει άνθρωπος. Έτσι καταγράφοντας τις θέσεις που ο ανιχνευτής μας απάντησε θετικά, έχουμε εντοπίσει τη θέση ή τις θέσεις που βρίσκεται άνθρωπος. Στην πράξη όμως συναντάμε κάποια προβλήματα με βασικότερο και σημαντικότερο αυτό του χρόνου. Αν σκεφτούμε πως οι πιθανές διαφορετικές θέσεις για το παράθυρο μας μέσα στην εικόνα κινούμενοι pixel pixel, ακόμα και για μία εικόνα μικρού μεγέθους είναι υπερβολικά πολλές θα αντιληφθούμε πως για μία εξαντλητική σάρωση απαιτείται ένα τεράστιο χρονικό διάστημα. Αν προσθέσουμε σε αυτό και το χρόνο λειτουργίας του αλγόριθμου μας για κάθε παράθυρο τα πράγματα γίνονται χειρότερα. Στο πρόβλημα του χρόνου προστίθεται και το πρόβλημα του μεγέθους, καθώς το παράθυρο μας ανιχνεύει μονάχα ανθρώπους που “γεμίζουν” τις διαστάσεις του. Στην πραγματικότητα όμως, σε μία εικόνα μπορούμε να έχουμε πολλούς ανθρώπους με τελείως διαφορετικά μεγέθη αναλόγως σε ποιο σημείο του προσκηνίου ή του φόντου βρίσκονται, όπως φαίνεται χαρακτηριστικά στο Σχήμα19. Οι προσεγγίσεις που δοκιμάσαμε ενδεικτικά σε αυτή την αναφορά για να περιορίσουμε το δυνατόν αυτά τα προβλήματα έχουν ως εξής. Όσον αφορά το πρόβλημα του χρόνου βελτιώσαμε κάπως τα πράγματα, όταν εγκαταλείψαμε την ιδέα της εξαντλητικής σάρωσης. Αντί να σαρώσουμε την εικόνα με το παράθυρο μας κινούμενοι pixel-pixel επιλέξαμε βήμα σάρωσης το ένα τέταρτο του παραθύρου για κάθε διάσταση αντίστοιχα. Τώρα, για να ξεπεράσουμε το πρόβλημα του μεγέθους μπορούμε να πραγματοποιήσουμε τη σάρωση μας σε μία πυραμίδα από εικόνες του ίδιου περιεχομένου αλλά διαφορετικής κλίμακας, Με αυτόν τον τρόπο θα επιτύχουμε ανίχνευση σε κάθε ορατού μεγέθους άνθρωπο, αλλά θα σπαταλήσουμε αρκετό χρόνο και θα επανέλθουμε στο αρχικό μας πρόβλημα. Για να κάνουμε την ανίχνευση μας σε μία πυραμίδα εικόνων πιο σύντομη θα μπορούσαμε να αφήσουμε την ιδέα της σάρωσης και να επικεντρωθούμε σε πιθανά σημεία ενδιαφέροντος, σημεία δηλαδή στα οποία αν κεντράρουμε το παράθυρο ανίχνευσης μας είναι περισσότερο πιθανό να υπάρχει άνθρωπος.

Για περισσότερες πληροφορίες σχετικά με τα σημεία ενδιαφέροντος παραπέμπουμε ενδεικτικά στα “keypoints” της μεθόδου SIFT (Lowe [9]).

Περαιτέρω επεκτάσεις της μεθόδου μας θα μπορούσαν να αφορούν ανίχνευση διαφορετικών αντικειμένων εκτός των ανθρώπων. Αντικείμενα όπως αυτοκίνητα, μοτοσικλέτες, κτίρια, αεροπλάνα και διάφορα άλλα αποτελούν αντικείμενα που κεντρίζουν το ενδιαφέρον μας και προσφέρονται για ανίχνευση. Επίσης, μία ενδιαφέρουσα επέκταση θα αφορούσε την εφαρμογή της μεθόδου σε ακολουθία εικόνων. Η συγκεκριμένη αναφορά περιορίστηκε στην ανίχνευση ανθρώπων σε ακίνητες εικόνες, οπότε είναι αναμενόμενο να επιθυμούμε την επέκταση της μεθόδου για ανίχνευση ανθρώπων σε βίντεο.



Σχήμα 19. Άνθρωποι στο προσκήνιο και στο φόντο μίας εικόνας δεν μπορούν να ανιχνευθούν από το ίδιο παράθυρο ανίχνευσης

6.2 Σύνοψη –Γενικά συμπεράσματα

Παρά τα ιδιαίτεως ενθαρρυντικά αποτελέσματα που παρήγαγε το σύστημα αυτό, εύκολα συνειδητοποιεί κανείς, ότι αυτά προήλθαν από εφαρμογή του συστήματος σε αυστηρά προκαθορισμένες συνθήκες και ελεγχόμενες μεταβολές στις παραμέτρους του. Για να χρησιμοποιηθεί το σύστημα σε πρακτικές εφαρμογές απαιτείται ανασχεδιασμός του, ώστε να γίνει πιο ευέλικτο και ανθεκτικό σε μεταβολές του περιβάλλοντος και να μπορεί να εφαρμοστεί σε πραγματικό χρόνο. Παρ' όλα αυτά δείξαμε πως χρησιμοποιώντας πληροφορία από τα HOGs μπορούμε να ανιχνεύσουμε με εντυπωσιακά αποτελεσματικό τρόπο ανθρώπους και να ξεπεράσουμε πολλές από τις συνηθισμένες προκλήσεις σε παρόμοια θέματα.

7 Βιβλιογραφία

[1] D. M. Gavrila and V. Philomin. Real-time object detection for smart vehicles. CVPR, Fort Collins, Colorado, USA, pages 87-93, 1999.

[2] S. Ioffe and D. A. Forsyth. Probabilistic methods for finding people. *IJCV*, 2001.

[3] W. T. Freeman and M. Roth. Orientation histograms for hand gesture recognition. *Intl. Workshop on Automatic Face and Gesture- Recognition, IEEE Computer Society, Zurich, Switzerland*, pages 296-301, June 1995.

[4] A. Mohan, C. Papageorgiou and T. Poggio. Example-based object detection in images by components. *PAMI*, pages: 349-361, April 2001.

[5] N.Sprague and J.Luo. Clothed People Detection in Still Images.Proceedings of the 16th International Conference on Pattern Recognition (ICPR'02) 2002 IEEE

[6] K. Mikolajczyk, C. Schmid, and A. Zisserman. Human detection based on a probabilistic assembly of robust part detectors. *The 8th ECCV, Prague, Czech Republic*, volume I, pages 69-81, 2004.

[7] P. Viola, M. J. Jones, and D. Snow. Detecting pedestrians using patterns of motion and appearance. *The 9th ICCV, Nice, France*, volume 1, pages 734-741, 2003.

[8] R. K. McConnell. Method of and apparatus for pattern recognition, January 1986. U.S. Patent No. 4,567,610.

[9] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91-110, 2004.

[10] Christopher J.C. Burges. A Tutorial on Support Vector Machines for Pattern Recognition, 1998

[11] Navneet Dalal and Bill Triggs. Histograms of Oriented Gradients for Human Detection. Conference on Computer Vision and Pattern Recognition (CVPR), 2005.

[12] Navneet Dalal. Finding People in Images and Videos. PhD Thesis, July 2006

[13] Arthur R. Pope and David G. Lowe, Probabilistic models of appearance for 3-D object recognition, *International Journal of Computer Vision*, 40, 2 (2000), pp. 149-167.

[14] M. Bichsel. Strategies of robust object recognition for the automatic identification of human faces. Ph.d. thesis, ETH Zurich, 1991.