



ΜΗΧΑΝΙΚΗ ΟΡΑΣΗ

Assignment 3: Detection

Μαυρογιώργης Δημήτρης, AM:2016030016
Κολομβάκη Αφροδίτη, AM:2016030158
Δελατόλας Θάνος, AM:2016030074

June 25, 2021

1. Introduction

Σκοπός της τρίτης εργαστηριακής άσκησης είναι να δημιουργήσουμε έναν object detector, ο οποίος βασίζεται στην εξαγωγή χαρακτηριστικών με βάση το Histograms of Gradient Orientations και χρησιμοποιεί τη μέθοδο των sliding windows για να κάνει classification. Πιο συγκεκριμένα, στόχος μας είναι να εξάγουμε κάποια positive και κάποια negative δείγματα από ένα set εικόνων, να τις μετατρέψουμε σε ένα σύνολο χαρακτηριστικών και να τα χρησιμοποιήσουμε για να κάνουμε train το μοντέλο μας. Η εξαγωγή χαρακτηριστικών από τα δείγματα γίνεται με τη αναπαράσταση HOG, με την οποία βγάζουμε ένα feature vector, το οποίο στη συνέχεια θα χρησιμοποιηθεί, για να γίνει το classification.

2. Implementation

1. Image Gradient

Αρχικά, επιλέξαμε ένα φίλτρο Sobel 3x3, για να υπολογίσουμε τη συνέλιξη της εικόνας με το φίλτρο, επειδή το συγκεκριμένο φίλτρο χρησιμοποιείται για τον υπολογισμό μιας προσεγγιστικής τιμής των παραγώγων. Πιο συγκεκριμένα, χρησιμοποιήσαμε το φίλτρο για να βρούμε τις οριζόντιες και κατακόρυφες αλλαγές που υπάρχουν στην εικόνα. Τα φίλτρα που χρησιμοποιήσαμε για να βρούμε τις κατακόρυφες και οριζόντιες διαφορές είναι τα παρακάτω:

$$G_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}, \quad G_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$$

Το φίλτρο G_x το δημιουργήσαμε στη matlab με τη συνάρτηση fspacial και το αποτέλεσμά της το πολλαπλασιάσαμε με -1 έτσι, ώστε να προκύψουν οι αρνητικές τιμές από τα αριστερά και οι θετικές από δεξιά. Επιπλέον, το φίλτρο G_y το υπολογίσαμε παίρνοντας τον ανάστροφο πίνακα του G_x . Στη συγκεκριμένη συνάρτηση που υλοποιήσαμε, υπολογίζουμε τη συνέλιξη των παραπάνω φίλτρων με την εικόνα, ενώ υπολογίζουμε το μαγνητιδε και το οριεντάτιον με βάση τους παρακάτω τύπους

$$\|\nabla f\| = \sqrt{D_x^2 + D_y^2}$$
$$\theta = \tan^{-1} \left(\frac{D_y}{D_x} \right)$$

2. Histograms of Gradient Orientations

Στη συγκεκριμένη συνάρτηση υπολογίσουμε το gradient της εικόνας και στη συνέχεια θέλουμε να βρούμε τα pixel της εικόνας που έχουν magnitude πάνω από το 10% του μέγιστου και το orientation που ανήκει σε κάποιο από τα 9 bins που έχουμε χωρίσει το διάστημα $[-\frac{\pi}{2}, \frac{\pi}{2}]$.

Για την υλοποίηση, δημιουργείται ένας boolean array για κάθε ένα από τα 9 bins. Αν το στοιχείο (i,j) είναι true τότε το pixel (i,j) έχει μια επιμυητή τιμή magnitude και το orientation του είναι μέσα στο διάστημα που ορίζεται από το bin. Τέλος, ανθρίζουμε και τα 9 bins του κάθε pixel και κανονικοποιούμε στο διάστημα [0, 1], για να μην επηρεαζόμαστε από διάφορες αλλαγές στη φωτεινότητα της εικόνας.

3. Detection

Αρχικά, υπολογίζεται το cross-correlation μεταξύ του feature map και του template. Στη συνέχεια, σκοπός είναι να βρούμε τα detection που έχουν το μεγαλύτερο cross-correlation και κατ' επέκταση το μικρότερο sum of square differences. Για την υλοποίηση αυτού ταξινομήσκαν τα cross-correlation με φθίνουσα σειρά και αφαιρέθηκαν τα detection που εχουν overlap πάνω από το 70% του template width.

Ως detection ορίζεται το σημείο (i,j) του πίνακα cross-correlation το οποίο είναι και κέντρο ενός block. Για να μην υπάρχει overlap πάνω από το όριο που αναφέρθηκε παραπάνω, υπολογίζεται η Ευκλείδεια απόσταση μεταξύ του νέου σήμειου και κάθε σήμειου που υπάρχει. Ηδη στη λίστα με τα top detection. Αν η απόσταση δεν ξεπερνά το παραπάνω όριο, τότε μπαίνει στη λίστα με τα top detection και αυξάνονται το συνολικό detection counter. Τέλος, αυτό που κάνουμε σε περίπτωση που δεν βρούμε τον επιθυμητό αριθμό από detections, είναι να επιστρέψουμε μόνο τις τιμές από 1 εως detection counter.

4. Detection Script

Στο συγκεκριμένο script, το οποίο είναι το main program της εργασίας, έχουμε θέσει δύο μεταβλητές με τις οποίες καθορίζουμε το width και height του template (γραμμές 26 και 27 αντίστοιχα). Από τις διαστάσεις του template καθορίζονται και οι διαστάσεις του patch. Στη συνέχεια υπολογίζουμε το αριστερό και δεξιό όριο του template, αλλά και των παραθύρων, τα κέντρα των οποίων υπολογίζουμε με τη συνάρητηση detect που υλοποιήσαμε.

Κατόπιν, η λογική του προγράμματος είναι η εξής:

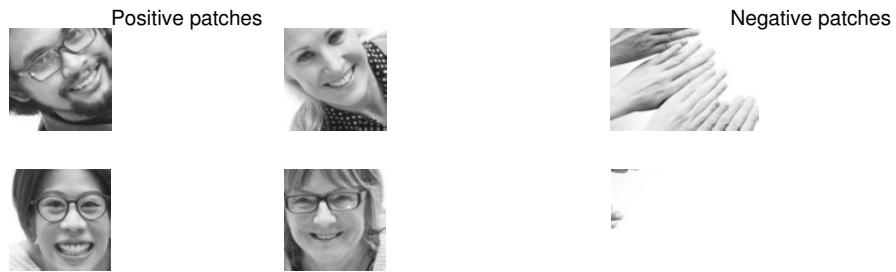
1. Διαβάζουμε μία εικόνα και επιλέγουμε έναν αριθμό από click που θα δώσει ο χρήστης για τα positive templates, τα οποία είναι τα αντικείμενα που θέλουμε να αναγνωρίσουμε.
2. Υπολογίζουμε το HOG της train εικόνας και κατόπιν υπολογίζουμε το μέσο όρο των positive templates.
3. Επαναλαμβάνουμε τα παραπάνω 2 βήματα για να καθορίσουμε τα negative templates, τα οποία είναι αντικείμενα στην εικόνα τα οποία δεν έχουν σχέση με αυτά που θελόμενε να αναγνωρίσουμε. Παραδείγματος χάρη μπορεί να είναι ένα λευκό background.
4. Ως template ορίζουμε τη διαφορά μεταξύ average positive template και negative template.
5. Τέλος, διαβάζουμε και μία άλλη test image και χρησιμοποιούμε το template, για να αναγνωρίσουμε σε αυτή την εικόνα τα αντικείμενα που επιλέξαμε στο πρώτο βήμα του αλγορίθμου.

Αφού μας επιστραφούν τα κέντρα από τη detect ζωγραφίζουμε τα rectangles στην ίδια εικόνα που διαβάσαμε ως test image.

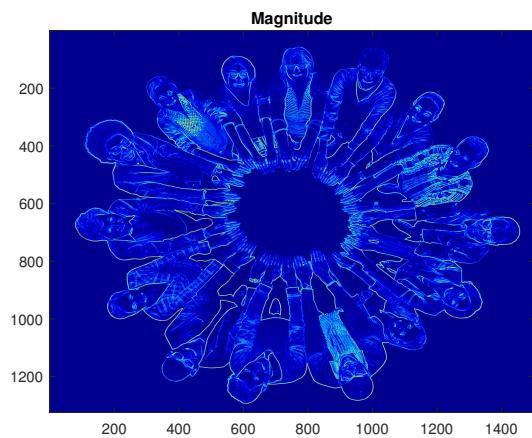
3. Results



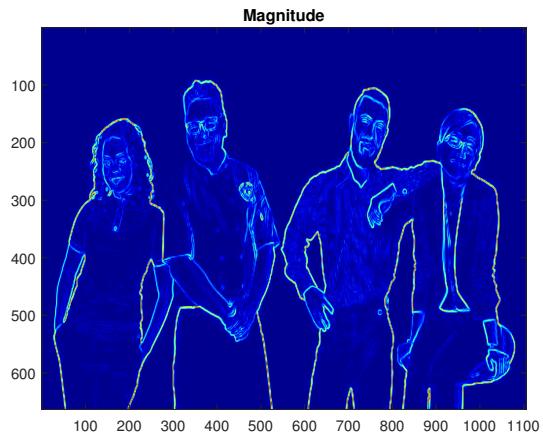
Σχήμα: 1. Input image



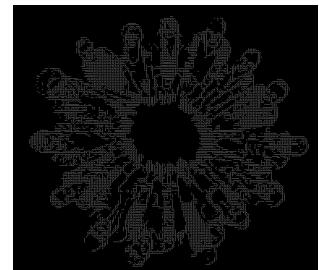
«i» Πρόσωπα προς εντοπισμό.



«iii» Magnitude of training image

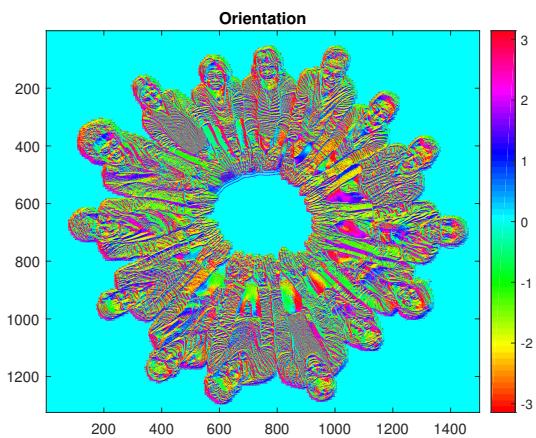


«v» Magnitude of test image

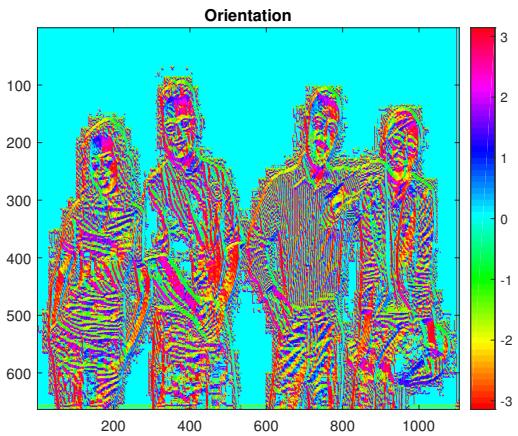


«vii» Hog_draw of training image

«ii» Μη επιθυμητά αντικείμενα.



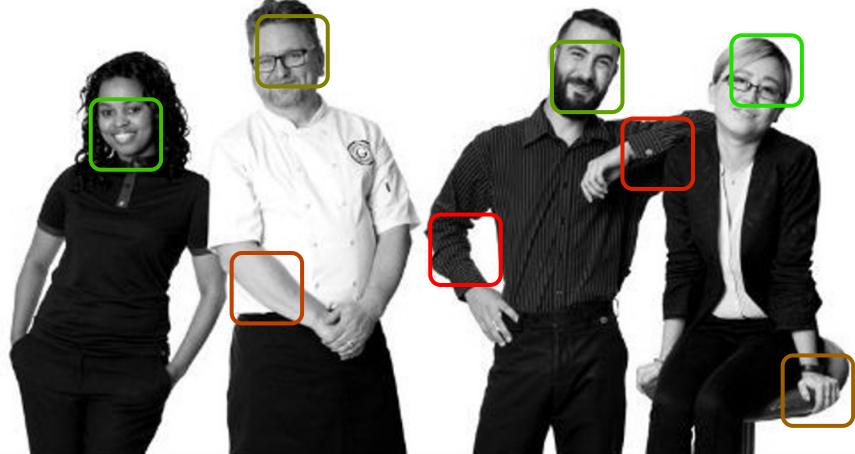
«iv» Orientation of training image



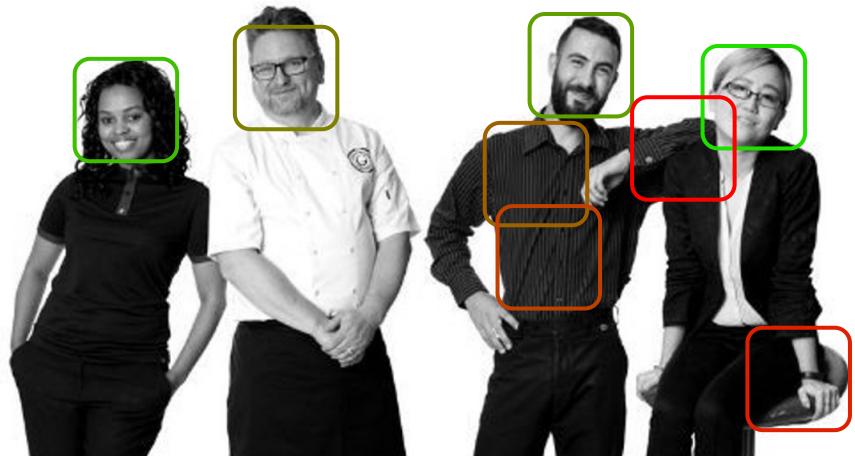
«vi» Orientation of test image



«viii» Hog_draw of test image



Σχήμα: 3. Detection with template 11x11



Σχήμα: 4. Detection with template 16x16



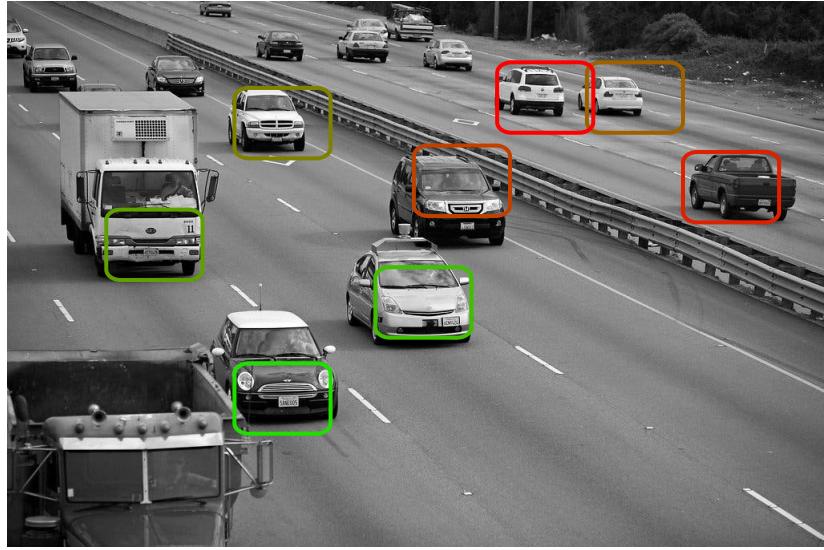
$\Sigma\chi\nu\alpha:$ 5. Input image



«i» Αυτοκίνητα προς εντοπισμό.

«ii» Μη επιθυμητά αντικείμενα.

$\Sigma\chi\nu\alpha:$ 6. Positive and Negative patches for image 5



Σχήμα: 7. Detection with template 15x11

Αρχικά, με βάση τα positive patches τα οποία φαίνονται στο σχήμα 2 (i), γίνεται η αναγνώριση κάποιων άλλων προσώπων στην test image τα οποία είναι διαφορετικά από αυτά που απεικονίζονται ως positive patches. Οι περιοχές-windows οι οποίες έχουν το μεγαλύτερο score (αυτές με πράσινο χρώμα) είναι σε πρόσωπα που έχουν χαρακτηριστικά τα οποία πρώτον είναι κοινά με τα positive patches και δευτερον δεν ανήκουν στο σύνολο των χαρακτηριστικών που δημιουργούν τα negative patches. Παράλληλα, βλέπουμε ότι υπάρχουν και κάποια windows που δεν βρίσκονται σε πρόσωπα και έχουν χαμηλότερο score.

Κατόπιν, παρατηρείται πως ακόμη και τα πιο κοντινά windows απέχουν σημαντικά μεταξύ τους. Αυτό οφείλεται στην συνθήκη του overlap που κληθήκαμε να υλοποιήσουμε και αναλύθηκε στην προηγούμενη ενότητα.

Επιπλέον, στην εικόνα 3 στην οποία το template είναι 11x11, παρατηρείται πως τα windows ο εντοπισμός των προσώπων γίνεται με μεγαλύτερη ακρίβεια σε σχέση με τα αντίστοιχα windows για template 16x16 (όπως φαίνεται στην εικόνα 4). Δηλαδή, αυτό που παρατηρούμε είναι ότι όσο μεγαλώνει το template size (width-height) τα windows συμπεριλαμβάνουν εκτός από τη σημαντική πληροφορία των προσώπων και πληροφορία από το λευκό background της εικόνας.

Τέλος, όπως φαίνεται και από την εικόνα 7, έχοντας ως μεταβλητές τα width, height του template έχουμε τη δημιατότητα να εντοπίσουμε αντικείμενα τα οποία έχουν σχήμα ορθοφωνίου παραλληλογράμου. Με βάση την εικόνα 7 βλέπουμε ότι έχουμε έναν αρκετά ικανοποιητικό εντοπισμό των αυτοκινήτων.