



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

**ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ
ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ
ΥΠΟΛΟΓΙΣΤΩΝ**

Ακαδημαϊκό έτος 2022-2023

ΟΡΑΣΗ ΥΠΟΛΟΓΙΣΤΩΝ

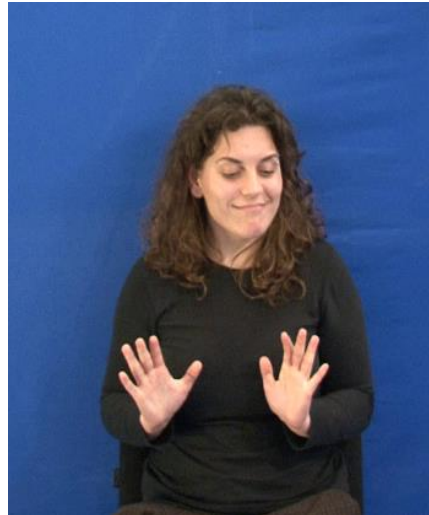
Αναφορά 2^{ης} Εργαστηριακής Άσκησης

Δαμιανός Δημήτρης – 03119825
Καπετανάκης Αναστάσιος – 03119048

Μέρος 1^ο:

1.1 Ανίχνευση Δέρματος Προσώπου και Χεριών

Σε αυτό το ερώτημα καλούμαστε να ανιχνεύσουμε τα σημεία δέρματος της παρακάτω εικόνας:



Για να το πετύχουμε αυτό, χρησιμοποιήσαμε τον χρωματικό χώρο YCbCr και συγκεκριμένα μόνο τις συνιστώσες Cb, Cr που περιγράφουν την ταυτότητα του χρώματος.

Για τον εντοπισμό του δέρματος, χρησιμοποιήσαμε την μοντελοποίηση:

$$P(\mathbf{c} = \text{skin}) = \frac{1}{\sqrt{|\Sigma|} (2\pi)^2} e^{-\frac{1}{2}(\mathbf{c}-\boldsymbol{\mu})\Sigma^{-1}(\mathbf{c}-\boldsymbol{\mu})'}$$

Όπου το διάνυσμα μέσης τιμής $\boldsymbol{\mu}$ και ο πίνακας συνδιακύμανσης Σ υπολογίστηκαν με την βοήθεια των δεδομένων που λάβαμε από το αρχείο skinSamplesRGB.mat.

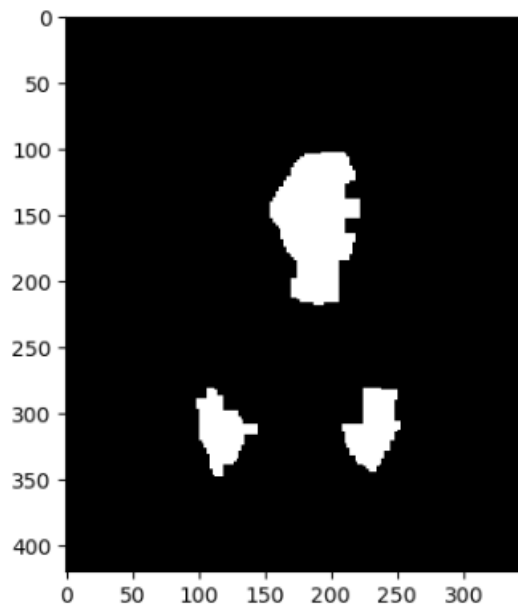
Οι τιμές τους είναι οι εξής:

$$\boldsymbol{\mu} = [157, 103]$$

$$\Sigma = \begin{bmatrix} 44 & -11 \\ -11 & 11 \end{bmatrix}$$

Για τη τελική ανίχνευση, χρησιμοποιήσαμε τα κατώφλια [0.0005, 0.25] (αυτά που μας έδινε η εκφώνηση δεν ήταν αρκετά καλά) ενώ εφαρμόσαμε opening με δομικό στοιχείο έναν 5x5 πίνακα με μονάδες σε κάθε σημείο, και closing με δομικό στοιχείο έναν 12x12 πίνακα με μονάδες σε κάθε σημείο.

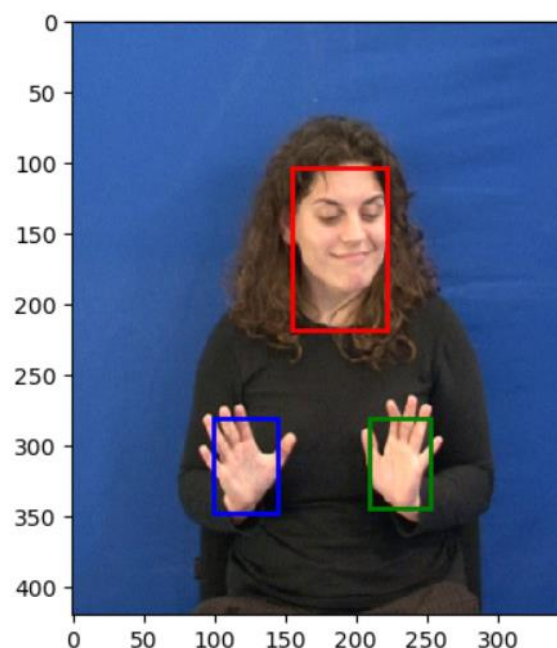
Τα αποτελέσματα της ανίχνευσης είναι τα παρακάτω:



Με την βοήθεια συναρτήσεων της `pytho`, εντοπίσαμε τις 3 διαφορετικές περιοχές της εικόνας και τοποθετήσαμε τα bounding boxes.

Η παραπάνω διαδικασία υλοποιείται μέσω της συνάρτησης **`fd(I, mu, cov)`**, που δέχεται σαν είσοδο την εικόνα, το διάνυσμα μέσων όρων και τον πίνακα συνδιακύμανσης, και επιστρέφει τις συντεταγμένες του bounding box για κάθε εντοπισμένο μέρος της εικόνας.

Παρακάτω τα αποτελέσματα:

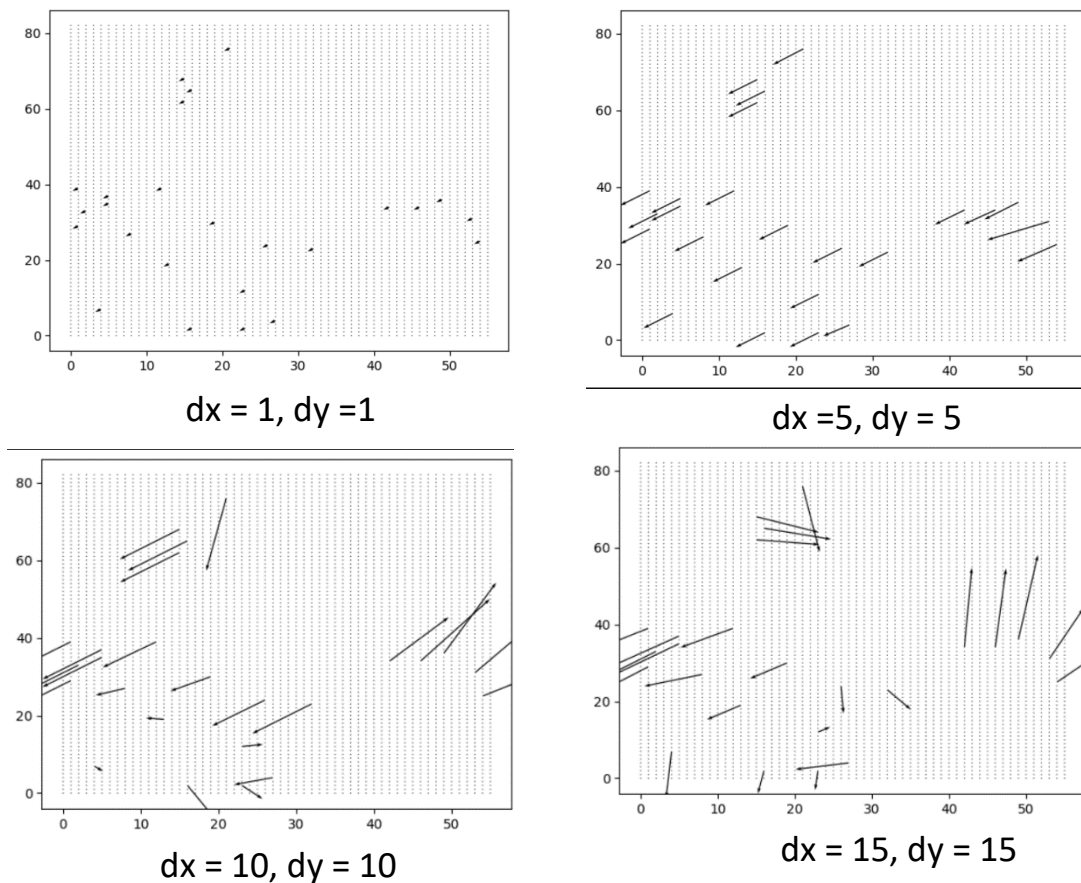


1.2 Παρακολούθηση Προσώπου και Χεριών

Υλοποιήσαμε τον αλγόριθμο Lucas-Kanade ώστε να υπολογίσουμε την οπτική ροή κάποιων σημείων ενδιαφέροντος.

Η συνάρτηση που δημιουργήσαμε, $lk()$, δέχεται ως είσοδο δύο εικόνες-διαδοχικά frames, τα σημεία ενδιαφέροντος, τις παραμέτρους ρ και ϵ καθώς και τις αρχικές εκτιμήσεις μας για την οπτική ροή. Η συνάρτηση εφαρμόζει τον αλγόριθμο Lucas-Kanade πάνω στα σημεία ενδιαφέροντος, και τερματίζει μετά από 100 επαναλήψεις.

Δοκιμάσαμε την $lk()$ για διάφορες μετατοπίσεις μεταξύ των δύο διαδοχικών εικόνων, ώστε να εξετάσουμε τα όρια του και την ικανότητα του να υπολογίζει καλά την οπτική ροή. Παρακάτω τα αποτελέσματα:



Παρατηρούμε ότι για μικρές μετατοπίσεις μεταξύ των εικόνων, η οπτική ροή των σημείων υπολογίζεται αρκετά καλά, ενώ για μεγάλες μετατοπίσεις ο αλγόριθμος δεν λειτουργεί καλά.

Για τον υπολογισμό της μετατόπισης του bounding box, έχουμε την $\text{displ}(dx,dy)$, η οποία επιστρέφει τον μέσο όρο των οπτικών ροών των σημείων που έχουν ενέργεια μεγαλύτερη από ένα κατώφλι.

Με την βοήθεια της, καταφέρνουμε να κάνουμε track το δεξί και αριστερό χέρι της κοπέλας, καθώς και το πρόσωπό της, κατά τη διάρκεια της κίνησής της.

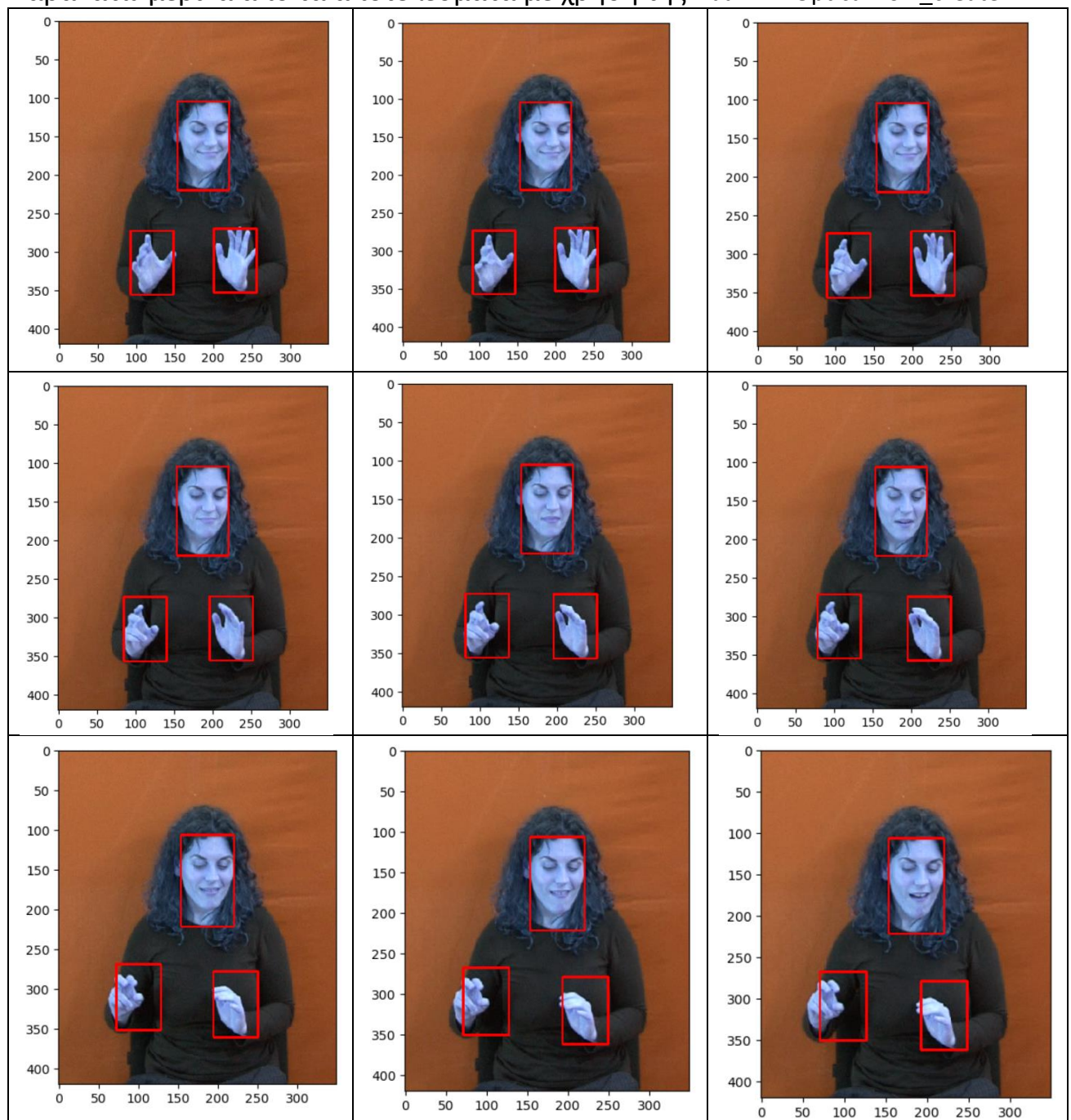
Παρακάτω μερικά από τα αποτελέσματά μας, με χρήση της δικής υλοποίησης του Lucas-Kanade:



Παρατηρούμε ότι στα δύο χέρια, τα οποία αλλάζουν την μορφή τους κατά την διάρκεια της κίνησης, δυσκολευόμαστε να κάνουμε πολύ καλό track. Το bounding box σε αυτές τις περιπτώσεις “χάνει” την κίνηση, και αυτό μάλλον οφείλεται στα διαφορετικά σημεία ενδιαφέροντος που εντοπίζονται μεταξύ των frames της έντονη αλλαγής.

Αντίθετα, στο πρόσωπο, που δεν έχει κάποια έντονη και απότομη αλλαγή, κάνουμε πολύ καλό track.

Παρακάτω μερικά από τα αποτελέσματα με χρήση της DualTVL1OpticalFlow_create:



Σε αυτό το σημείο θα εξετάσουμε τον πολυκλιμακωτό Lucas-Kanade.

Η υλοποίηση του γίνεται ως εξής:

Εφαρμόζουμε ένα 2D γκαουσιανό φίλτρο στην εικόνα (ώστε να μειώσουμε το aliasing), και εντοπίζουμε τα σημεία ενδιαφέροντος.

Έπειτα, μειώνουμε την κλίμακα της εικόνας και εφαρμόζουμε τον μονοκλιμακωτό Lucas-Kanade.

Αυτό επαναλαμβάνεται για όλες τις κλίμακες θέλουμε (εδώ 3).

Παρακάτω τα αποτελέσματα του tracking με χρήση του πολυκλιμακωτού Lucas-Kanade:



Παρατηρούμε ότι στον πολυκλιμακωτό Lucas-Kanade έχουμε πιο αργή σύγκλιση της οπτικής ροής. Παράλληλα, ενώ έχουμε καλύτερα αποτελέσματα στα bounding boxes για το αριστερό χέρι, για το δεξί χέρι τα αποτελέσματα είναι λίγο χειρότερα. Αυτό πιθανότατα θα μπορούσε να λυθεί με fine tuning των παραμέτρων του Lucas-Kanade, καθώς και των παραμέτρων της `displ()`.

Τέλος, στο `lk()` πειραματιστήκαμε και με άλλο κριτήριο σύγκλισης. Πέρα από τις 100 επαναλήψεις, θέσαμε σαν κριτήριο σύγκλισης η διαφορά μεταξύ δυο διαδοχικών `dx` και `dy` να είναι μικρότερη κάποιου threshold.

Για μικρές μετατοπίσεις και αλλαγές μεταξύ των frames, είχαμε γρήγορη σύγκλιση. Για μεγάλες μετατοπίσεις και αλλαγές όμως ο αλγόριθμος αργεί πολύ να συγκλίνει.

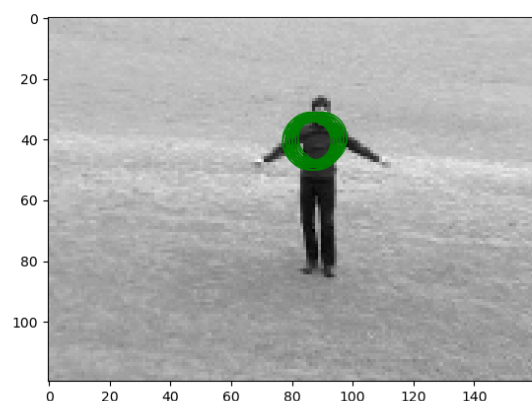
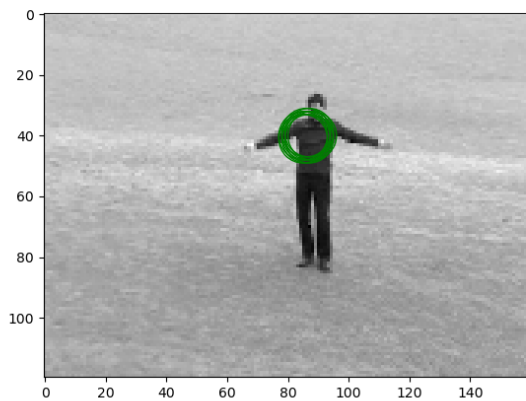
Μέρος 2^ο :

2.1 Χωρο-χρονικά Σημεία Ενδιαφέροντος

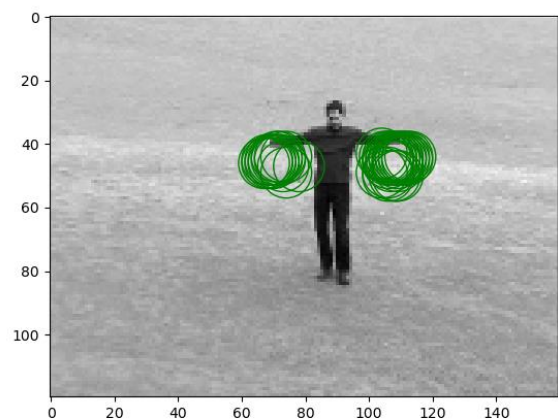
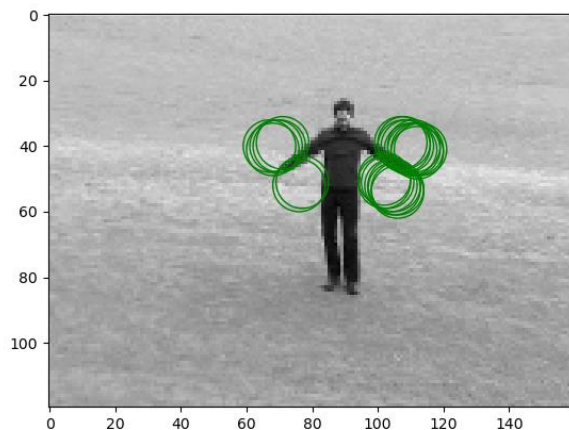
Για την εύρεση σημείων ενδιαφέροντος στο χρόνο και στο χώρο, υλοποιήσαμε το HarrisDetector για 3 διαστάσεις, καθώς και τον Gabor detector, όπως περιγράφεται στην εκφώνηση.

Τα αποτελέσματα στα frames του handwaving βίντεο είναι τα παρακάτω.

Με την χρήση του HarrisDetector:

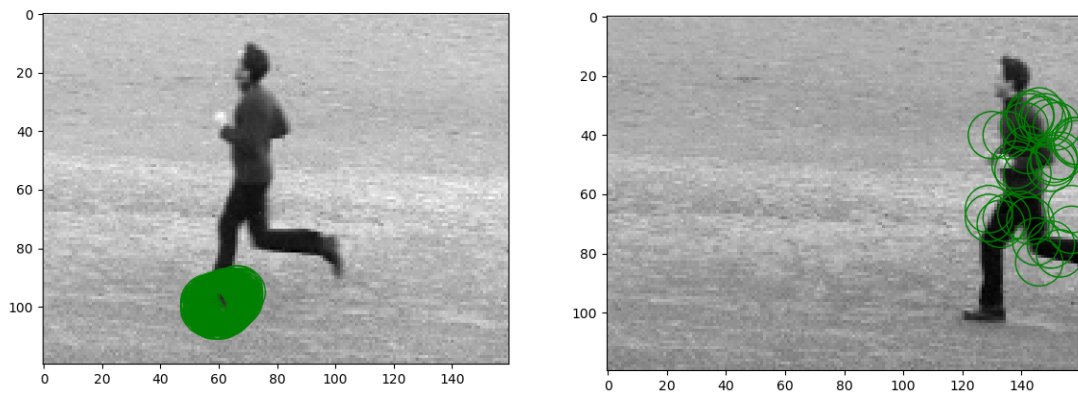


Με την χρήση του Gabor:

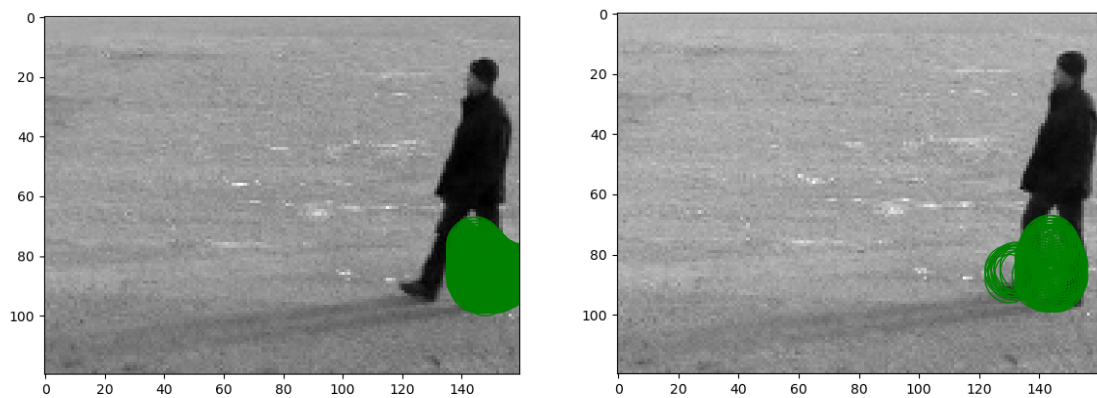


Παρατηρούμε ότι ο HarrisDetector εντοπίζει σημεία ενδιαφέροντος στο σώμα ενώ αντίθετα ο Gabor εντοπίζει σημεία στα χέρια. Συνεπώς, ο Gabor θα είναι πιο χρήσιμος για τον εντοπισμό κίνησης σε βίντεο.

Παρακάτω τα αποτελέσματα για τα άλλα δύο βίντεο:
Για τα running, harrisDetector και Gabor αντίστοιχα:



Για τα walking, harrisDetector και Gabor αντίστοιχα:



Παρατηρούμε ότι πάλι ο HarrisDetector εντοπίζει σημεία πάνω στο σώμα, ενώ ο Gabor βρίσκει περισσότερα σημεία.

2.2 Χωρο-χρονικοί Ιστογραφικοί Περιγραφητές

Σε αυτό το σημείο θα εξαγάγουμε χωρο-χρονικούς περιγραφητές με δύο διαφορετικούς τρόπους: σύμφωνα με την κατευθυντική παράγωγο (HOG) και σύμφωνα με την οπτική ροή (HOF).

Για γίνει αυτό, μετά τον υπολογισμό της παραγώγου και της οπτικής ροής, εξαγάγουμε τετραγωνικές περιοχές πλευράς σ (η σ που χρησιμοποιήθηκε για την εξαγωγή σημείων ενδιαφέροντος) τόσο για τον x άξονα όσο και για τον y , στο οποίο θα υπολογίσουμε τους ιστογραφικούς περιγραφητές.

2.3 Κατασκευή BoVW και χρήση SVM για ταξινόμηση δράσεων

Μετά την εξαγωγή των περιγραφητών (στα βίντεο εκπαίδευσης και επαλήθευσης που μας δόθηκαν) χρησιμοποιούμε τις συναρτήσεις που μας δόθηκαν, `bag_of_words()` και `svm_train_test()`, για να υπολογίσουμε το accuracy της εκπαίδευσης μας.

Δοκιμάζοντας διάφορους συνδυασμούς εξαγωγής χαρακτηριστικών (Harris, Gabor) και περιγραφητών (HOG, HOF, HOG/HOF), είχαμε μερικά αποτελέσματα όπως τα παρακάτω:

```
100%|████████████████████████████████████████████████████████████████████████████████| 36/36 [09:29<00:00, 15.82s/it]
100%|████████████████████████████████████████████████████████████████████████████████| 12/12 [03:02<00:00, 15.24s/it]

Using Harris and HOG, acc:0.3333333333333333

100%|████████████████████████████████████████████████████████████████████████████████| 36/36 [09:35<00:00, 16.00s/it]
100%|████████████████████████████████████████████████████████████████████████████████| 12/12 [02:50<00:00, 14.23s/it]

Using Gabor and HOG, acc: 0.25

100%|████████████████████████████████████████████████████████████████████████████████| 36/36 [2:10:21<00:00, 217.25s/it]
100%|████████████████████████████████████████████████████████████████████████████████| 12/12 [40:21<00:00, 201.82s/it]

Using Harris and HOF, acc:0.5833333333333334
```

Τα καλύτερα αποτελέσματα που λάβαμε ήταν για τον συνδυασμό Gabor-HOF, με accuracy 0.91. Αυτό που παρατηρήσαμε είναι ότι για χαρακτηριστικά από Gabor και HOF περιγραφητές είχαμε καλύτερα αποτελέσματα.

Μέρος 3^ο:

Βήμα 1:

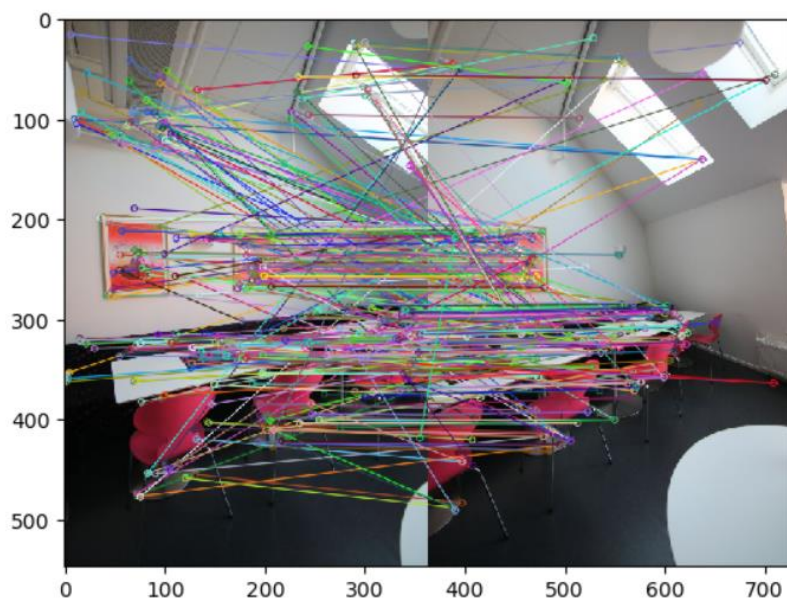
Σε αυτό το βήμα εξάγουμε τα σημεία ενδιαφέροντος από τις 2 εικόνες, με την χρήση της `SIFT_create()`.

Τα σημεία που εντοπίσαμε για τις 2 πρώτες εικόνες είναι τα εξής:



Βήμα 2:

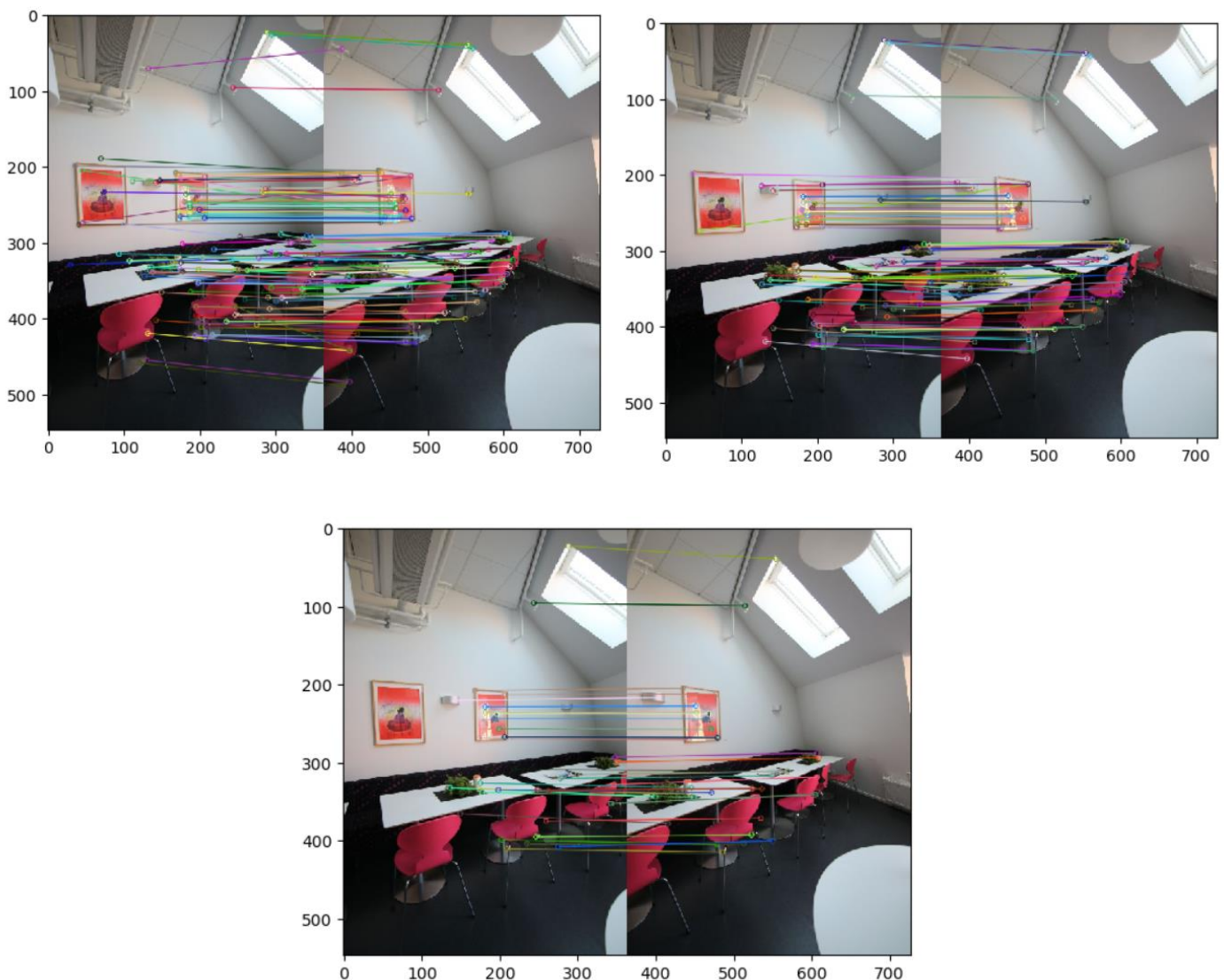
Σε αυτό το βήμα, θα ταιριάζουμε τα σημεία που εντοπίσαμε στο προηγούμενο βήμα μεταξύ των δύο εικόνων (όλα τα σημεία μεταξύ τους):



Βήμα 3:

Παρατηρούμε ότι το ταίριασμα όλων των σημείων με όλα οδηγεί σε λάθος ταιριάσματα. Παράλληλα, για να κάνουμε καλό matching, θα πρέπει να επιλέξουμε τα πιο «κοντινά» σημεία, ώστε, μετά την παραμόρφωση που θα δεχτούν οι εικόνες να υπάρχουν ακόμα κοινά σημεία ώστε να γίνει το stitching. Για να το πετύχουμε αυτό, εφαρμόζουμε το κριτήριο Lowe για διάφορες τιμές ratio.

Παρακάτω τα αποτελέσματα του matching για ratio 0.8, 0.5 και 0.3:



Στην δική μας υλοποίηση χρησιμοποιήσαμε Lowe's ratio = 0.5.

Βήμα 4:

Σε αυτό το βήμα , με την βοήθεια της RANSAC , θα υπολογίσουμε το πίνακα ομογραφίας H . Ο πίνακας H μας είναι χρήσιμος, γιατί μας δείχνει πως απεικονίζονται τα σημεία της 1^{ης} εικόνας στον σύστημα συντεταγμένων της 2^{ης}.

Βήμα 5:

Σε αυτό το βήμα θα πρέπει να μετασχηματίσουμε την 1^η εικόνα σύμφωνα με τον πίνακα ομογραφίας H , ώστε να απεικονιστεί στο χώρο της 2^{ης} εικόνας.

Μετά τον υπολογισμό των νέων 4 ακριανών σημείων της εικόνας (με την βοήθεια της H), θα πρέπει να υπολογίσουμε τον μετασχηματισμό, που θα μεταφέρει την 1^η εικόνα από τα αρχικά ακριανά σημεία στα νέα.

Ο νέος μετασχηματισμός M θα πρέπει από την μία να λάβει υπόψη τον πίνακα ομογραφίας αλλά και την μετατόπιση που μπορεί να πρέπει να εφαρμοστεί ώστε η μετασχηματισμένη 1^η εικόνα να «ευθυγραμμιστεί» με την 2^η.

Για αυτό, έχουμε τον πίνακα μετατόπισης

$$T = \begin{bmatrix} 1 & 0 & -A[0] \\ 0 & 1 & -A[1] \\ 0 & 0 & 1 \end{bmatrix}$$

Τα $A[0], A[1]$ είναι οι (x,y) συντεταγμένες του πάνω αριστερά γωνιακού σημείου της 1^η εικόνας. Με βάση την T , η 1^η εικόνας μετατοπίζεται κατάλληλα.

Συνεπώς, ο μετασχηματισμός M είναι $M = T * H$ (ώστε να εφαρμόσουμε και τους 2 μετασχηματισμούς).

Το αποτέλεσμα μετασχηματισμού της 1^{ης} εικόνας είναι το παρακάτω:



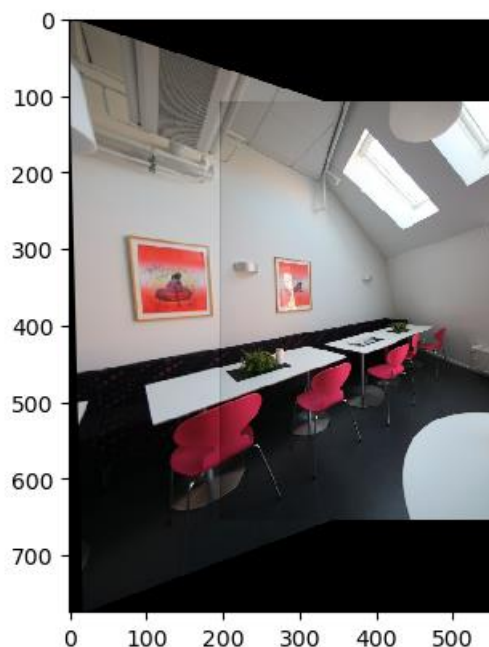
Βήμα 6:

Σε αυτό το βήμα θα συνενώσουμε τις 2 εικόνες, με την βοήθεια των προηγούμενων βημάτων.

Συγκεκριμένα, μετατοπίζουμε την 2^η εικόνα σύμφωνα με τον ίδιο πίνακα T του βήματος 5 (ώστε να υπάρξει ευθυγράμμιση των δύο εικόνων) και έπειτα υπολογίζουμε το μέγεθος της εικόνας που θα παραχθεί από την συνένωση.

Μεταφέρουμε τις εικόνες στο νέο μέγεθος, και τις ενώνουμε (μηδενίζοντας τα pixels της μίας εκεί που υπάρχει επικάλυψη).

Το αποτέλεσμα της συνένωσης των δύο πρώτων εικόνων είναι το παρακάτω:



Τελική Συνένωση:

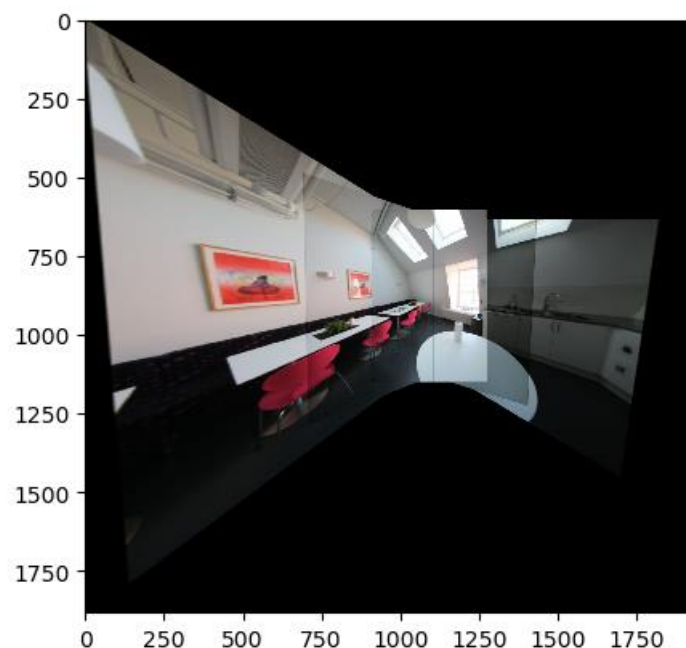
Για την τελική συνένωση των εικόνων, είναι σημαντική η σειρά με την οποία θα γίνει η συνένωση, ώστε να έχουμε καλό αποτέλεσμα.

Μετά από κάποιους πειραματισμούς, κάναμε τις συνενώσεις με την παρακάτω σειρά:

Εικόνες 1-2-3, Εικόνες 5-4-3 και τελικά τις δύο εικόνες μεταξύ τους.

Ο λόγος αυτής τη ακολουθίας είναι ότι και στις δύο παραγόμενες εικόνες έχουν την εικόνα 3 ως «κοινή» και μη παραμορφωμένη. Έτσι, όταν εφαρμοστεί η παραπάνω διαδικασία, οι δύο εικόνες θα έχουν κοινά σημεία τα σημεία της 3.

Το αποτέλεσμα είναι το παρακάτω:



Παρατηρούμε ότι το αποτέλεσμα είναι αρκετά ικανοποιητικό. Μια βελτίωση του αποτελέσματος θα ήταν να βελτιώσουμε την φωτεινότητα του τελικού αποτελέσματος, ώστε να μην έχουμε έντονες διαφορές μεταξύ των εικόνων.