2ο Εργαστήριο Όραση Υπολογιστών - Ακ. Έτος 2020-2021

Εκτίμηση Οπτικής Ροής (Optical Flow)
Εξαγωγή Χαρακτηριστικών σε Βίντεο
Αναγνώριση Δράσεων

Τσακανίκα Χριστίνα ΑΜ: 03117012
Παυλάκη Παρασκευή-Ευγενία ΑΜ: 03117190

Μέρος 1: Παρακολούθηση Προσώπου και Χεριών με Χρήση της Μεθόδου Οπτικής Ροής των Lucas-Kanade

1.1 Ανίχνευση Δέρματος Προσώπου και Χεριών

Στο πρώτο ερώτημα μελετήθηκε η ανίχνευση σημείων δέρματος στο πρώτο πλαίσιο της ακολουθίας και η τελική επιλογή της περιοχής του προσώπου και των χεριών. Για την ανίχνευση των σημείων δέρματος χρησιμοποιείται ο χρωματικός χώρος YCbCr, αφαιρώντας την πληροφορία της φωτεινότητας Y και διατηρώντας τα κανάλια Cb και Cr που περιγράφουν την ταυτότητα του χρώματος.

Η μοντελοποίηση του χρώματος του δέρματος με δισδιάστατη Γκαουσιανή κατανομή αποδίδεται παρακάτω:

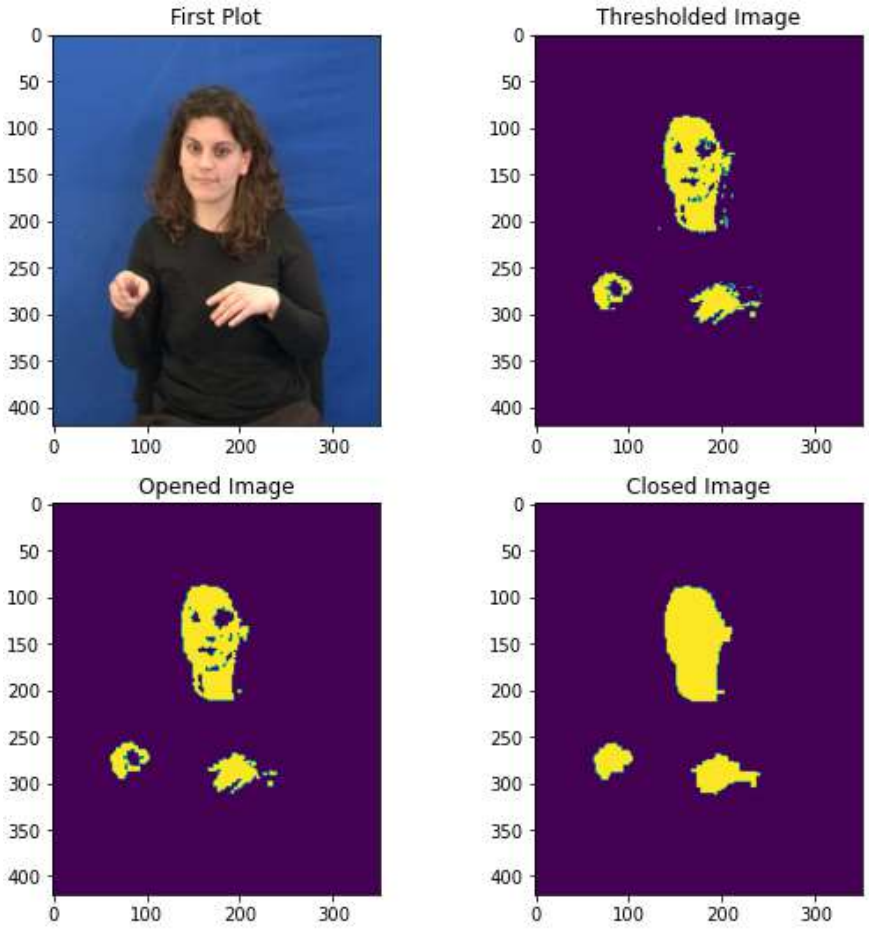
$$P(c = skin) = \frac{e^{-\frac{1}{2}(c-\mu)\sum^{-1}(c-\mu)'}}{|\sum(2\pi)^2|^{1/2}}$$

Στην συνέχεια, ορίσαμε threshold και υπολογίζουμε με κατωφλιοποίηση την δυαδική εικόνα ανίχνευσης δέρματος. Η μέση τιμή καθώς και η συνδιακύμανση της Γκαουσιανής κατανομής, υπολογίζονται από το δεδομένο αρχείο *matlab*, το οποίο περιέχει *RGB* δείγματα ανθρώπινου δέρματος.

Στόχος ήταν η κατηγοριοποίηση των χαρακτηριστικών του δέρματος σε τρεις κατηγορίες: πρόσωπο, δεξί χέρι και αριστερό χέρι.

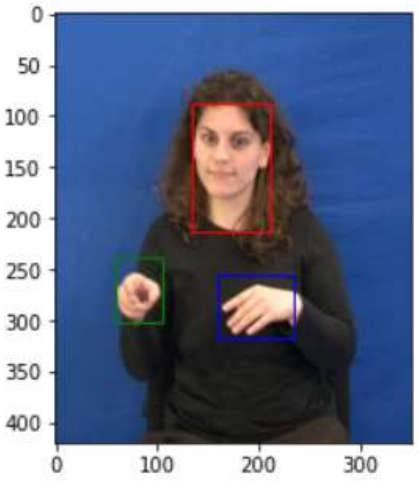
Γι αυτό χρειάστηκε να καλύψουμε τις οπές που εμφανίζονται, εφαρμόζοντας opening με ένα πολύ μικρό δομικό στοιχείο - δίσκο (3x3) και closing με ένα μεγαλύτερο δομικό στοιχείο (18x18) στη δυαδική εικόνα. Τα νούμερα προέκυψαν πειραματικά.

Παρακάτω φαίνεται αναλυτικά σε βήματα η διαδικασία που περιγράψαμε:



Χρησιμοποιώντας τη συνάρτηση *patches.Rectangle* σχηματίσαμε το ορθογώνιο πλαίσιο γύρω από τις περιοχές ενδιαφέροντος στην δυαδική εικόνα.

Συνεπώς, το αποτέλεσμα επί της αρχικής εικόνας θα είναι το παρακάτω:



Η εφαρμογή των bounding boxes σε κάθε frame καθώς και η εξαγωγή της εικόνας κατωφλιοποίησης, της ανοιγμένης και της κλειστής εικόνας (thresholded image, opened image, closed image) επιτεύχθηκε με τη συνάρτηση *fd*. Η προαναφερθείσα συνάρτηση δέχεται ως εισόδους μια εικόνα (στην προκειμένη περίπτωση, την πρώτη της ακολουθίας βίντεο), τη μέση τιμή, *mean*, και τη συνδιακύμανση, *cov*, της Γκαουσιανής κατανομής, ενώ επιστρέφει το πλαίσιο οριοθέτησης της περιοχής ενδιαφέροντος στη μορφή [x, y, width, height], όπου x, y οι συντεταγμένες του πάνω αριστερά σημείου.

1.2 Παρακολούθηση Προσώπου και Χεριών

Η αρχικοποίηση των Bounding Boxes που περιλαμβάνει το πρόσωπο και τα χέρια της νοηματίστριας στη μορφή [x, y, width, height] επιλέχθηκε να είναι η παρακάτω:

- Πρόσωπο: [134, 88, 77, 126]
- Αριστερό χέρι: [58, 239, 47, 64]
- Δεξί χέρι: [159, 256, 75, 62]

ό,τι επέστρεψε δηλαδή η συνάρτηση *fd* του μέρους 1.1.

1.2.1 Υλοποίηση του Αλγόριθμου των Lucas-Kanade

Ο αλγόριθμος αποτελεί αυτόνομη συνάρτηση, που να δέχεται ως εισόδους δύο εικόνες κομμένα παράθυρα με βάση το bounding box του ερωτήματος 1.1 από δύο διαδοχικά πλαίσια του βίντεο. Οι δύο αυτές διαδοχικές εικόνες απεικονίζουν είτε τα χέρια είτε το κεφάλι της νοηματίστριας. Το σύνολο των σημείων ενδιαφέροντος (*features*) εντός του παραπάνω παραθύρου αφορούν τη δεύτερη εκ των δύο εικόνων (I_2). Τέλος, προστίθενται οι είσοδοι *rho*, το οποίο εκφράζει το εύρος του γκαουσιανού παραθύρου, η θετική σταθερά κανονικοποίησης *epsilon* και την αρχική εκτίμηση d_0 για το πεδίο οπτικής ροής. Ο αλγόριθμος υπολογίζει τη λύση των ελαχίστων τετραγώνων, *u*, για μια

περιοχή γύρω από κάθε *feature*, την οποία εμείς ορίσαμε ως 2 *pixels*, και ανανεώνει το διάνυσμα οπτικής ροής d, σύμφωνα με τη σχέση: $d_{i+1} = d_i + u$,

όπου:

$$\mathbf{u}(\mathbf{x}) = \begin{bmatrix} (G_\rho * A_1^2)(\mathbf{x}) + \epsilon & (G_\rho * (A_1 A_2))(\mathbf{x}) \\ (G_\rho * (A_1 A_2))(\mathbf{x}) & (G_\rho * A_2^2)(\mathbf{x}) + \epsilon \end{bmatrix}^{-1} \cdot \begin{bmatrix} (G_\rho * (A_1 E))(\mathbf{x}) \\ (G_\rho * (A_2 E))(\mathbf{x}) \end{bmatrix}$$

ενώ:

$$A(\mathbf{x}) = \begin{bmatrix} A_1(\mathbf{x}) & A_2(\mathbf{x}) \end{bmatrix} = \begin{bmatrix} \frac{\partial I_{n-1}(\mathbf{x}+\mathbf{d}_i)}{\partial x} & \frac{\partial I_{n-1}(\mathbf{x}+\mathbf{d}_i)}{\partial y} \end{bmatrix}$$

$$E(\mathbf{x}) = I_n(\mathbf{x}) - I_{n-1}(\mathbf{x}+\mathbf{d}_i)$$

Η παραπάνω ανανέωση του διανύσματος συμβαίνει έως ότου να επέλθει η σύγκλιση, η οποία έχει καθοριστεί να σημειώνεται είτε στις πεντακόσιες επαναλήψεις, είτε εάν η νόρμα του διανύσματος $d_{i+1} - d_i$ λαμβάνει τιμήμικρότερη του κατωφλίου 0.005.
Κατά αυτόν τον τρόπο, επιστρέφεται το διάνυσμα οπτικής ροής των χαρακτηριστικών *d*.

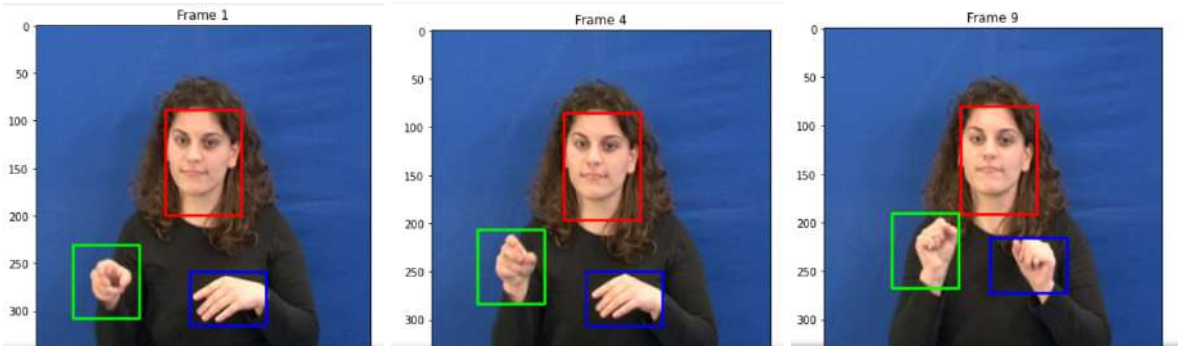
1.2.2 Υπολογισμός της Μετατόπισης των Παραθύρων από τα Διανύσματα Οπτικής Ροής

Αφού εξάγαμε την οπτική ροή για κάθε ένα από τα χαρακτηριστικά του προσώπου και των χεριών, χρησιμοποιώντας την συνάρτηση displacement (*displ*), επιτυγχάνεται η μετατόπιση του bounding κατά την ακολουθία των 66 frames. Με αυτή την συνάρτηση εντοπίζουμε τον μέσο όρο της μετατόπισης τον άξονα x και y μέσω της οπτικής ροής, καθώς η πλειονότητα των σημείων ενδιαφέροντος (στην περίπτωσή μας οι γωνίες) βρίσκονται σε σημεία με έντονη υφή. Συγκεκριμένα, για την επίτευξη μέγιστης ακρίβειας, ενώ παράλληλα για την απόρριψη των *outliers,* ο παραπάνω μέσος όρος αφορά αποκλειστικά τα χαρακτηριστικά των οποίων η ενέργεια διανύσματος υπερβαίνει το κατώφλι που ορίστηκε στο 0.1, όταν δηλαδή ισχύει:

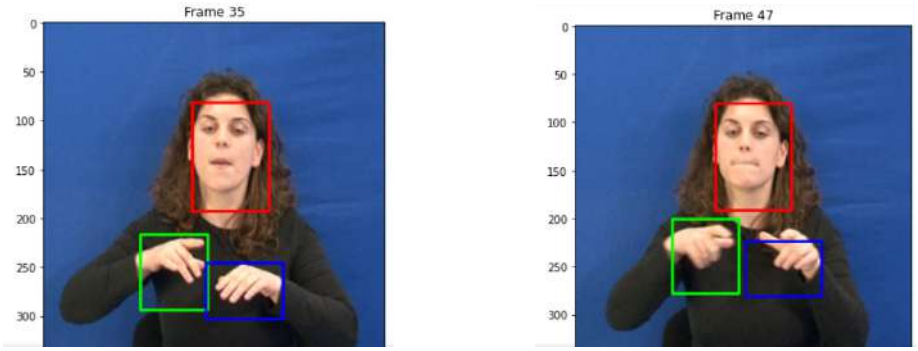
$$\|d\|^2 = d_x^2 + d_y^2 \geq 0.1$$

Παρακάτω, παρατίθεται ένα δείγμα από τις τελικές εικόνες της παρακολούθησης προσώπου και χεριών με το ορθογώνιο παρακολούθησης σχεδιασμένο σε κάθε μία. Το πλήρες δείγμα για την ακολουθία των 66 *frames* παρέχεται στον εργαστηριακό κώδικα του *jupyter notebook*.

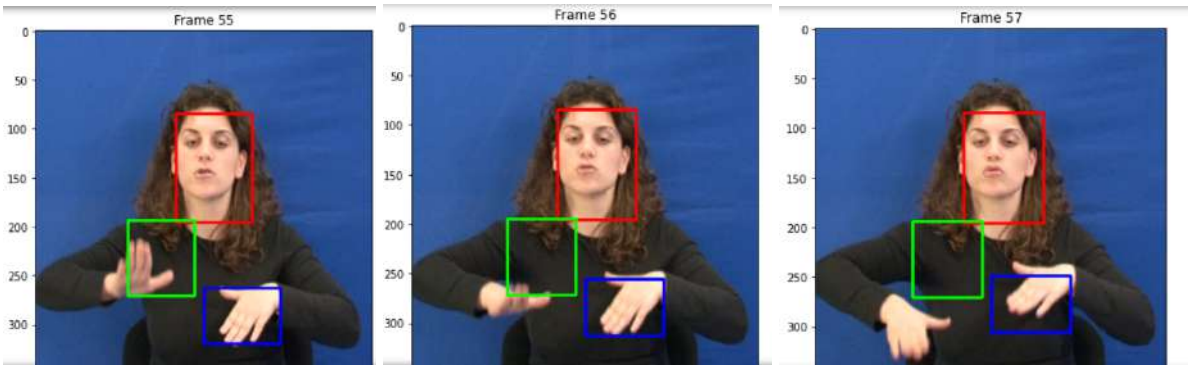
Frame 1 *Frame 4* *Frame 9*



Frame 35 *Frame 47*



Frame 55 *Frame 56* *Frame 57*



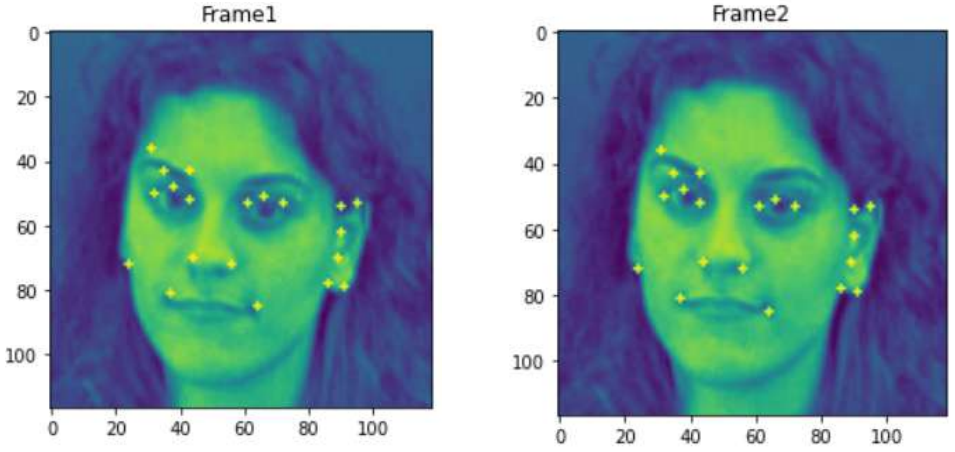
Ενώ παρατηρούμε, πως ο αλγόριθμος αποδίδει επιθυμητά αποτελέσματα για την πλειονότητα των frames, δηλαδή τόσο το πρόσωπο όσο και τα χέρια βρίσκονται εντός των *bounding boxes*, για την ακολουθία των *frames* 55-56-57, υπάρχει μικρή απόκλιση για την κίνηση του αριστερού χεριού (πράσινο μπλοκ). Συγκεκριμένα, φαίνεται πως ο αλγόριθμος αδυνατεί να ανιχνεύσει την κίνηση του αριστερού χειρου, εφόσον αυτό δεν περιβάλλεται από το πράσινο πλαίσιο. Η παραπάνω αστοχία του αλγορίθμου ήταν αναμενόμενη , διότι ο αλγόριθμος *Lucas Kanade* είναι σχεδιασμένος να υπολογίζει την οπτική ροή ανάμεσα σε διαδοχικά *frames*, τα οποία παρεκκλίνουν μόλις λίγα πίξελς (από 1 έως και το πολύ 5). Ωστόσο, από τις παραπάνω τρεις

φωτογραφίες γίνεται φανερό, ότι η νοηματίστρια κινείται πολύ έντονα μεταξύ των *frames* 55-56 και 56-57, με το εύρος κίνησης να υπερβαίνει κατά πολύ 5 πίξελ που είχαμε θέσει ως άνω όριο (φαίνεται μάλιστα το αριστερό χέρι να μετακινείται περίπου 50 *pixels*). Αυτός είναι ο λόγος που παρατηρείται απόκλιση αδυναμία ορθής μετατόπισης παραθύρου.

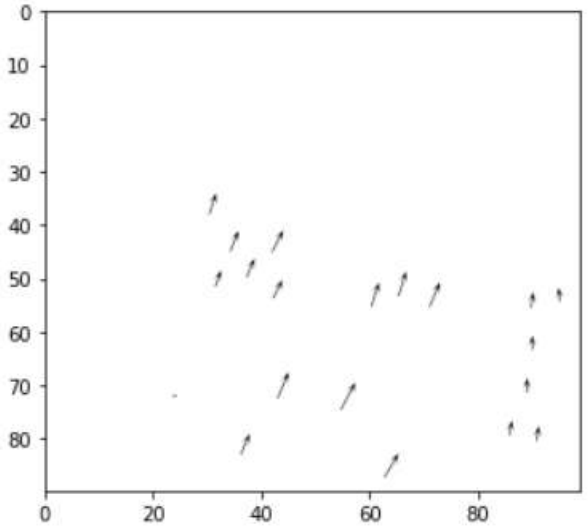
Για την οπτικοποίηση της διαδικασίας, παρατίθενται ορισμένα παραδείγματα δύο διαδοχικών frames και η αντίστοιχη απόδοση του διανύσματος οπτικής ροής για τα χαρακτηριστικά τους.

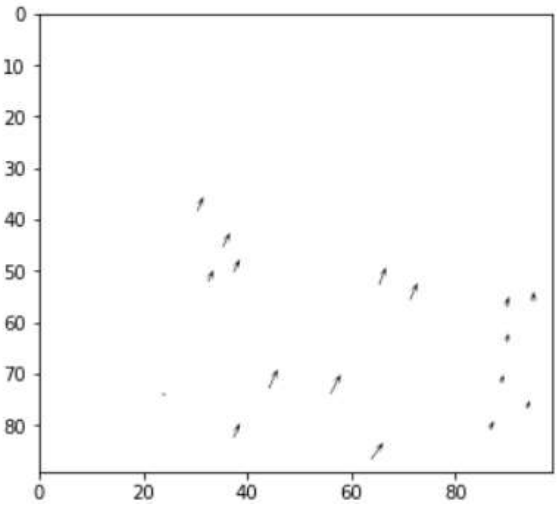
Πρόσωπο:

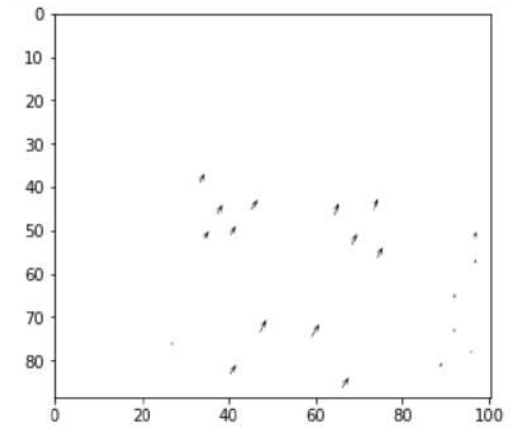
frames 1-2:

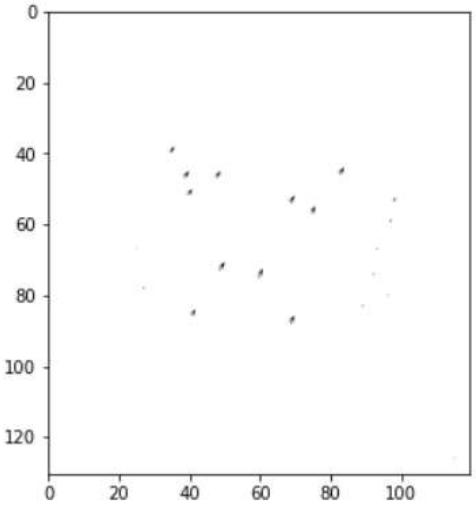


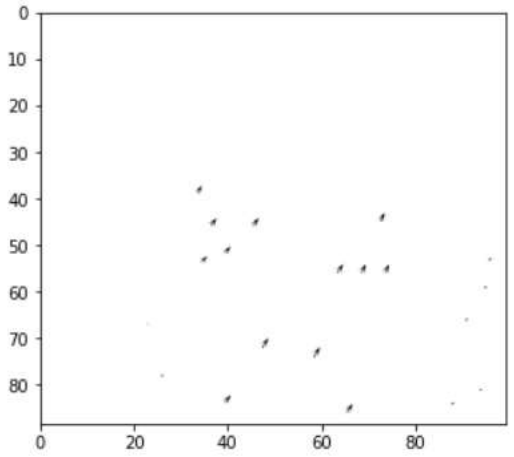
διανύσματα οπτικής ροής *d* για τα *frames* 1-2, 2-3, 3-4, 4-5, 5-6 αντίστοιχα:




0
10
20
30
40
50
60
70
80
0
20
40
60
80

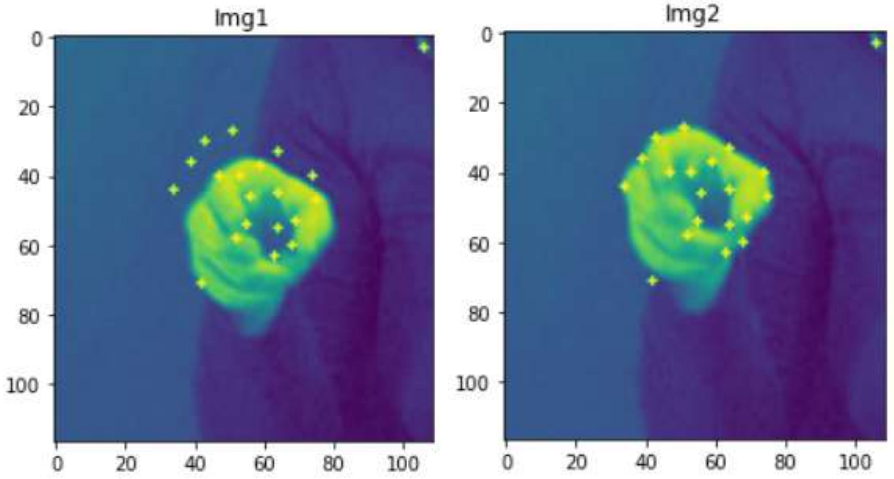

0
10
20
30
40
50
60
70
80
0
20
40
60
80
100


0
20
40
60
80
100
120
0
20
40
60
80
100

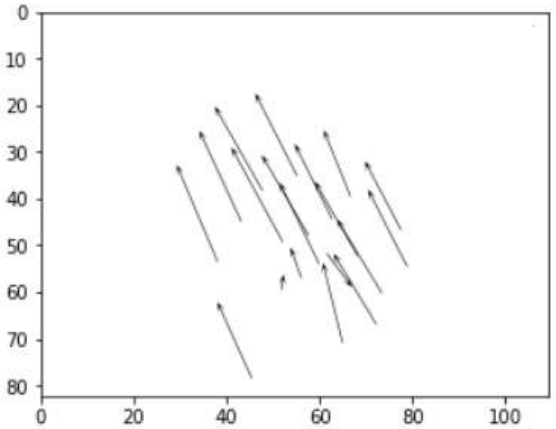


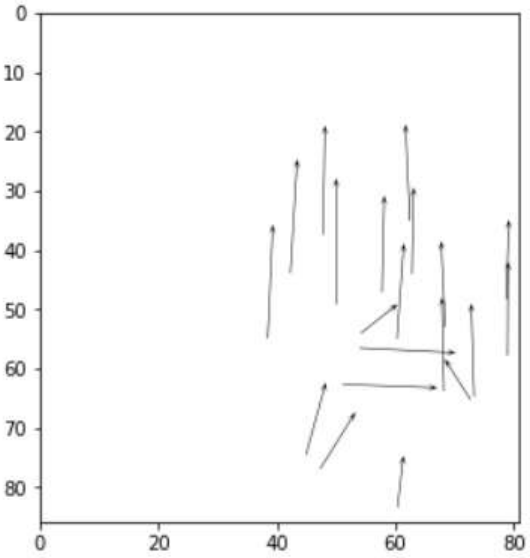
Δεξί Χέρι:

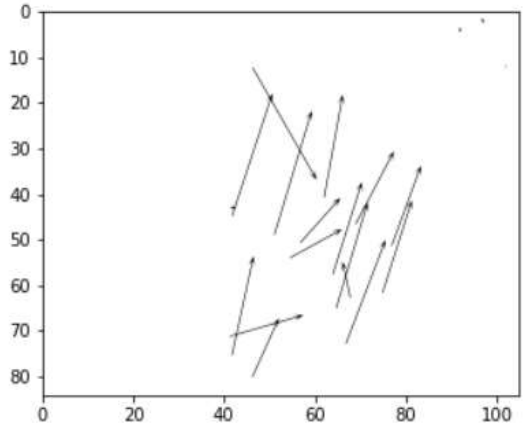
frames 1-2:

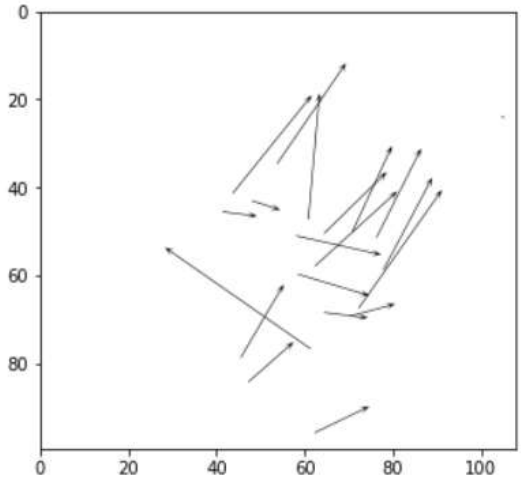


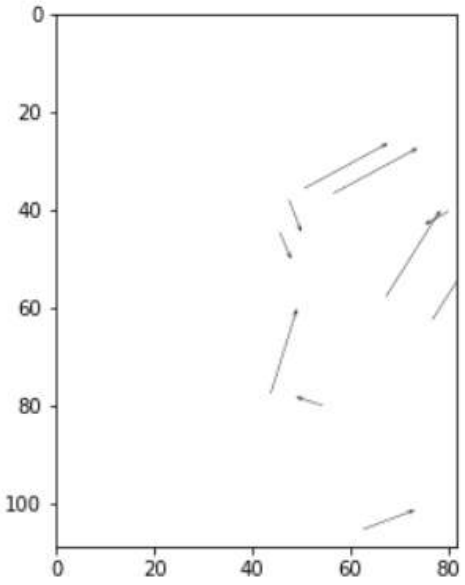
διανύσματα οπτικής ροής *d* για τα *frames* 1-2, 2-3, 3-4, 4-5, 5-6 αντίστοιχα:




0
10
20
30
40
50
60
70
80
0
20
40
60
80

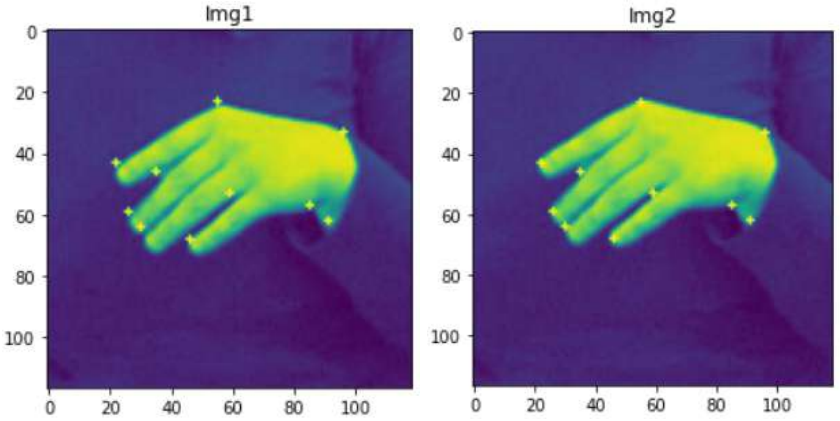

0
10
20
30
40
50
60
70
80
0
20
40
60
80
100


0
20
40
60
80
0
20
40
60
80
100

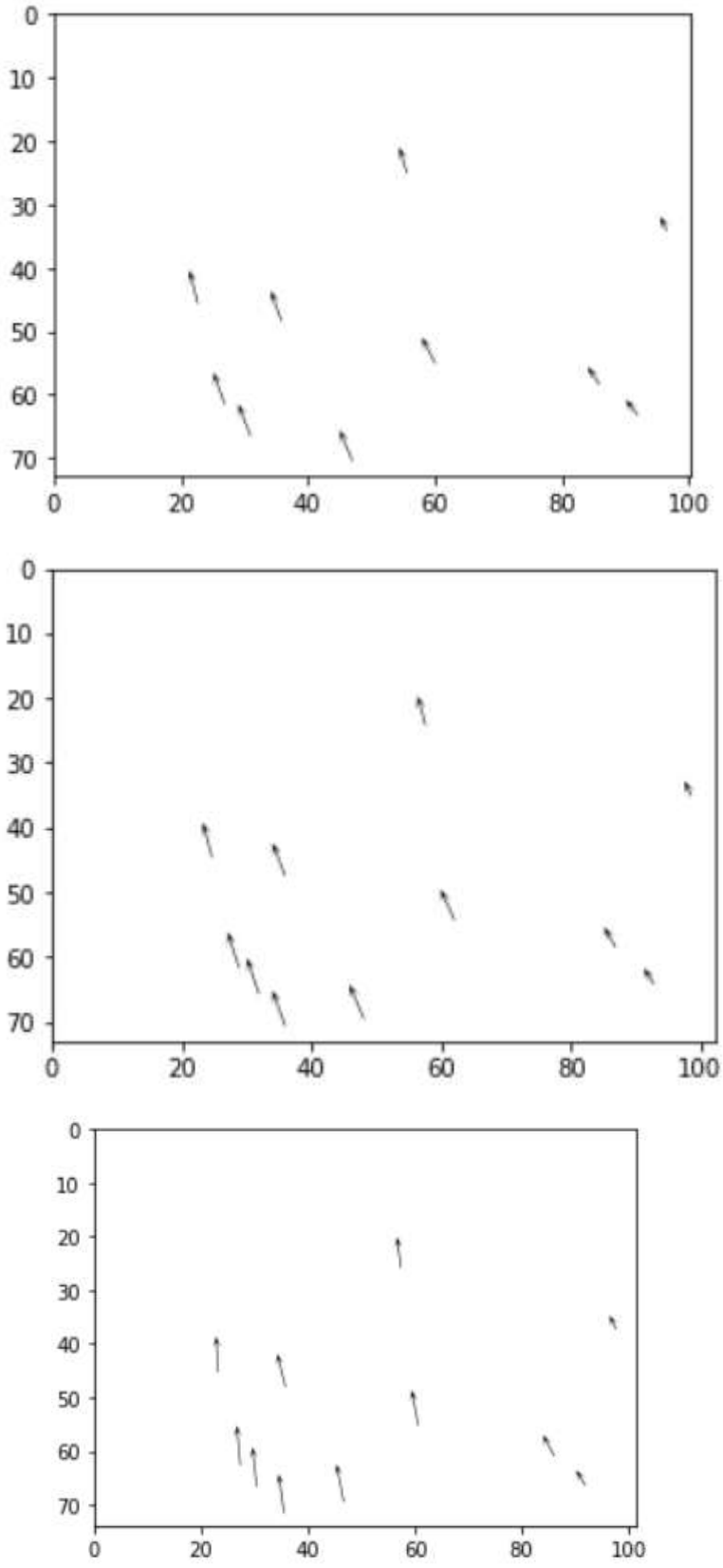


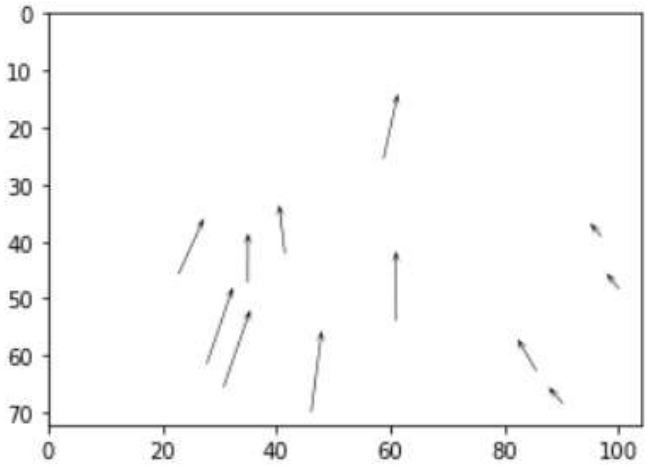
Αριστερό Χέρι:

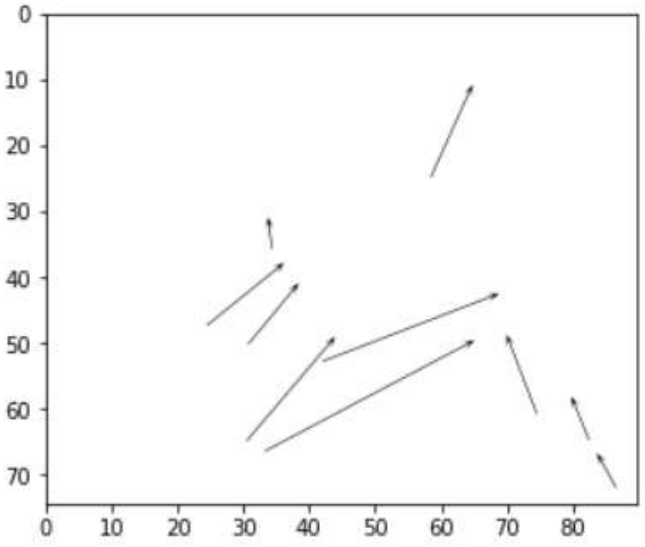
frames 1-2:



διανύσματα οπτικής ροής *d* για τα *frames* 1-2, 2-3, 3-4, 4-5, 5-6 αντίστοιχα:







<u>Πειραματισμός με τις παραμέτρους rho & epsilon</u>

Οι τιμές που επιλέγουμε για τα rho και epsilon καθορίζονται ανάλογα το είδος της κίνησης που θέλουμε να εντοπίσουμε.

Για μεγαλύτερο rho αυξάνεται το μήκος των διανυσμάτων της οπτικής ροης που αντιλαμβάνεται ο αλγόριθμος, ενώ για μικρότερο ανιχνεύει πιο έντονα τις επιμέρους αλλαγές, τις μικρότερες δηλαδή μικρότερες κινήσεις. Αν η τιμή του rho μειωθεί πάρα πολύ, τα διανύσματα της οπτικής ροής μικραίνουν. Έτσι αλγόριθμος καταλαβαίνει πολύ μικρή κίνηση με αποτέλεσμα γειτονικά διανύσματα να τέμνονται. Έτσι ανίχνευση της κίνησης δεν καθίσταται εφικτή.

Αν πάλι η τιμή αυξηθεί πάρα πολύ, ο αλγόριθμος αντιλαμβάνεται το ορθογώνιο κομμάτι της εικόνας ως ένα αυτούσιο χαρακτηριστικό, οπότε, στην πραγματικότητα, η οπτική ροή γίνεται απλά η μετακίνηση ολόκληρης της περιοχής ενδιαφέροντος και δεν μπορούμε να ξεχωρίσουμε τα επιμέρους χαρακτηριστικά gestures.

Αντίστοιχα, με το *epsilon*, θέλουμε να εντοπίσουμε το πού υφίσταται κίνηση. Για μεγάλο *epsilon* εντοπίζεται κίνηση σε λιγότερες περιοχές
(άρα τα διανύσματα οπτικής ροής είναι πιο απότομα), ενώ με μικρό *epsilon* , την εντοπίζουμε σε περισσότερες περιοχές.

Δεδομένου ότι στο πρόσωπο είναι επιθυμητή η ανίχνευση κίνησης σε μια ευρεία περιοχή, δίχως όμως τις επιμέρους λεπτομέρειες ούτε μεγάλο εύρος κινήσεων, προτιμάται η επιλογή μικρότερου *rho* και μεγαλύτερου *epsilon*.

Αντιθέτως στα χέρια επειδή θέλουμε να απεικονίσουμε την χαρακτηριστική κίνηση της νοηματήστριας, χωρίς όμως να μας ενδιαφέρει η κίνηση σε ολόκληρο το ορθογώνιο αλλά μόνο εκεί που αυτή γίνεται πιο έντονη, αποφασίσαμε να χρησιμοποιήσουμε μεγαλύτερο rho και μικρότερο epsilon.

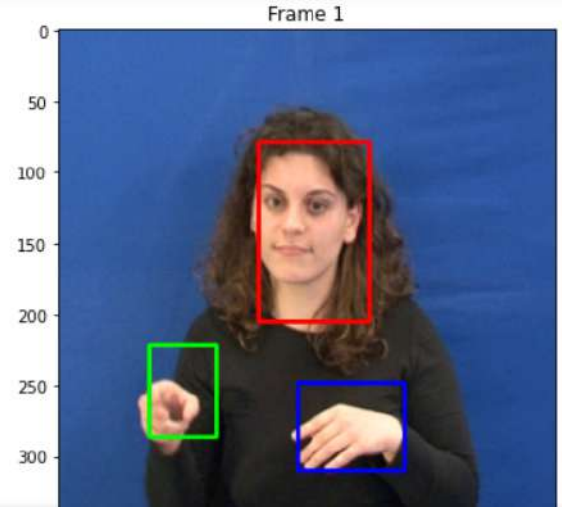
1.2.3 Πολυ-Κλιμακωτός Υπολογισμός Οπτικής Ροής

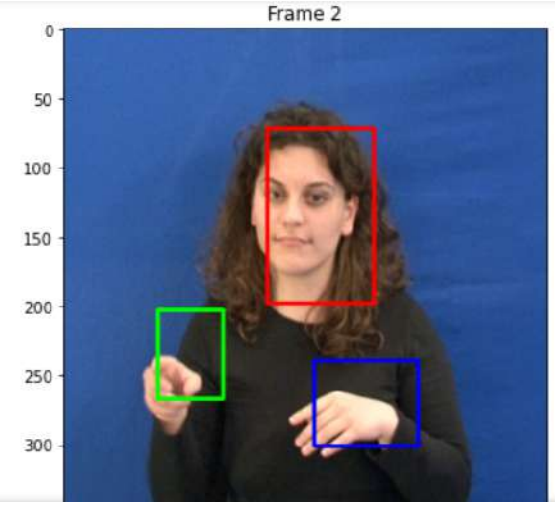
Η πολυκλιμακωτή εκδοχή του αλγορίθμου Lucas Kanade αναλύει τις αρχικές εικόνες σε γκαουσιανές πυραμίδες, δεχόμενος ως είσοδο τον επιθυμητό αριθμό των κλιμάκων της πυραμίδας, ενώ ύστερα υπολογίζει την οπτική ροή από τις πιο μικρές (τραχείς) στις πιο μεγάλες (λεπτομερείς) κλίμακες, χρησιμοποιώντας τη λύση της μικρής κλίμακας ως αρχική συνθήκη για τη μεγάλη κλίμακα.

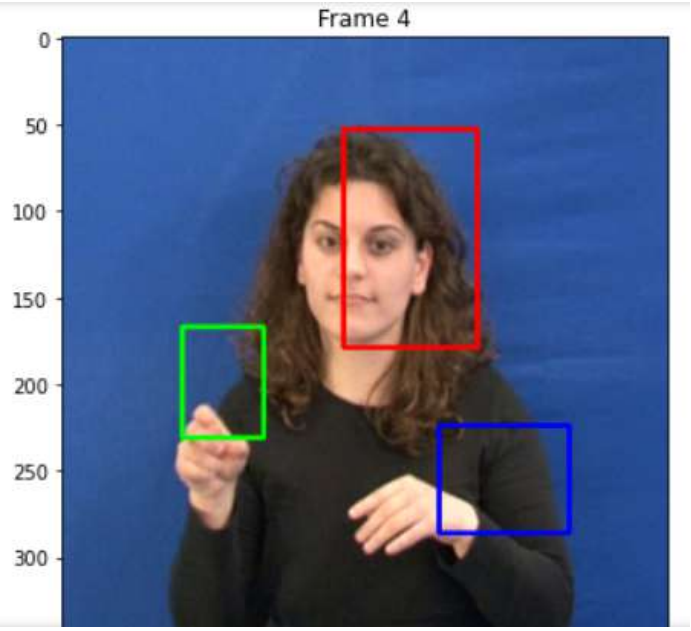
Για τη μετάβαση από μεγάλες σε μικρές κλίμακες κατά την κατασκευή της γκαουσιανής πυραμίδας η εικόνα φιλτράρεται με βαθυπερατό φίλτρο γκαουσιανή τυπικής απόκλισης τριών pixels πριν την υποδειγματοληψία, για την αποφυγή του aliasing. Αντίστοιχα, κατα τη μετάβαση από μικρές σε μεγάλες κλίμακες, διπλασιάζεται το διάνυσμα οπτικής ροής d.

Για τον υπολογισμό της πολυκλιμακωτής μορφής του αλγορίθμου *Lucas-Kanade*, γίνεται μια μικρή τροποποίηση στον μονοκλιμακωτό αλγόριθμο *Lucas-Kanade,* ούτως ώστε να επιστρέφει το διάνυσμα οπτικής ροής για όλα όλα τα πίξελ της εικόνας. Συγκεκριμένα, μόνο στα *pixels* των γωνιών αποδόθηκε τιμή στο διάνυσμα *d*, ενώ όλα τα άλλα *pixels* λαμβάνουν τιμή μηδέν. Κατά συνέπεια, και ο αλγόριθμος *displ* υπέστη μικρή τροποποίηση ώστε να μπορεί να επεξεργάζεται τον *Lucas-Kanade,* και να εξάγει το μέσο όρο του διανύσματος οπτικής ροής.

Τέλος, παραθέτουμε και τα αντίστοιχα δείγμα από τις τελικές εικόνες της παρακολούθησης προσώπου και χεριών με το ορθογώνιο παρακολούθησης σχεδιασμένο σε κάθε μία για τον πολυκλιμακωτό αλγόριθμο *Lucas Kanade*. Τόσο στον κώδικα όσο και στο παρόν δε παρουσιάζεται η πλήρης ακολουθία των 66 *frames*, διότι ο αλγόριθμος δείχνει να παρεκκλίνει πολύ νωρίς και δεν έχει νόημα η συνέχεια της διεξαγωγής του.







Μέρος 2: Εντοπισμός Χωρο-χρονικών Σημείων Ενδιαφέροντος και Εξαγωγή Χαρακτηριστικών σε Βίντεο Ανθρωπίνων Δράσεων

Το 2ο εργαστηριακό μέρος του πραγματεύεται την εξαγωγή χωροχρονικών χαρακτηριστικών με στόχο την εφαρμογή τους στο πρόβλημα κατηγοριοποίησης βίντεο που περιέχουν ανθρώπινες δράσεις.

Στο πλαίσιο της άσκησης αυτής δόθηκαν βίντεο από 3 κλάσεις δράσεων (walking, running, boxing) από στα οποία θα εκπαιδεύσουμε και έπειτα θα εξετάσουμε χωρο-χρονικούς περιγραφητές με σκοπό την κατηγοριοποίηση των δράσεων που απεικονίζουν.

2.1 Χωροχρονικά Σημεία Ενδιαφέροντος

Στην εργαστηριακή αυτή άσκηση χρησιμοποιήθηκαν δύο ανιχνευτές χωρο-χρονικών σημείων ενδιαφέροντος: Harris Detector και Gabor Detector. Η αναπαράσταση με χρήση τοπικών χαρακτηριστικών έχει επικρατήσει και στην αναγνώριση ανθρώπινων δράσεων, όπου γίνεται μια επιλογή από δεδομένα που αφ΄ενός μεν μειώνουν κατά πολύ τη διάσταση των βίντεο και αφ΄ετέρου δε τα μετασχηματίζουν σε μια αναπαράσταση που τα κάνει διαχωρίσιμα.
Περισσότερες λεπτομέρειες για κάθε έναν από αυτούς παραθέτουμε παρακάτω:

2.1.1 Harris Detector

Σε αυτό το ερώτημα υλοποιήσαμε τον ανιχνευτή Harris με σκοπό να εξάγουμε χώρο-χρονικά σημεία ενδιαφέροντος από διαδοχικά frames.
Η τρισδιάστατη έκδοσή του έχει τη δυνατότητα να ανιχνεύει, ακολουθώντας την ίδια λογική με την οποία ο δισδιάστατος Harris ανιχνεύει σημεία που επιδεικνύουν υψηλή μεταβολή των τιμών της εικόνας στο χώρο, σημεία που επιδεικνύουν και μη σταθερή κίνηση στο χρόνο. Για την ακρίβεια ο Harris3D πυροδοτεί ανιχνεύσεις σε περιοχές που πληρούν και τις δύο παραπάνω προϋποθέσεις, δηλαδή παρουσιάζουν διακριτική εμφάνιση στο χώρο και διέπονται από μη σταθερή κίνηση στο χρόνο. Τέτοια χαρακτηριστικά αντιστοιχούν σε γεγονότα με έντονο πληροφοριακό περιεχόμενο στο video εισόδου. Ο Harris3D ψάχνει στο video εισόδου για σημεία που μεγιστοποιούν μια συνάρτηση χωροχρονικής γωνιότητας.

Στο παρόν εργαστήριο αναζητούμε σημεία τα οποία όχι μόνο μεταβάλλονται στις δύο διευθύνσεις αλλα και στον χρόνο.

Για κάθε voxel του βίντεο υπολογίσαμε τον 3 × 3 πίνακα M (x, y, t) προσθέτοντας στον δισδιάστατο δομικό τανυστή και τη χρονική παράγωγο σε μορφή πινάκων σύμφωνα με τον τύπο:

$$M(x, y, t; \sigma, \tau) = g(x, y, t; s\sigma, s\tau) * \begin{pmatrix} L_x^2 & L_x L_y & L_x L_t \\ L_x L_y & L_y^2 & L_y L_t \\ L_x L_t & L_y L_t & L_t^2 \end{pmatrix}$$

όπου ο πρώτος συντελεστής είναι ένας τρισδιάστατος γκαουσιανός πυρήνας ομαλοποίησης και το δεύτερο οι χωροχρονικές παράγωγοι για την χωρική κλίμακα σ και την χρονική κλίμακα τ.

Για τις κατευθυντικές παραγώγους, εφαρμώσαμε συνέλιξη ως προς κάθε διάσταση με έναν πυρήνα κεντρικών διαφορών [-101] T. Το αποτέλεσμα της συνέλιξης της μίας διάστασης αποτελεί είσοδο στη συνέλιξη της επόμενης.

Τέλος, το κριτήριο γωνιότητας που χρησιμοποιήσαμε έχει τη μορφή:

$$H(x, y, t) = det(M(x, y, t)) - k \cdot trace^3(M(x, y, t))$$

2.1.2 Gabor Detector

Η επέκταση του ανιχνευτή Gabor βασίζεται στο χρονικό φιλτράρισμα του βίντεο μέσω ενός περιττού και ενός άρτιου φίλτρου Gabor. Συνεπώς, δημιουργήσαμε ένα ζεύγος φίλτρων σύμφωνα με τις σχέσεις:

$$h_{ev}(t; \tau, \omega) = \cos(2\pi t\omega) \exp(-t^2/2\tau^2)$$

$$h_{od}(t; \tau, \omega) = \sin(2\pi t\omega) \exp(-t^2/2\tau^2)$$

σε διάστημα [-2τ, 2τ], η συχνότητα ω σε κάθε φίλτρο σχετίζεται με την χρονική κλίμακα τ μέσω της σχέσης $\omega = 4/\tau$.

Φιλτράροντας χρονικά τα βιντεο με τα παραπάνω φίλτρα και υπολογίζοντας τη συνέλιξη των frames με τα φίλτρα αυτά και τον gaussian πυρήνα, καταλήγουμε στο κριτήριο τετραγωνικής ενέργειας:

$$H(x, y, t) = (I(x, y, t) * g * h_{ev})^2 + (I(x, y, t) * g * h_{od})^2$$

2.1.3

Εντός των συναρτήσεων των τοπικών περιγραφητών που υλοποιηθηκαν στα παραπάνω ερωτήματα (*def Harris_Stephens* και *def Gabor*) έχει υλοποιηθεί το κριτήριο που υπολογίζει τα σημεία ενδιαφέροντος ως τα τοπικά μέγιστα του κριτηρίου σημαντικότητας. Κατά συνέπεια, οι δύο αυτές συναρτήσεις επιστρέφουν τα N σημεία με τις μεγαλύτερες τιμές του κριτηρίου σημαντικότητας (π.χ. τα 500-600 πρώτα).

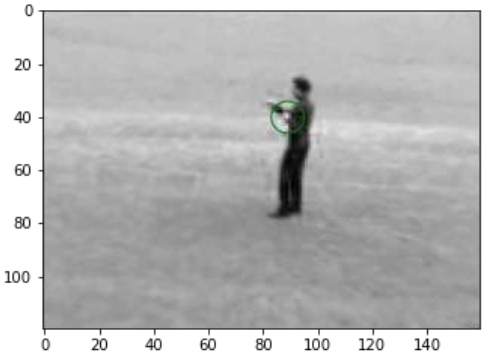
Εν συνεχεία, χρησιμοποιώντας τη συνάρτηση
show_detection(Video, DetectorPoints, save_path)
αποδίδουμε για ένα βίντεο από κάθε δραστηριότητα και για κάθε ανιχνευτή τα σημεία που προκύπτουν. Τέλος, στη μεταβλητή *save_path,* ορίζουμε το directory (εφόσον πρώτα το έχουμε ορίσει) στο οποίο επιθυμούμε να αποθηκευτεί η ακολουθία των frames με τα σημεία.

Ακολουθούν ενδεικτικά frames από τα δεδομένα videos.

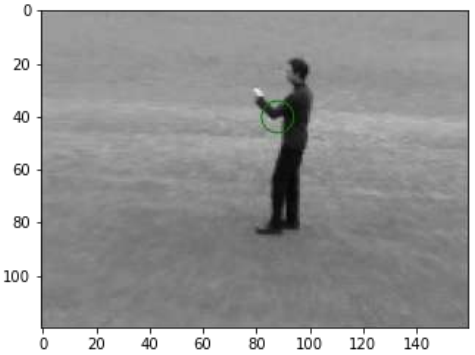
Harris-Stefens Detector

Boxing (video: person01_boxing_d2_uncomp.avi)

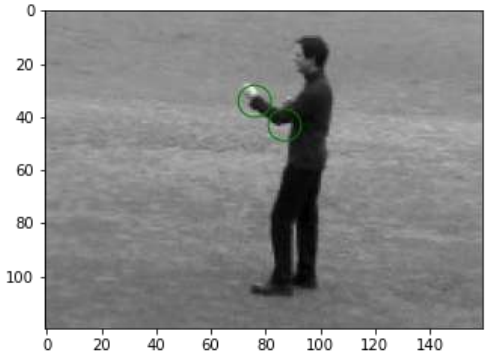
Frame 1



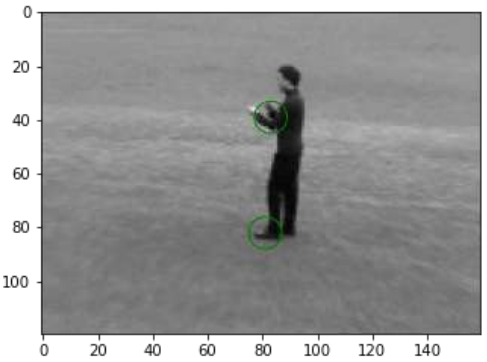
Frame 2



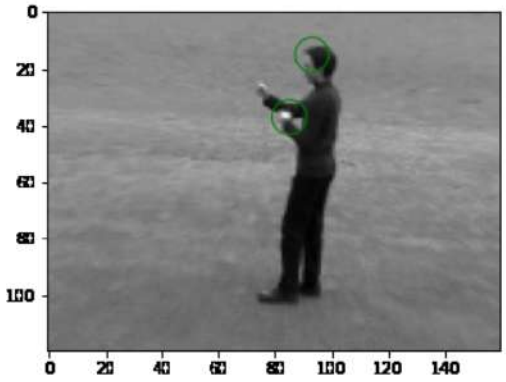
Frame 3



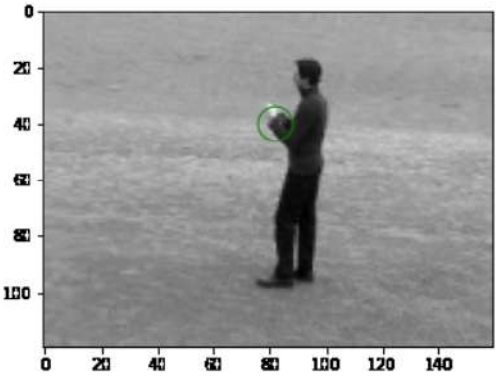
Frame 4



Frame 13

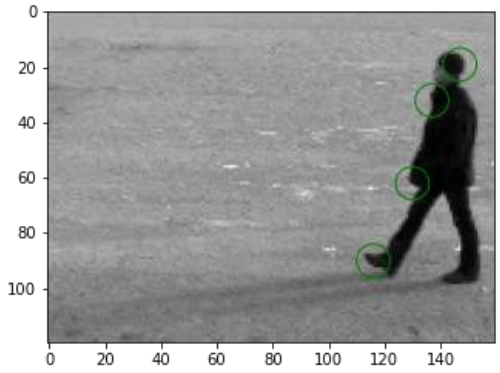


Frame 48

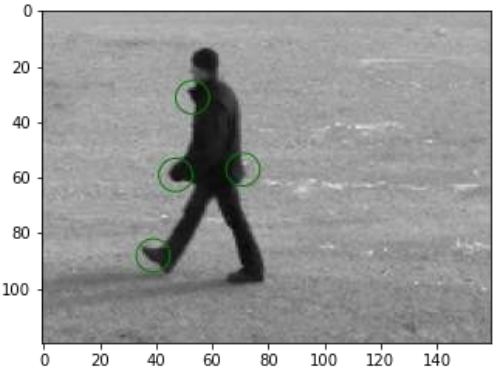


Walking (video: person04_walking_d1_uncomp.avi)

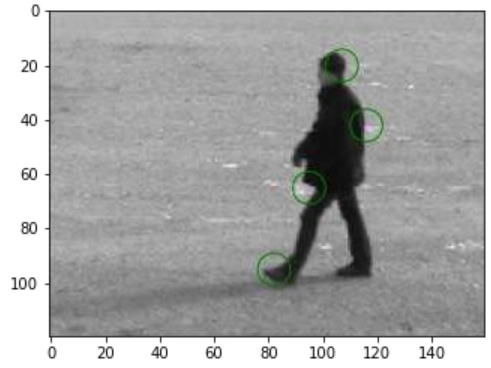
Frame 1



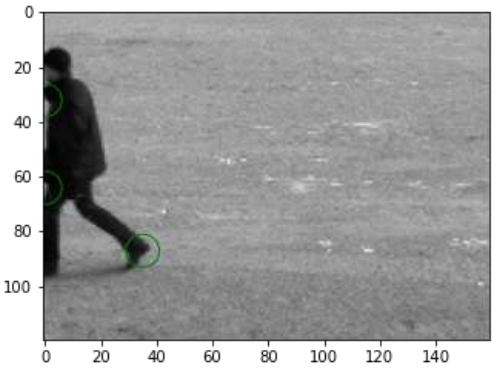
Frame 2



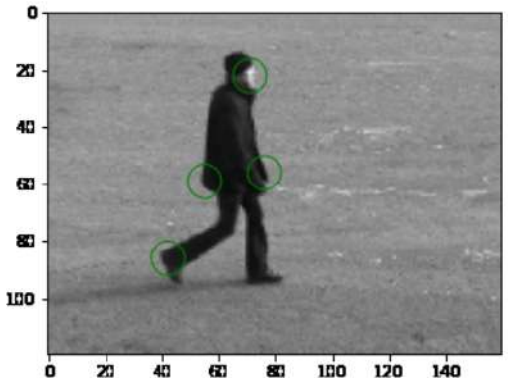
Frame 7



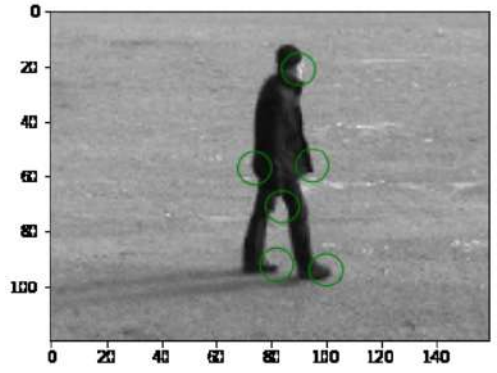
Frame 14



Frame 148

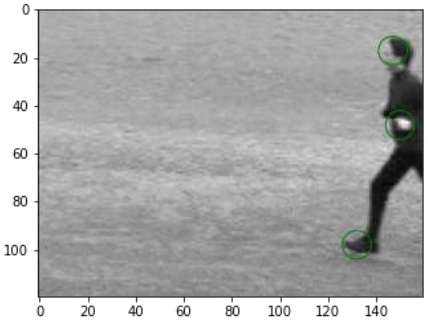


Frame 155

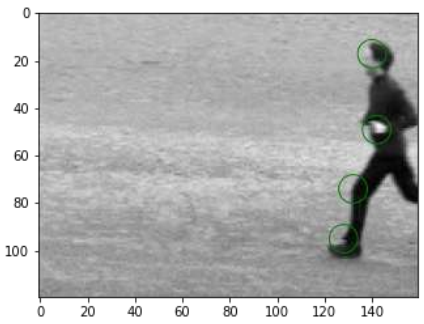


Running (video: person01_running_d1_uncomp.avi)

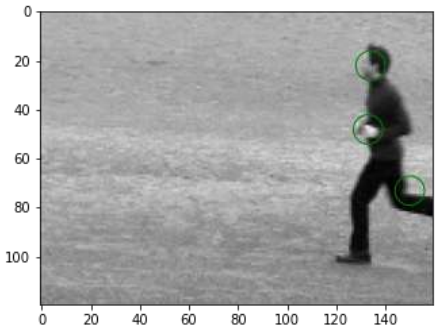
Frame 5



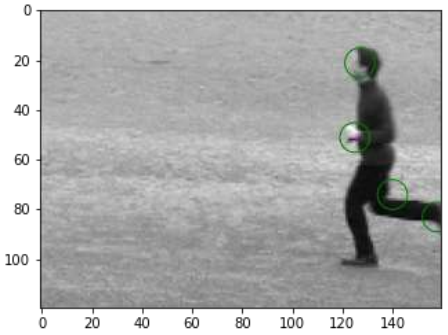
Frame 6



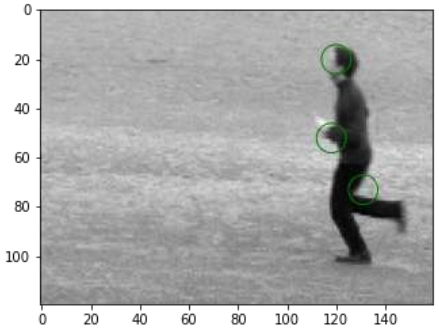
Frame 7



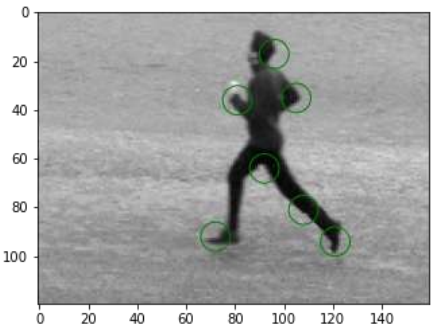
Frame 8



Frame 9



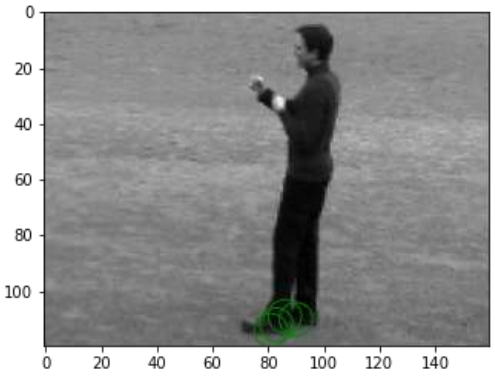
Frame 14



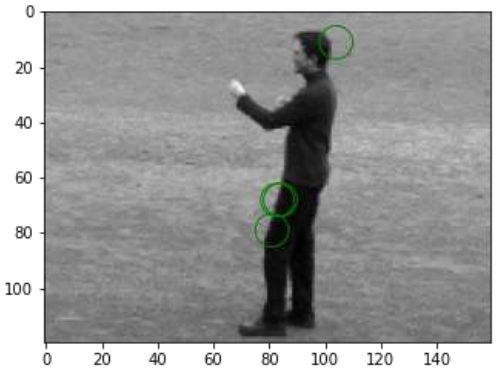
Σχετικά με τον Harris Detector, παρατηρούμε παρόμοια συμπεριφορά με τον αλγόριθμο Gabor, χωρικά. Χρονικά, για πολύ μικρές τιμές χρονικής κλίμακας παρατηρείται ολική αδυναμία εντοπισμού κίνησης διότι δεν λαμβάνεται αρκετή πληροφορία από τα επόμενα – προηγούμενα frames.

Gabor Detector (video: person01_boxing_d2_uncomp.avi)
Boxing

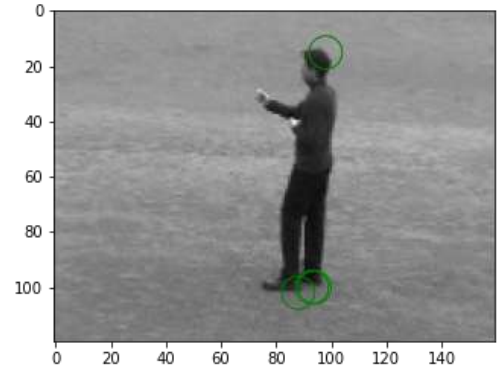
Frame 1



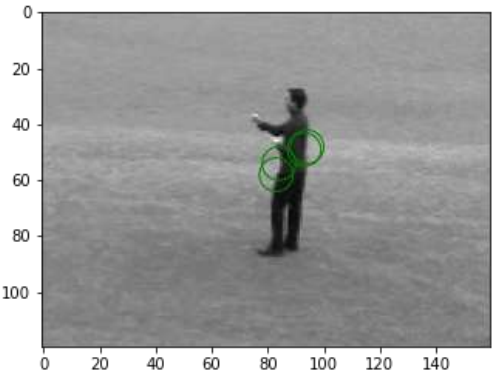
Frame 2



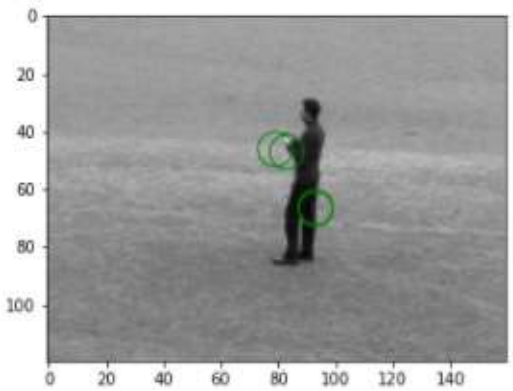
Frame 3



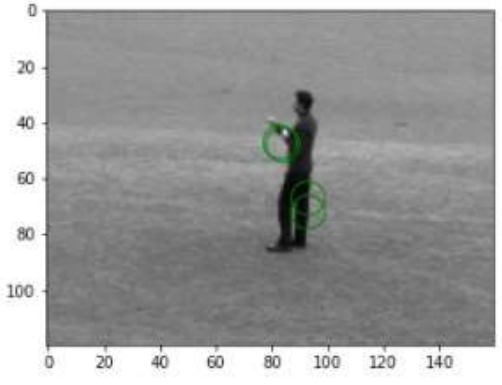
Frame 4



Frame 5

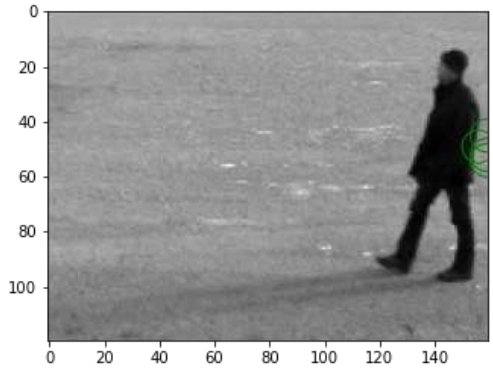


Frame 6

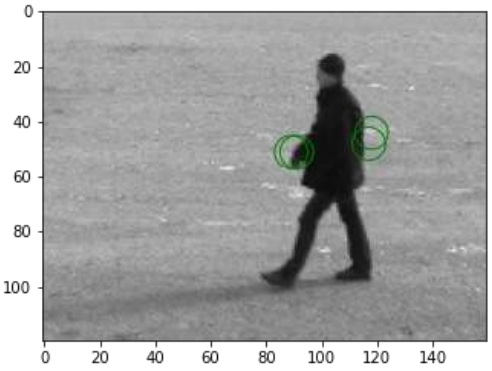


Walking (video: person04_walking_d1_uncomp.avi)

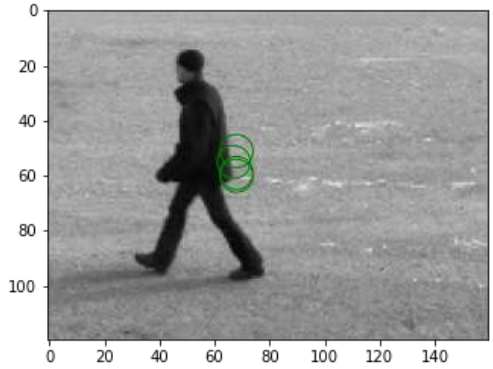
Frame 18



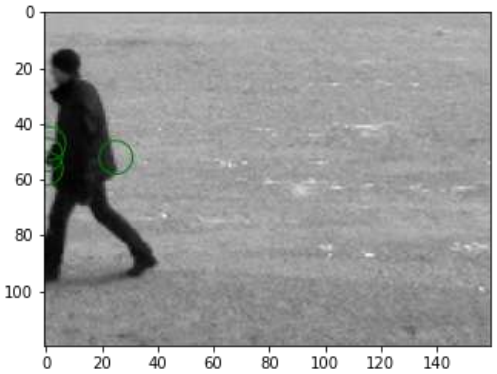
Frame 32



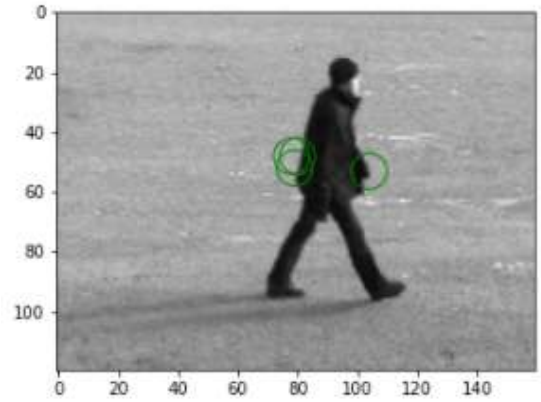
Frame 49



Frame 61



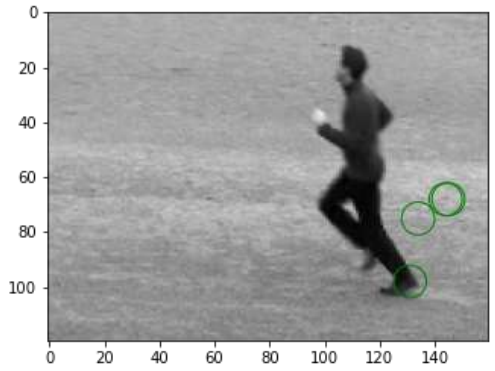
Frame 155



Frame 156

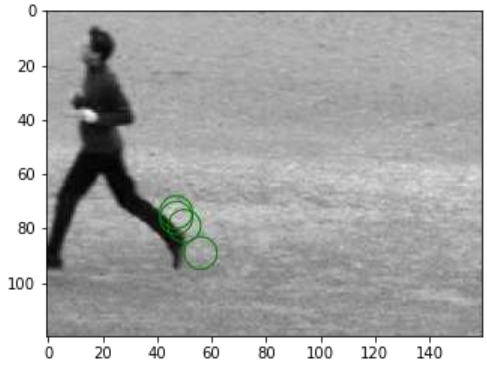
Running

Frame 11

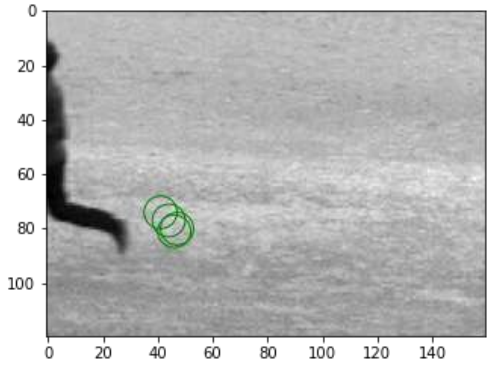


Frame 18

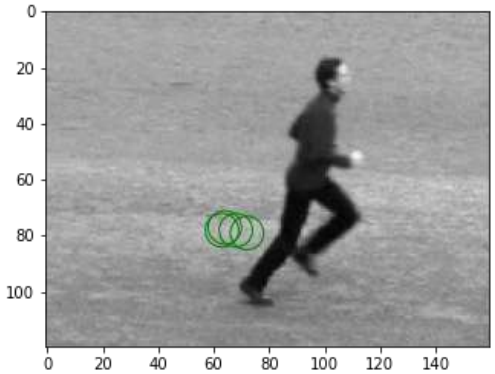
Frame 25



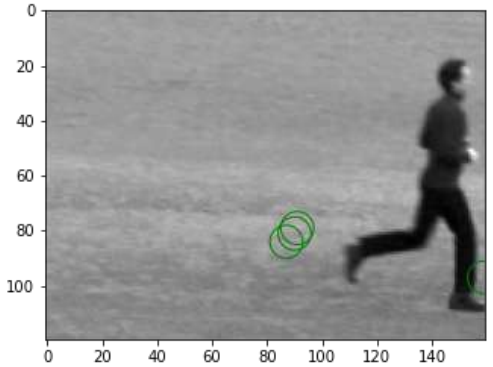
Frame 28



Frame 112



Frame 118



Όσον αφορά τον Gabor Detector, παρατηρούμε γενικά καλά αποτελέσματα για μεγάλες τιμές της χωρικής κλίμακας. Σχετικά με την χρονική κλίμακα εδώ η χρονική σταθερά αυξάνεται, παράγονται χειρότερα αποτελέσματα.

Συγκεντρωτικά, για τους δύο ανιχνευτές:

Και οι δύο ανιχνευτές καταφέρνουν να εντοπίσουν την κίνηση στα σημεία ενδιαφέροντος (χέρια και πόδια), με τον ανιχνευτή Gabor είναι αρκετά πιο ακριβής, αφού σπάνια "χάνει" σημεία ενδιαφέροντος σε frame στο οποίο εμφανίζεται άνθρωπος Με μια πιο λεπτομερή εξέταση των αποτελεσμάτων βλέπουμε, ότι ο Gabor εντοπίζει σημεία ενδιαφέροντος σε όλα τα επιμέρους σημεία ενδιαφέροντος στο σώμα του ανθρώπου που εμφανίζει κίνηση. Αντίθετα, τα σημεία ενδιαφέροντος που εντοπίζει ο ανιχνευτής Harris δεν είναι αρκετά αντιπροσωπευτικά, τόσο ως προς το πλήθος των πλαισίων όσο και ως προς την πυκνότητα σε κάθε πλαίσιο.

Και οι δύο ανιχνευτές φαίνεται να ανταποκρίνονται σωστά στην ανίχνευση έντονων κινήσεων, ανεξάρτητα από το πόσο γρήγορες ή αργές είναι αυτές (μιλώντας πάντα σε πλαίσια ανθρώπινης κίνησης) ωστόσο σε ορισμένες περιπτώσεις πιάνουν και πολύ μικρές κινήσεις, όπως την κίνηση των ρούχων του ανθρώπου, η μικρές κινήσεις ισορροπίας που κάνει ο άνθρωπος ασυναίσθητα.

2.2 Χωρο-χρονικοί Ιστογραφικοί Περιγραφητές

2.2.1

Για κάθε frame του βίντεο υπολογίσαμε το διάνυσμα κλίσης και την TVL1 οπτική ροή χρησιμοποιώντας τη συνάρτηση *cv2.DualTVL1OpticalFlow_create*.

2.2.2

Στη συνέχεια χρησιμοποιήσαμε τη συνάρτηση orientation_histogram από το συμπληρωματικό υλικό του μέρους αυτού προκειμένου να υπολογίσουμε τους 2 ιστογραφικούς περιγραφητές.

Ορίσαμε τετράγωνα με διαστάσεις 4σ εκατέρωθεν του σημείου ενδιαφέροντος. Στη συνέχεια υπολογίσαμε τους τοπικούς περιγραφητές ανάλογα, HOG για κατευθυντικές παραγώγους, HOF για κατεύθυνση ροής. Η έξοδος αποθηκεύτηκε στη γραμμή ενός πίνακα, επαναλάβαμε τη διαδικασία για τα υπόλοιπα σημεία προσθέτοντας τις γραμμές στον εκάστοτε πίνακα.

2.3: Κατασκευή Bag of Visual Words και χρήση Support Vector Machines για την ταξινόμηση δράσεων

2.3.1
Διαχωρίσαμε το σύνολο των βίντεο σε σύνολο εκπαίδευσης (train set) και σύνολο δοκιμής (test set) με βάση το αρχείο που μας δίνεται στο συμπληρωματικό υλικό, το οποίο περιέχει το σύνολο των videos που ανήκουν στο train set. Η συνάρτηση *split_sets()*, ανάλογα τις ανάγκες των αλγορίθμων επιστρέφει το train & test set των videos καθώς επίσης και τα ονόματα των videos σε κάθε περίπτωση.

2.3.2
Για κάθε τίτλο στο training και testing set καλέσαμε την διαδικασία του ερωτήματος 2.2. Έπειτα εκτελέσαμε όλους τους πιθανούς συνδυασμούς ανάμεσα στους δύο ταξινομητές και περιγραφητές (Harris_HOG, Harris_HOF, Gabor_HOG, Gabor_HOF).

Υπολογίσαμε την τελική αναπαράσταση (global representation) για κάθε βίντεο με την bag of visual words (BoVW) τεχνική που περιγράφεται στην 1η εργαστηριακή άσκηση, χρησιμοποιώντας μόνο τα βίντεο εκπαίδευσης. Για τον υπολογισμό των BoVW ιστογραμμάτων χρησιμοποιήστε τη συνάρτηση bag_of_words συνάρτηση από το συμπληρωματικό υλικό αυτού του μέρους.

2.3.3
Για τον HOG περιγραφητή, βλέπουμε από την εκτέλεση ότι παραμένει αμετάβλητος σε φωτομετρικές και γεωμετρικές μεταβολές λόγω του υπολογισμού του σε μικρές γειτονιές.

Ο περιγραφητής HOF ο οποίος βασίζεται στον υπολογισμό της οπτικής ροής δεν απέδωσε τόσο καλά αποτελέσματα όσο ο HOG. Με κατάλληλες όμως τροποποιήσεις στις παραμέτρους της Lucas - Kanade πετύχαμε υψηλά ποσοστά αναγνώρισης.

2.3.4
Τρέξαμε τους αλγορίθμους αρκετές φορές και τα βέλτιστα αποτελέσματα που λάβαμε ήταν τα ακόλουθα:

```
Harris_HOG:  0.775
Harris_HOF:  0.683
Harris_HOG_HOF:  0.937
Gabor_HOG:  0.788
Gabor_HOF:  0.745
Gabor_HOG_HOF:  0.792
```

Ο καλύτερος συνδυασμός ήταν αυτός των περιγραφητών Harris HOG HOF με ακρίβεια 0.937. Το συνδυασμένο σχήμα HOG/HOF εκμεταλλεύεται την πληροφορία των

κατευθυνόμενων gradients αλλά και της οπτικής ροής, σε μια δομή πλέγματος στην οποία διαμερίζεται το κάθε τοπικό τεμάχιο. Τα εν λόγω ιστογράμματα στοχεύουν στο να συλλάβουν την τοπική εμφάνιση και κίνηση αντίστοιχα, στις γειτονιές των σημείων ενδιαφέροντος. Ο περιγραφητής HOG/HOF συνίσταται στον υπολογισμό ιστογραμμάτων χωρικών gradients και οπτικής ροής στη χωροχρονική γειτονιά των σημείων ενδιαφέροντος που έχουν προηγουμένως εξαχθεί από κάποιον ανιχνευτή στο video εισόδου.

Για την παραπάνω ταξινόμηση χρησιμοποιούνται επιπλέον οι Μηχανές Διανυσμάτων Υποστήριξης (SVMs). Πρόκειται για ταξινομητές μηχανικής μάθησης που επιτυγχάνουν μεγάλα περιθώρια διαχωρισμού ενώ παρουσιάζουν επιτυχία σε προβλήματα αναγνώρισης οπτικών προτύπων. Λόγω του ότι συνδυάζονται άψογα με τα ιστογράμματα BoW και αποτελούν τη συνήθη τεχνική ταξινόμησης ανθρώπινων δράσεων σε video.

2.3.5

Τέλος, κατά τον πειραματισμό με διαφορετικούς διαμερισμούς των δεδομένων σε train και test set αναμέναμε εκπαιδεύοντας τους ταξινομητές με videos που περιέχουν περισσότερα σημεία ενδιαφέροντος, παρουσιάζουν δηλαδή εντονότερη κινητικότητα, να πετύχουμε στη συνέχεια υψηλότερη ακρίβεια στο test set. Αυτή μας η σκέψη όμως, δεν ήταν απόλυτα ορθή διότι παρατηρήσαμε, τρέχοντας αρκετές φορές του αλγορίθμους ταξινόμησης στα δύο set, τα ποσοστά ακρίβειας να ποικίλουν και να μην είναι σημαντικά υψηλότερα του αρχικού διαμοιρασμού που μας είχε δοθεί. Να παρουσιάζουν δηλαδή μια τυχαιότητα ως προς τις τιμές τους, διατηρώντας ωστόσο την διάταξη επιδόσεων όπως ακριβώς έχει σημειωθεί παραπάνω (πρώτο σε επίδοση πάντα οι HOG/HOF έπειτα οι HOG και τέλος οι HOF).