

”Εθνικό Μετσόβιο Πολυτεχνείο”

Σχολή Ηλεκτρολόγων Μηχανικών  
&  
Μηχανικών Η/Υ

## ΎΡΑΣΗ ΥΠΟΛΟΓΙΣΤΩΝ

---



8ο Εξάμηνο

Ροή Σ

### 3<sup>η</sup> Εργαστηριακή Άσκηση

*Εντοπισμός Χωρο-χρονικών Σημείων Ενδιαφέροντος και Εξαγωγή Χαρακτηριστικών σε  
Βίντεο Ανθρωπίνων Δράσεων*

*Αθανασίου Νικόλαος*

*03112074*

*Ματακιάς Δημήτρης*

*03110142*

*12/7/2016*

# Μέρος 1

## Χωρο-χρονικά Σημεία Ενδιαφέροντος

Στο συγκεκριμένο μέρος της άσκησης καλούμαστε αρχικά να υπολογίσουμε κάποιους ανιχνευτές συγκεκριμένα 2 ώστε αρχικά να εξάγουμε μία έξοδο με τη βοήθεια τους και της εικόνας την οποία τελικά θα χρησιμοποιήσουμε για την εξαγωγή σημείων ενδιαφέροντος από διαδοχικά frames όπως μεγάλες μεταβολές ή σύνθετη κίνηση με βάση κάποιο κριτήριο σημαντικότητας διαφορετικό για κάθε μέθοδο.

### 1.1 Υλοποίηση 3D ανιχνευτή Harris με επέκταση του Harris Stephens (ανίχνευση γωνιών)

Ο ανιχνευτής γωνιών του Harris διατηρεί την κεντρική ιδέα ότι για την εύρεση γωνιών ψάχνει σημεία στα οποία οι τιμές των pixel μεταβάλλονται πάνω από ένα κατώφλι και στις δύο κατευθύνσεις έτσι στον 3D ψάχνει για μεταβολές και στις 3 διαστάσεις με τη διαφορά ότι η αναζήτηση τέτοιων σημείων στο χρόνο αναζητούνται σε διαφορετικές κλίμακες από ότι στον χώρο. Με βάση αυτή τη λογική συνελίσσουμε τον πίνακα με τις παραγώγους (γινόμενα πρωτων παραγώγων) με τον ανισοτροπικό πυρήνα  $g$  ώστε με βάση τις ιδιοτιμές του πίνακα  $M$  που δημιουργούμε σύμφωνα με την (1) να αντιληφθούμε τη μεταβολή του βίντεο σε 3 διευθύνσεις.

Αρχικά καλούμαστε να υλοποιήσουμε των ανιχνευτή Harris σε 3 διαστάσεις υπολογίζοντας για κάθε voxel των  $3 \times 3$  πίνακα:

$$M(x, y, t; \sigma, \tau) = g(x, y, t; \sigma, \tau) * (\nabla L(x, y, t; \sigma, \tau) (\nabla L(x, y, t; \sigma, \tau))^T) \rightarrow$$

$$M(x, y, t; \sigma, \tau) = g(x, y, t; \sigma, \tau) * \begin{bmatrix} L_x^2 & L_x L_y & L_x L_t \\ L_x L_y & L_y^2 & L_y L_t \\ L_x L_t & L_y L_t & L_t^2 \end{bmatrix} \quad (1)$$

Η τελική μορφή του τύπου 1 μας δίνει εύκολα την οριζουσα του πίνακα ποθ θα χρησιμοποιηθεί στο κριτήριο γωνιότητας.

Όπου :

$g(x, y, t; \sigma, \tau)$ : Gaussian πυρήνας ομαλοποίησης

$\nabla L(x, y, t; \sigma, \tau)$ : Οι χωρο-χρονικές παράγωγοι (δηλαδή ως προς  $\sigma$  και  $\tau$ ) με βάση υπολογισμού τον πυρήνα  $[-1 \ 0 \ 1]^T$ . Με κριτήριο γωνιότητας αντίστοιχα

Από την (1) βρίσκουμε την οριζουσα του πίνακα  $M$  να προκύπτει εύκολα:

$$|M| = J_{xx} * J_{yy} * J_{tt} + 2 * J_{xy} * J_{xt} * J_{yt} - J_{xx} * J_{yt}^2 - J_{yy} * J_{xt}^2 - J_{tt} * J_{xy}^2$$

Ωστε σύμφωνα με την (2) να υπολογιστεί εύκολα το κριτήριο γωνιότητας.

$$H(x, y, t) = \det(M(x, y, t)) - k \cdot \text{trace}^3(M(x, y, t)) \quad (2)$$

Στη συνέχεια εφαρμόζουμε kernel σφαίρα λόγω του ότι δουλεύουμε στις 3 διαστάσεις και βρίσκουμε τα επιθυμητά μέγιστα της H χρησιμοποιώντας κατώφλι 0.005 και τις θέσεις τους με τη βοήθεια της συνάρτησης του Matlab *ind2sub()*.

### 1.2 Υλοποίηση ανιχνευτή Gabor

Με αντίστοιχη λογική στο παρόν ερώτημα υπολογίζουμε τον ανιχνευτή Gabor αρχικά εξομαλύνοντας την εικόνα με κατάλληλο πυρήνα  $g$  και στη συνέχεια φιλτράροντας αρχικά την εικόνα με το ζεύγος (άρτιο και περιττό) των παρακάτω φίλτρων των οποίων η κρουστική απόκριση υπολογίζεται σε διάστημα  $[-2\tau, 2\tau]$  και κανονικοποιείται με την L1 νόρμα

$$h_{ev}(t; \tau, \omega) = -\cos(2\pi t\omega) e^{-t^2/2\tau^2} \quad (3)$$

$$h_{od}(t; \tau, \omega) = -\sin(2\pi t\omega) e^{-t^2/2\tau^2} \quad (4)$$

Ως κριτήριο σημαντικότητας προκύπτει η τετραγωνική ενέργεια του ζεύγους των Gabor συνελιγμένων με την εξομαλυμένη εικόνα σύμφωνα με την (5):

$$H(x, y, t) = (I(x, y, t) * g * h_{ev})^2 + (I(x, y, t) * g * h_{od})^2 \quad (5)$$

Από την παραπάνω διατύπωση του ανιχνευτή με το ζεύγος των gabor φίλτρων παρατηρούμε ότι ανιχνεύονται οι περιοδικές στη συχνότητα κινήσεις οι οποίες παράγουν και τη τις μεγαλύτερες αποκρίσεις παρ'όλα αυτά ανιχνεύονται και οι σύνθετες κινήσεις όχι απαραίτητα περιοδικές. Επιπλέον διαπιστώνεται ότι η μεταφορική κίνηση δεν επιφέρει μεγάλες αποκρίσεις αφού δημιουργούνται ουσιαστικά ακμές και οι αλλαγές είναι «αργές» για δεδομένες χωρικές γειτονιές. Τέλος, όπως περιμένουμε από τον ανιχνευτή δεν παράγουν μεγάλες τιμές σημαντικότητας περιοχές με έλλειψη κίνησης.

### 1.3 Υπολογισμός σημείων ενδιαφέροντος με βάση κριτήριο σημαντικότητας

Στη συνέχεια παρατίθεται απεικόνιση των σημείων ενδιαφέροντος στην περίπτωση του Harris αλλά και του Gabor για διάφορες χρονικές και χωρικές κλίμακες για τον Harris με κατώφλι ίσο με 0.00048 και  $k=0.0047$  ενώ στην περίπτωση του Gabor με κατώφλι ίσο με 0.047. Να σημειωθεί ότι οι ιδανικότερες τιμές όπως θα φανεί και από τα παρακάτω σχήματα είναι η χωρική κλίμακα ίση με περίπου 2 (~1.85-2.15) και η χρονική για τον Harris ~0.7 ενώ για τον Gabor ~1.6.

```
default_xy_scale = 2;
default_harris_time_scale = 0.7;
default_gabor_time_scale = 1.6;
default_th_harris= 0.00048;
default_th_gabor= 0.047;

default_k = 0.0047;
default_neighbor_sphere_maxima = 2;
```

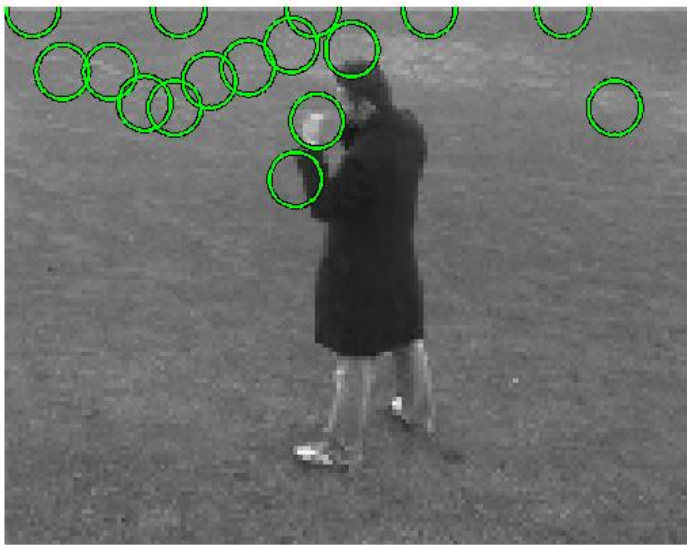
Gabor sxy=4 st=1.6 th=0.047 b=2



Gabor sxy=4 st=1.6 th=0.047 b=2



Gabor sxy=2 st=2.5 th=0.047 b=2



Gabor sxy=2 st=2.5 th=0.047 b=2



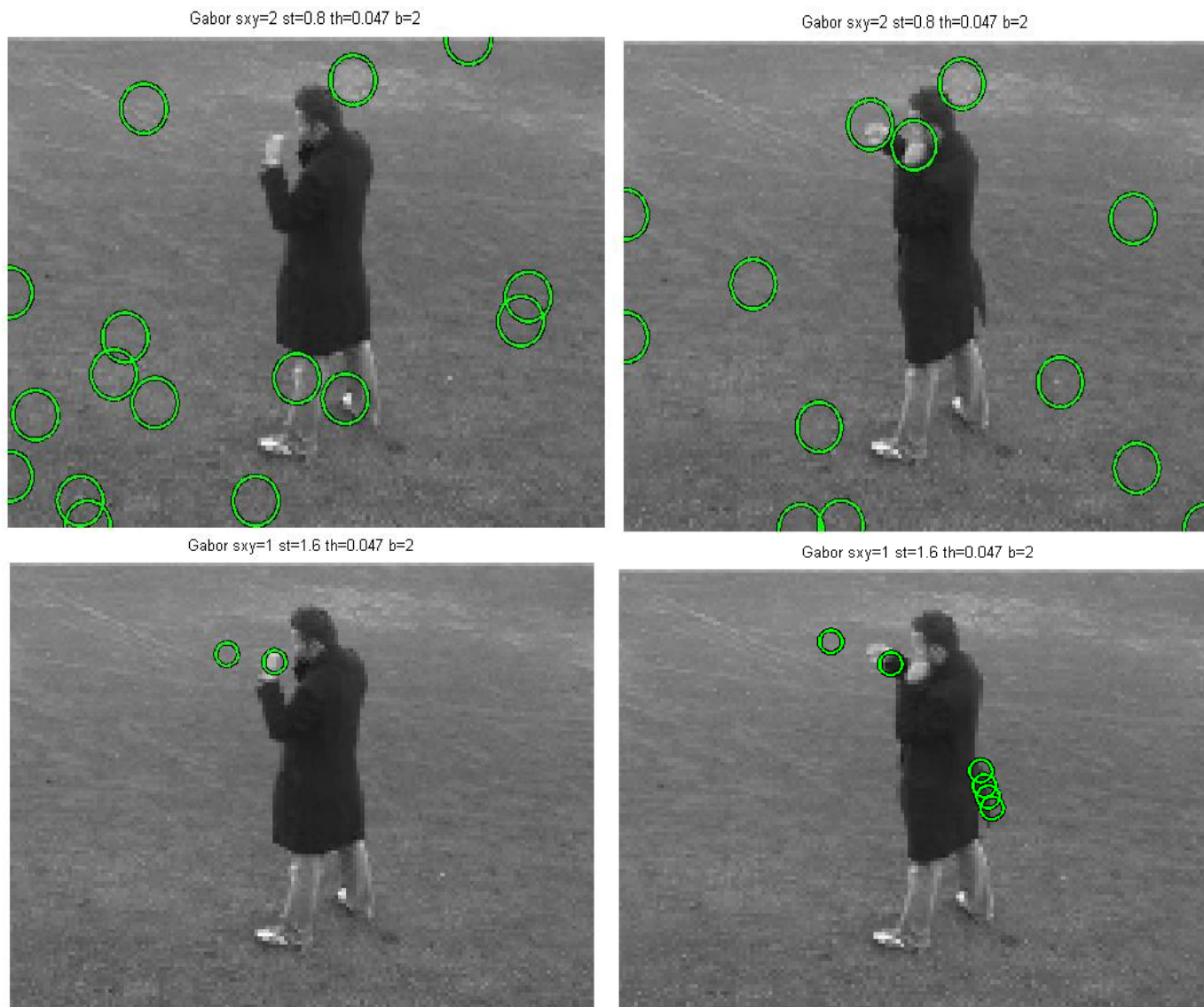
Gabor sxy=2 st=1.6 th=0.047 b=2



Gabor sxy=2 st=1.6 th=0.047 b=2



**Σχήμα 1.** Απεικόνιση σημείων ενδιαφέροντος της μεθόδου Gabor για μεγάλες τιμές χωρικής(πάνω ζεύγος frames)και χρονική κλίμακας (κάτω ζεύγος frames) και για μεσαίες-καλές τιμές(ενδιάμεσο ζεύγος frames)(χωρ.=4,2(«ιδανική τιμή») χρ.=1.6(«ιδανική τιμή»),2.5 κατώφλι=0.047 και κλίμακα 2) στην περίπτωση του boxing.



**Σχήμα 2.** Απεικόνιση σημείων ενδιαφέροντος της μεθόδου Gabor για μικρές τιμές χωρικής-δηλαδή 2-(πάνω ζεύγος frames) και χρονική κλίμακας-δηλαδή 1-(κάτω ζεύγος frames) στην περίπτωση του boxing.



**Σχήμα 3.** Απεικόνιση σημείων ενδιαφέροντος της μεθόδου Harris για μεγάλες τιμές χρονικής κλίμακας-δηλαδή 2-στην περίπτωση του boxing.



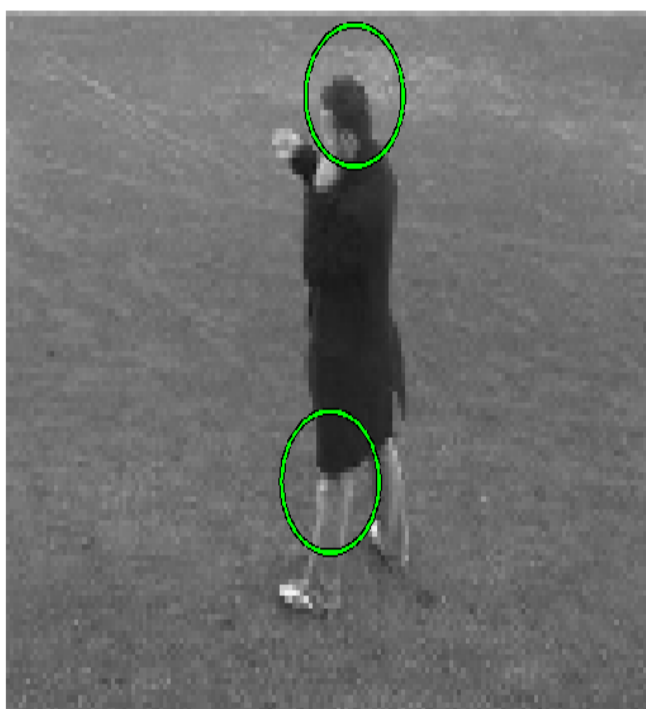
Harris sxy=2 st=0.7 th=0.00048 b=2



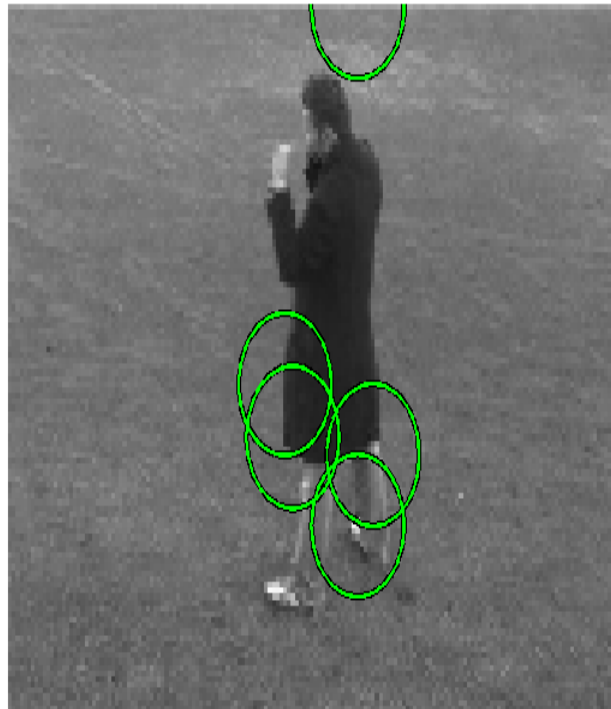
Harris sxy=2 st=0.7 th=0.00048 b=2



Harris sxy=4 st=0.7 th=0.00048 b=2



Harris sxy=4 st=0.7 th=0.00048 b=2



**Σχήμα 4.** Απεικόνιση σημείων ενδιαφέροντος της μεθόδου Harris για μεγάλες τιμές χωρικής(κάτω ζεύγος frames) για μεσαίες-καλές τιμές(πάνω ζεύγος frames) και πάλι στην περίπτωση του boxing.

Harris sxy=2 st=0.4 th=0.00048 b=2



Harris sxy=2 st=0.4 th=0.00048 b=2



Harris sxy=1 st=0.7 th=0.00048 b=2

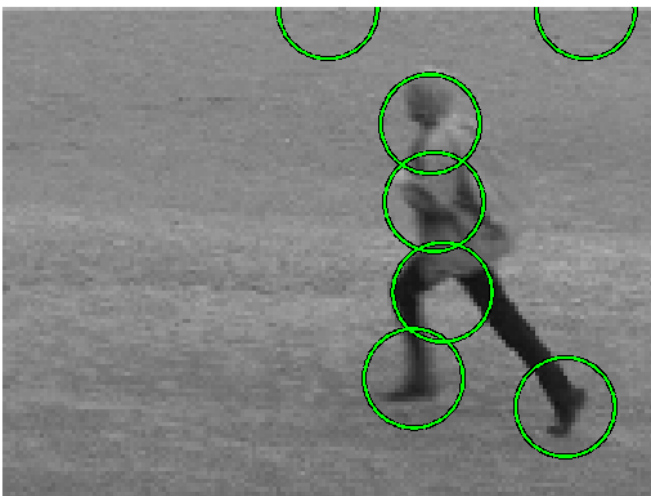


Harris sxy=1 st=0.7 th=0.00048 b=2

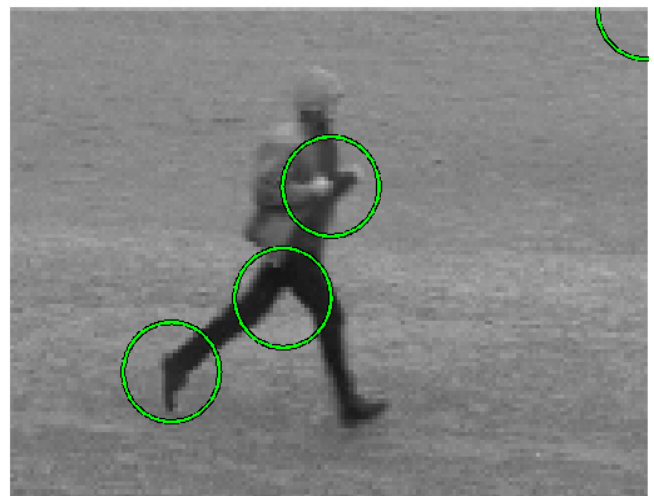


**Σχήμα 5.** Απεικόνιση σημείων ενδιαφέροντος της μεθόδου Harris για μικρές τιμές χωρικής(κάτω ζεύγος frames)και χρονική κλίμακας (πάνω ζεύγος frames).

Harris sxy=4 st=0.7 th=0.00048 b=2

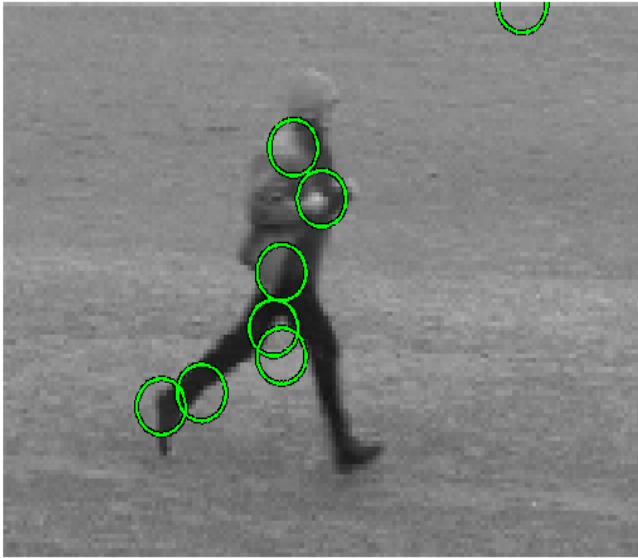


Harris sxy=4 st=0.7 th=0.00048 b=2



**Σχήμα 6.** Απεικόνιση σημείων ενδιαφέροντος της μεθόδου Harris για μεγάλες τιμές χωρικής στην περίπτωση του running.

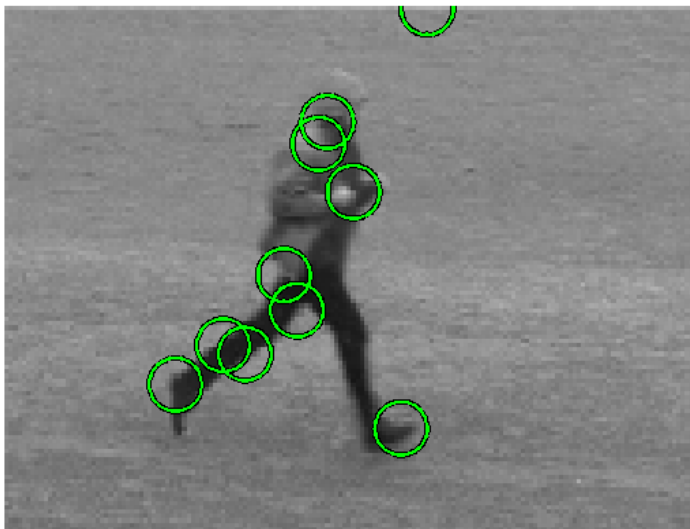
Harris sxy=2 st=2 th=0.00048 b=2



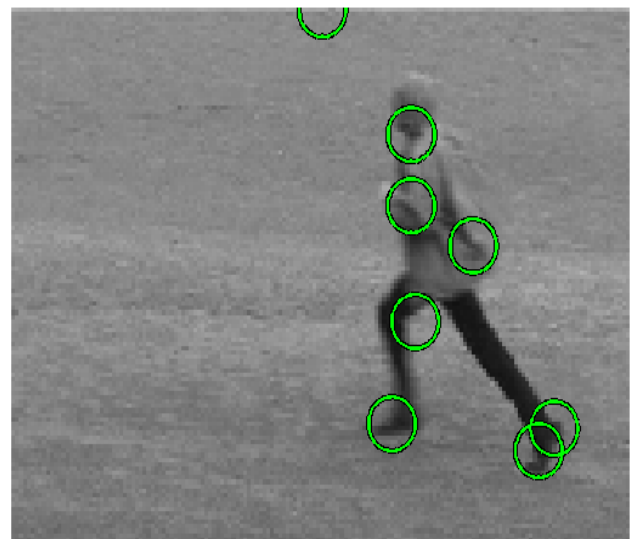
Harris sxy=2 st=2 th=0.00048 b=2



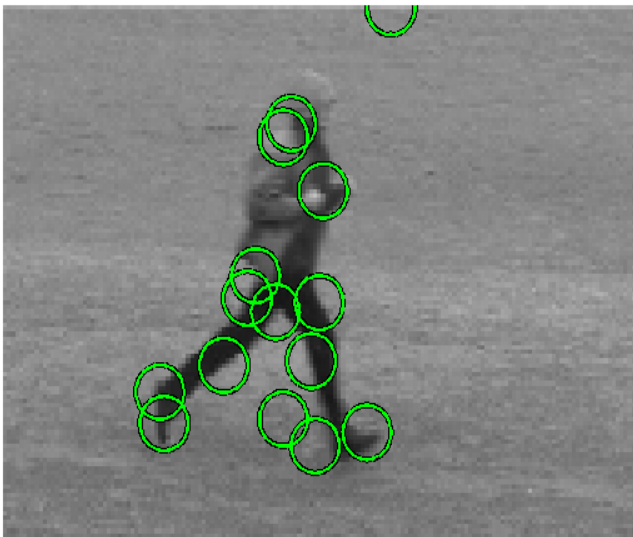
Harris sxy=2 st=0.7 th=0.00048 b=2



Harris sxy=2 st=0.7 th=0.00048 b=2



Harris sxy=2 st=0.4 th=0.00048 b=2

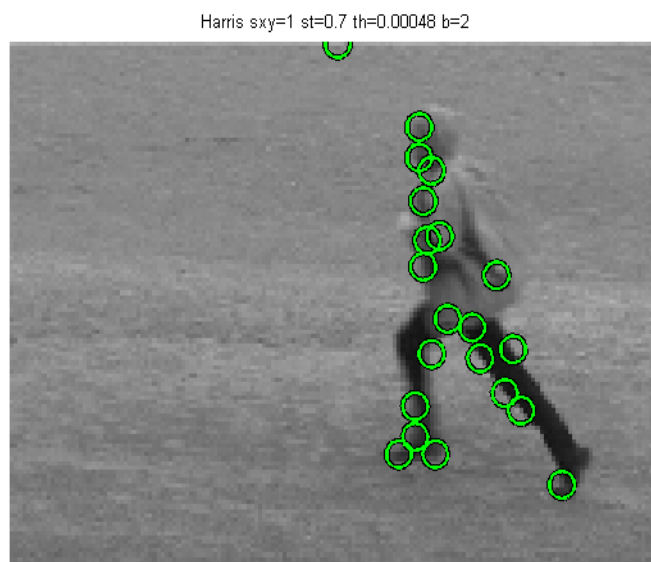
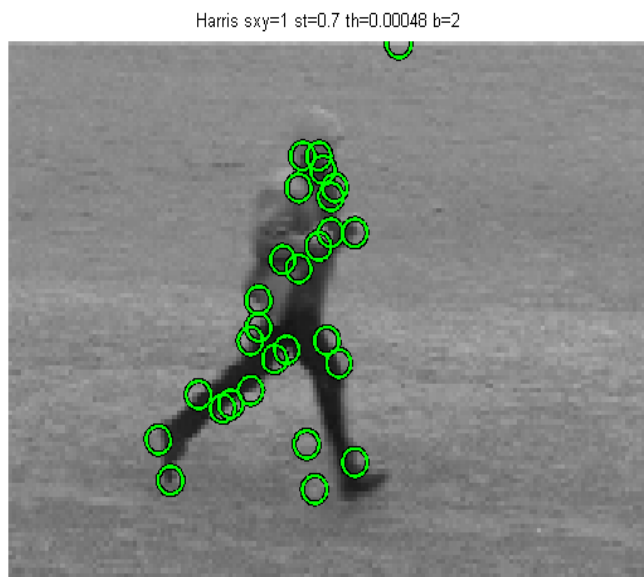


Harris sxy=2 st=0.4 th=0.00048 b=2

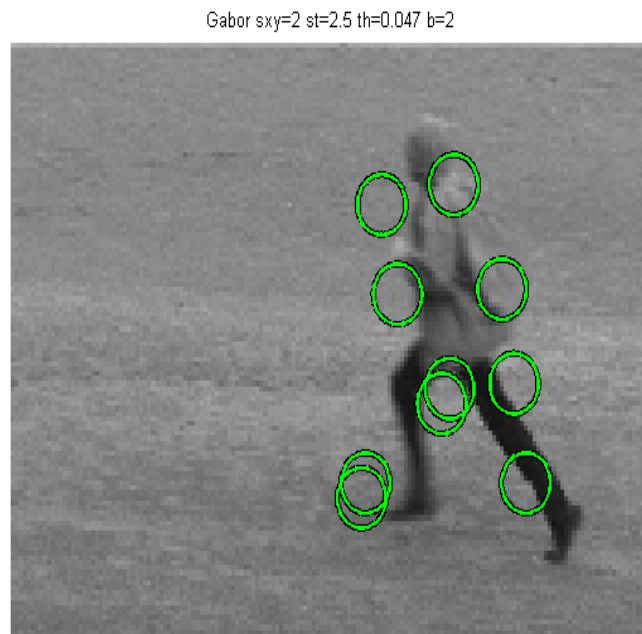
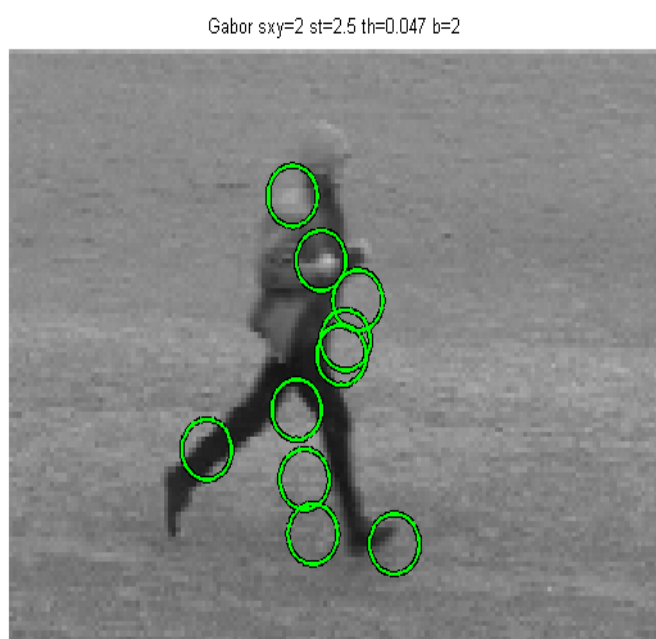
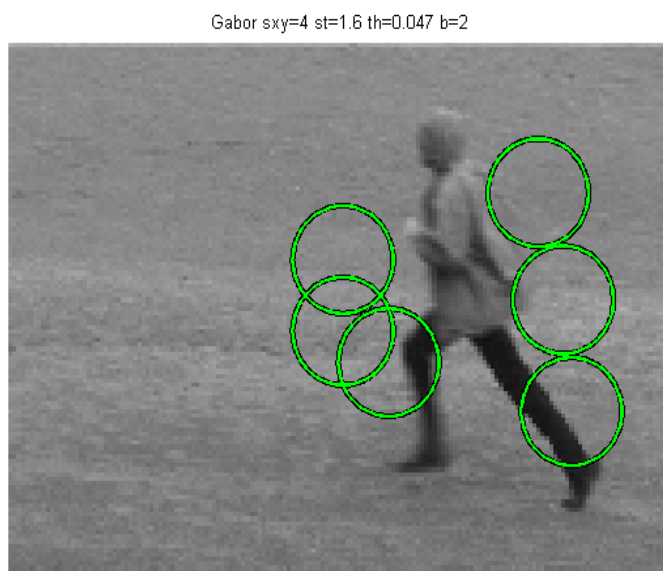


**Σχήμα 7.** Απεικόνιση σημείων ενδιαφέροντος(running) της μεθόδου Harris για μεγάλες τιμές χρονικής κλίμακας (πάνω ζεύγος frames), για μικρές τιμές χρονικής κλίμακας(κάτω ζεύγος) για μεσαίες-καλές τιμές(ενδιάμεσο ζεύγος frames).



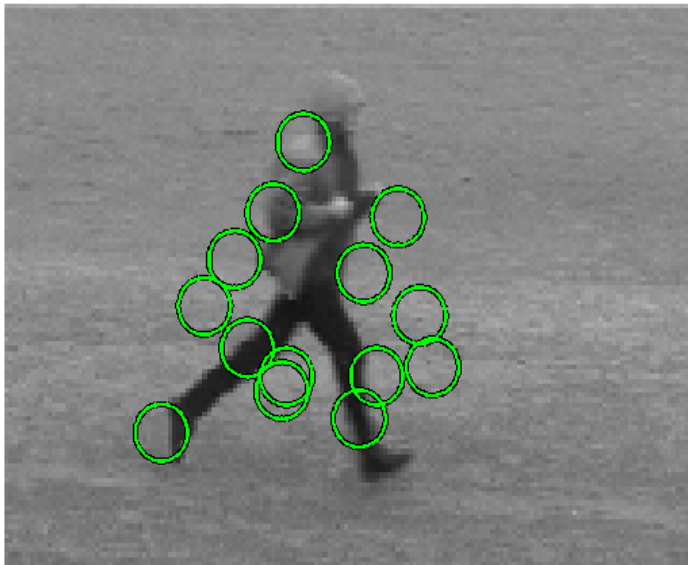


**Σχήμα 8.** Απεικόνιση σημείων ενδιαφέροντος της μεθόδου Harris για μικρή χωρική κλίμακα(running).



**Σχήμα 9.** Απεικόνιση σημείων ενδιαφέροντος της μεθόδου Gabor για μεγάλη χωρική κλίμακα (πάνω ζεύγος frames) και χρονική κλίμακας (κάτω ζεύγος frames).(running)

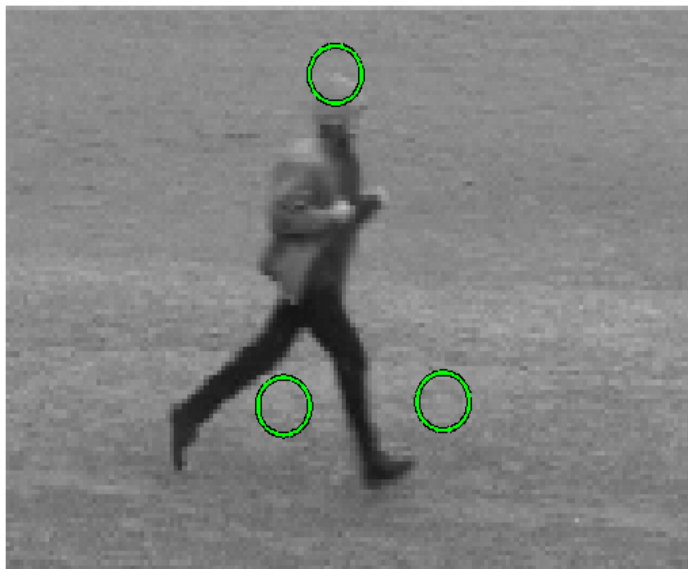
Gabor  $sxy=2$   $st=1.6$   $th=0.047$   $b=2$



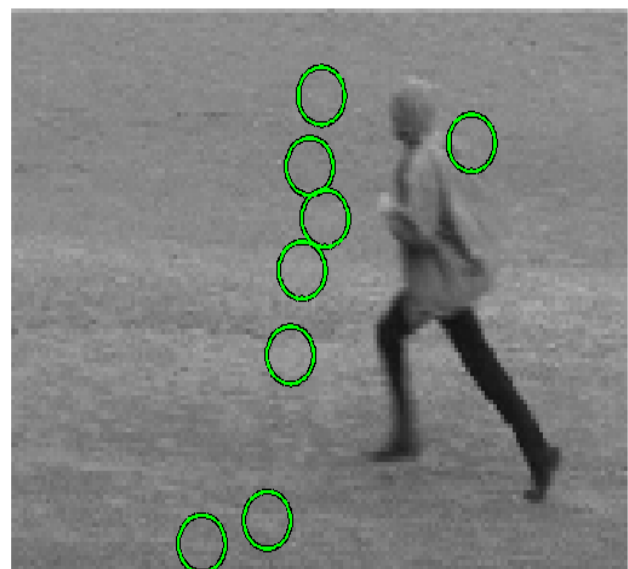
Gabor  $sxy=2$   $st=1.6$   $th=0.047$   $b=2$



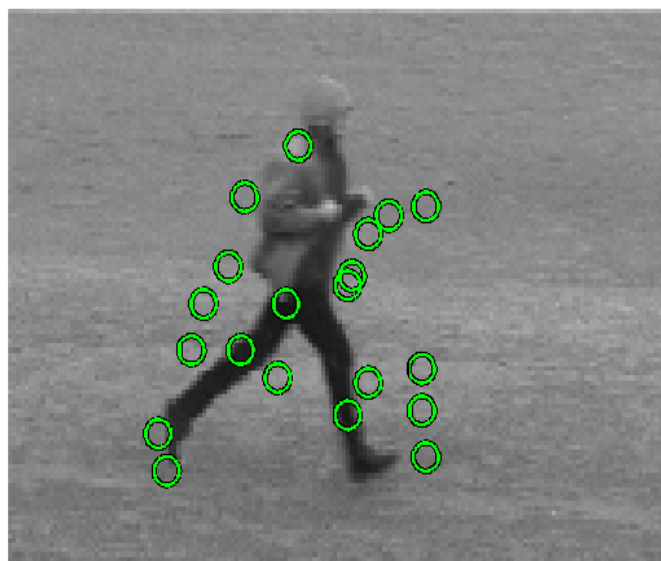
Gabor  $sxy=2$   $st=0.8$   $th=0.047$   $b=2$



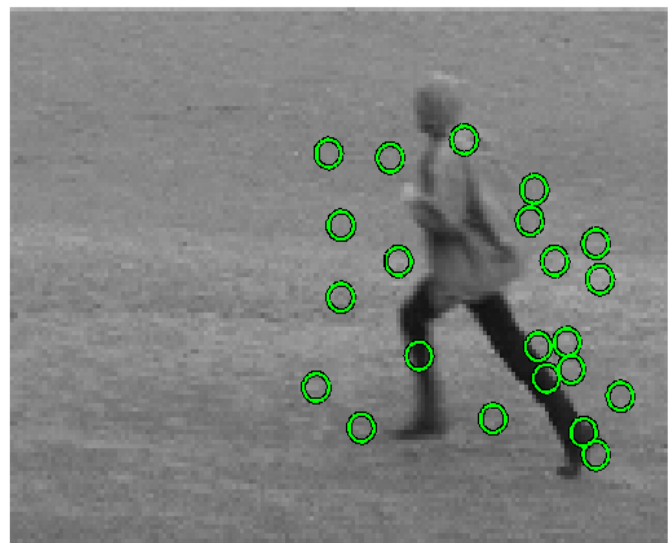
Gabor  $sxy=2$   $st=0.8$   $th=0.047$   $b=2$



Gabor  $sxy=1$   $st=1.6$   $th=0.047$   $b=2$

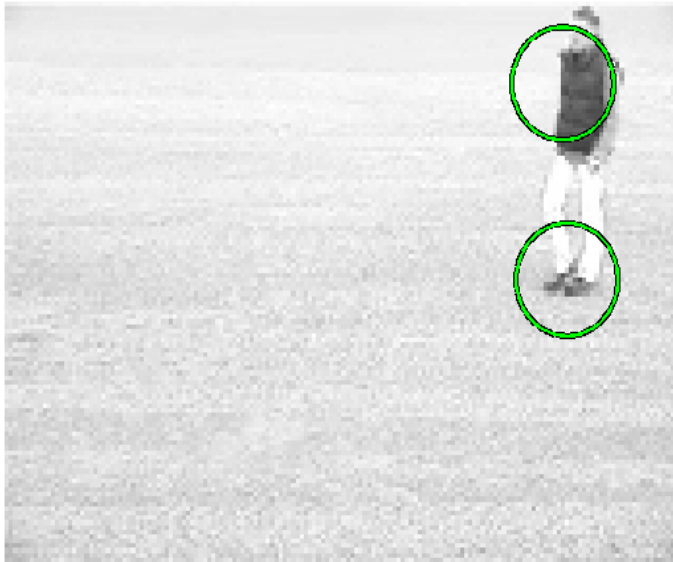


Gabor  $sxy=1$   $st=1.6$   $th=0.047$   $b=2$



**Σχήμα 10.** Απεικόνιση σημείων ενδιαφέροντος της μεθόδου Gabor για μικρές τιμές χωρικής(κάτω ζεύγος frames)και χρονική κλίμακας (ενδιάμεσο ζεύγος frames) και για μεσαίες-καλές τιμές(πάνω ζεύγος frames).(running)

Harris sxy=4 st=0.7 th=0.00048 b=2



Harris sxy=4 st=0.7 th=0.00048 b=2



Harris sxy=2 st=2 th=0.00048 b=2



Harris sxy=2 st=2 th=0.00048 b=2



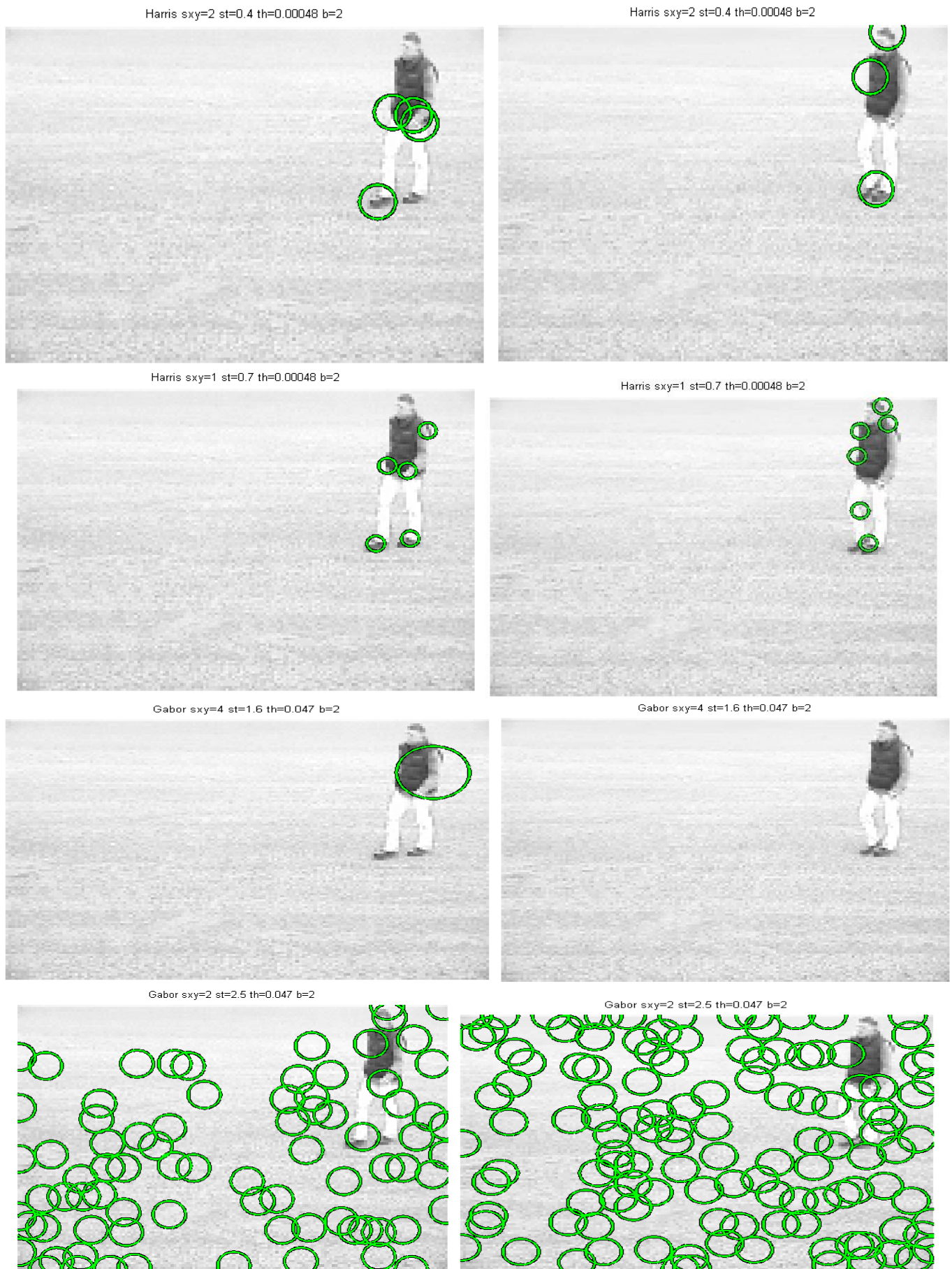
Harris sxy=2 st=0.7 th=0.00048 b=2



Harris sxy=2 st=0.7 th=0.00048 b=2

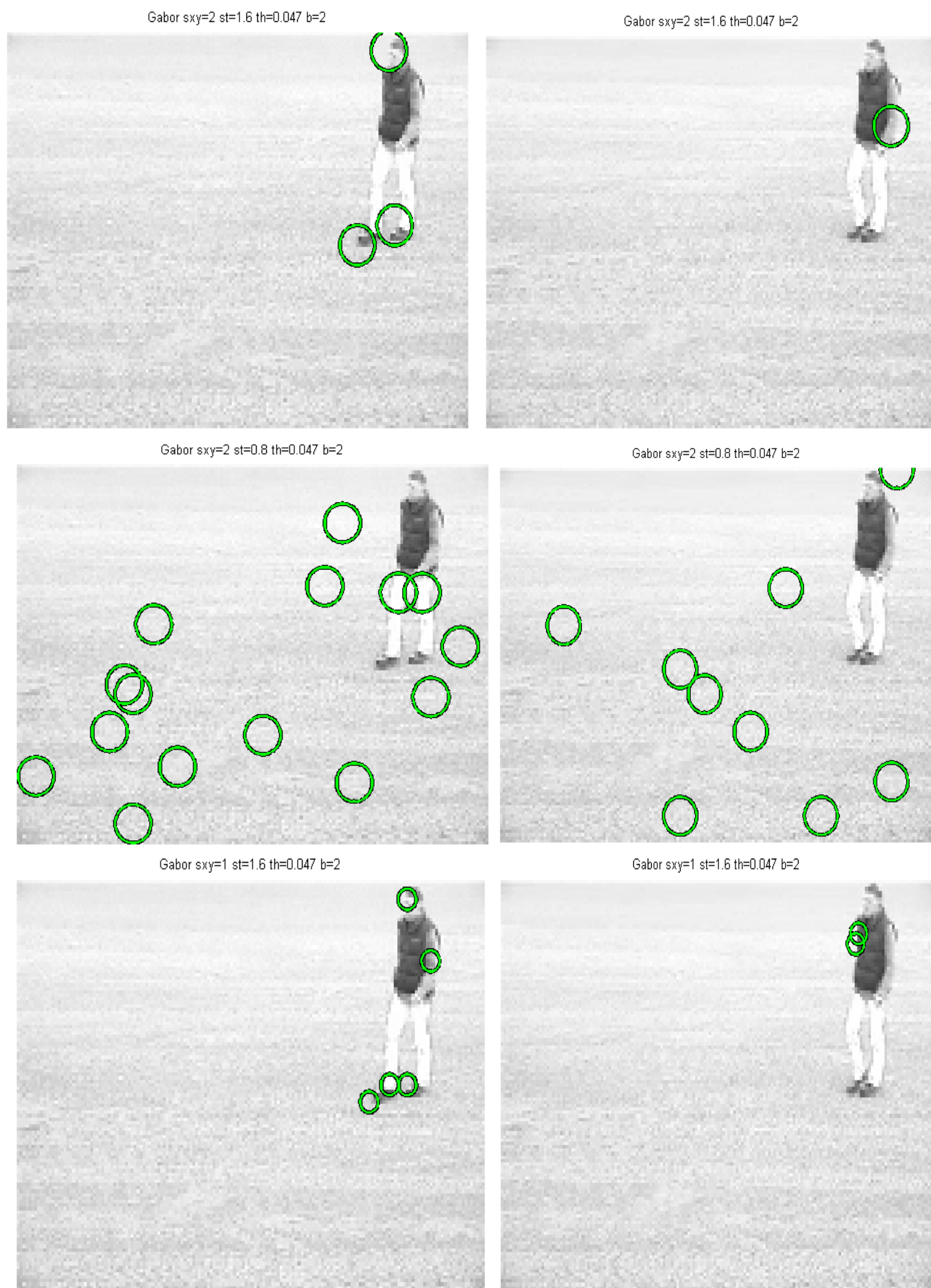


**Σχήμα 11.** Απεικόνιση σημείων ενδιαφέροντος της μεθόδου Harris για μεγάλες τιμές χωρικής(πάνω ζεύγος frames)και χρονική κλίμακας (κάτω ζεύγος frames) και για μεσαίες-καλές τιμές(ενδιάμεσο ζεύγος frames).(walking)



**Σχήμα 12.** Απεικόνιση σημείων ενδιαφέροντος της μεθόδου Harris για μικρές τιμές χωρικής(2<sup>ο</sup> ζεύγος) και χρονικής(1<sup>ο</sup> ζεύγος) και της μεθόδου Gabor για μεγάλες τιμές χωρικής (3<sup>ο</sup> ζεύγος) και χρονικής (4<sup>ο</sup> ζεύγος).(walking)





**Σχήμα 13.** Απεικόνιση σημείων ενδιαφέροντος της μεθόδου Gabor για μικρές τιμές χρονικής (2° ζεύγος) και χωρικής(3° ζεύγος) κλίμακας και για ενδιάμεσες-ιδανικές τιμές(1° ζεύγος)(walking).

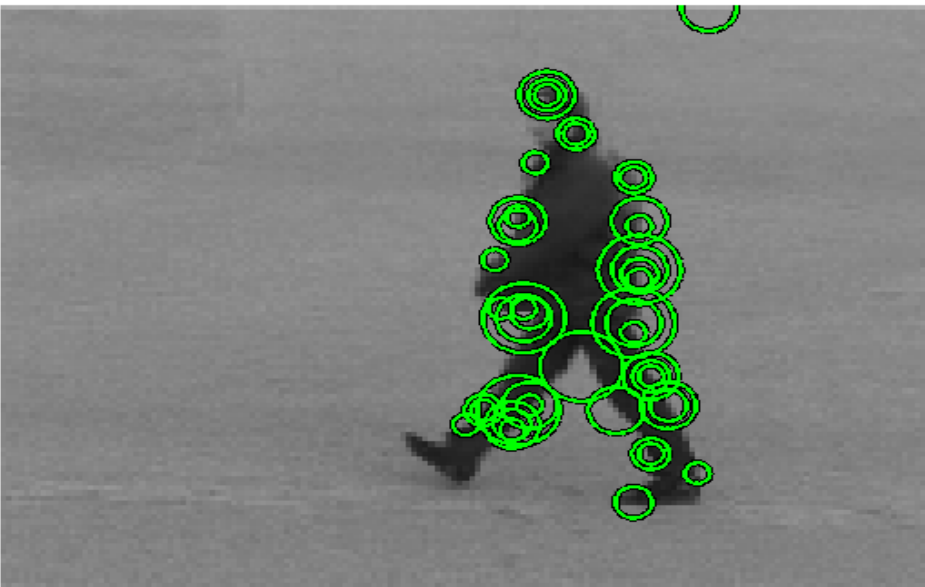
boxing



running

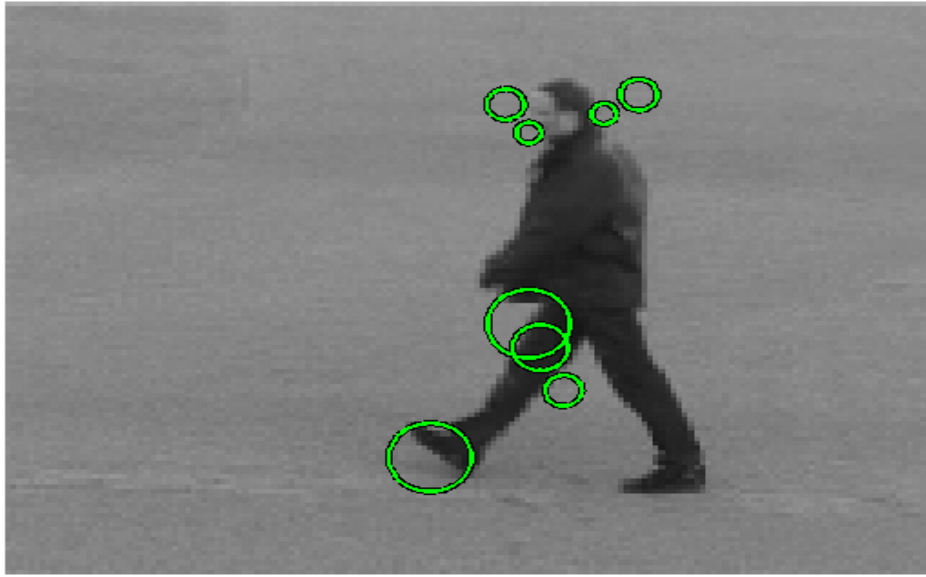


walking



**Σχήμα 14.** Απεικόνιση σημείων ενδιαφέροντος της μεθόδου Harris για πολλαπλές χωρικές κλίμακες

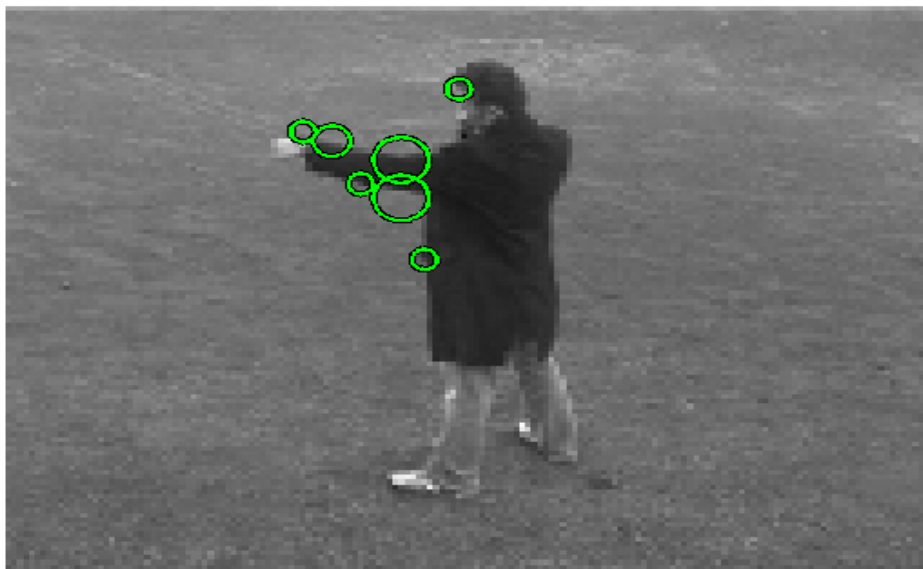
walking



running



boxing



Σχήμα 15. Απεικόνιση σημείων ενδιαφέροντος της μεθόδου Gabor για πολλαπλές χωρικές κλίμακες

# Συμπεράσματα και Παρατηρήσεις πάνω στις δύο μεθόδους και στους επιμέρους πειραματισμούς

## 1. Αλγόριθμος Gabor

- a. **Χωρική κλίμακα:** Παρατηρούμε γενικά καλά αποτελέσματα για μεγάλες τιμές της χωρικής κλίμακας, αλλά υπάρχει δυσκολία ανίχνευσης μικρών κινήσεων όπως της κίνησης της καπαρντίνας στο βίντεο με γροθιές.  
Αντιθέτως για μικρή κλίμακα οδηγούμαστε στη περίπτωση των μεγάλων κινήσεων σε ανίχνευση λανθασμένων σημείων. Η εισαγωγή αυστηρότερου κριτηρίου αφαιρεί πολλά από αυτά τα σημεία. Βέβαια βελτιώνεται η ικανότητα ανίχνευσης μικρότερων κινήσεων.
- b. **Χρονική κλίμακα:** Παρατηρούμε ότι η χρονική κλίμακα (από την σχέση (5)) βρίσκεται στον εκθέτη του αποσβενύμενου εκθετικού του φίλτρου gabor. Η μεταβολή του επιφέρει μεγάλες αλλαγές στο κατώφλι που χρησιμοποιούμε στο κριτήριο σημαντικότητας.  
Για μεγάλες τιμές της χρονικής σταθεράς παράγονται χειρότερα αποτελέσματα αφού συμπεριλαμβάνονται frames αρκετά μεταγενέστερα για την επιλογή σημείων ενδιαφέροντος. Μόνο για πολύ αργές κινήσεις θα είχε νόημα μια μεγάλη χρονική σταθερά.

## 2. Αλγόριθμος Harris

- a. **Χωρική κλίμακα:** Παρατηρούμε παρόμοια συμπεριφορά με τον αλγόριθμο Gabor. Για μικρότερες χωρικές κλίμακες απαιτείται η εισαγωγή αυστηρότερου κριτηρίου για την αποφυγή εντοπισμού λανθασμένων σημείων.
- b. **Χρονική κλίμακα:** Μεγάλες τιμές εισάγουν σφάλματα διότι περιλαμβάνεται περισσότερη πληροφορία για την επιλογή σημείων ενδιαφέροντος από παλαιότερα και μελλοντικά frame.  
Για πολύ μικρές τιμές χρονικής κλίμακας παρατηρείται ολική αδυναμία εντοπισμού κίνησης διότι δεν λαμβάνεται αρκετή πληροφορία από τα επόμενα – προηγούμενα frames με το απαιτούμενο συντελεστή ώστε να ανιχνευτεί η κίνηση.

## 3. Πολυκλιμακωτή υλοποίηση

Όπως παρατηρούμε και αναμέναμε στην περίπτωση των πολλαπλών κλιμάκων τα αποτελέσματα που λαμβάνουμε είναι σαφώς καλύτερα και μπορούν να βελτιωθούν ακόμη



περισσότερο αν πειράζουμε την παράμετρο `frame_index` στον αντίστοιχο κ.

## 4. Σύγκριση περιγραφητών

Όσον αφορά την καθολική σύγκριση των δύο περιγραφητών παρατηρούμε ότι η μέθοδος Gabor ανιχνεύει καλύτερα τις περιοδικές κινήσεις αλλά και σύνθετες κινήσεις αραιά καλά χωρίς να περιορίζεται μόνο στις δημιουργούμενες γωνίες από `frame` σε `frame`. Εκεί που αποτυγχάνει δηλαδή στη μεταφορική κίνηση με σταθερή ταχύτητα ο Harris επιτυγχάνει καλύτερο εντοπισμό.

## Μέρος 2

### *Χωρο-χρονικοί Ιστογραφικοί Περιγραφητές*

Σκοπός του δεύτερου μέρους είναι με βάση την έξοδο του πρώτου μέρους δηλαδή τα αντίστοιχα σημεία ενδιαφέροντος να υπολογίσουμε κάποιους χωρο-χρονικούς περιγραφητές (HOG,HOF).

#### 2.1 Υπολογισμός Οπτικής Ροής

Αρχικά υπολογίζουμε κατά τα γνωστά από την 2<sup>η</sup> άσκηση την οπτική ροή μέσω της διαφορετικότητας της θέσης των `pixel` σε σχέση με δύο καρέ το παρόν και το προηγούμενο και απεικονίζουμε μερικά αποτελέσματα

#### 2.2 Υπολογισμός Ιστογραφικών περιγραφητών HOG HOF

Για τον υπολογισμό των ιστογραφικών περιγραφητών όλων των `video` ακολουθούμε τα παρακάτω βήματα:

- Ανάγνωση `filenames` με την αναδρομική συνάρτηση `getAllFiles`
- Ανάγνωση αρχείων βίντεο και αποθήκευση τους σε αρχεία `mat` σε μορφή 3d πίνακα
- Εφαρμογή συναρτήσεων πρώτου μέρους για τον υπολογισμό σημείων ενδιαφέροντος
- Εκτέλεση συνάρτησης `CalculateDescriptors` η οποία είναι υπεύθυνη για την επιλογή κατάλληλης γειτονιάς διανυσματικού πεδίου γύρω από τα σημεία ενδιαφέροντος και τη παροχή τους στην ***OrientationHistogram.p***. Οι υπολογισμοί πραγματοποιούνται και για τους 6 συνδυασμούς περιγραφητών – μεθόδων εντοπισμού σημείων ενδιαφέροντος
  - HOG Harris
  - HOF Harris
  - HOG+HOF Harris
  - HOG Gabor
  - HOF Gabor
  - HOG+HOF Gabor

Ο περιγραφητής HOG βασίζεται στην ιδέα ότι το σχήμα και η μορφή των αντικειμένων σε μια γειτονιά μπορεί να περιγραφεί από την κατανομή των κατευθύνσεων των ακμών δηλαδή των κλίσεων

των διαφόρων Pixel της γειτονιάς. Επομένως κατά τη διακριτή επεξεργασία επικεντρωνόμαστε σε υπολογισμό της κλίσης με βάση διαφορών γειτονικών κλίσεων και την απεικόνιση της κατανομής τους με ιστογράμματα. Το σαφές πλεονέκτημα του HOG περιγραφητή είναι όπως επισημάνθηκε και στην 1<sup>η</sup> εργαστηριακή άσκηση η αμετάβλητοτητα του σε αλλαγές φωτισμού σκιάς(φωτομετρικές μεταβολές) αλλά και σε περιστροφές, μετατοπίσεις(γεωμετρικές μεταβολές) με μικρές αλλαγές στην κλίμακα και το μέγεθος του κελιού. Συμπερασματικά, διατηρεί μια σχετική αμεταβλητότητα σε φωτομετρικές και γεωμετρικές μεταβολές λόγω του υπολογισμού του σε μικρές γειτονίες.

Ο περιγραφητής HOF υλοποιείται με την ίδια λογική που υλοποιείται ο περιγραφητής HOG λαμβάνοντας σε μικρές γειτονίες(περιοχές) στην εικόνα την κατανομή της οπτικής ροής απεικονίζοντας την σε αντίστοιχα ιστογράμματα. Η ακρίβεια του και η αξία του ως περιγραφητή κυρίως εξαρτάται από τον υπολογισμό της οπτικής ροής.

Στη συνέχεια χρησιμοποιώντας την συνάρτηση ***OrientationHistogram.p*** υπολογίζουμε τους δύο περιγραφητές και η συνένωση τους μας δίνει τον HOG/HOF περιγραφητή.

### 2.3 Υπολογισμός global representation μες βάση την BoVW

Στη συνέχεια με βάση την BoVW της 1<sup>ης</sup> άσκηση υπολογίζουμε την τελική αναπαράσταση του βίντεο χωρίς διαχωρισμό δεδομένων σε data,train αλλά ένα σύνολο δεδομένων από το οποίο εξαγάγουμε αρχικά τις οπτικές λέξεις και στη συνέχεια τα ιστογράμματα εμφάνισης.

Για την κατασκευή του Bag of Visual Words ακολουθήσαμε τα παρακάτω βήματα:

- Συνένωση του συνόλου των περιγραφητών σε ένα διάνυσμα. Το παραπάνω επιτεύχθηκε με την περιστροφή των περιγραφητών ώστε κάθε γραμμή του διανύσματος να αντιστοιχεί σε περιγραφητή ενός σημείου ενδιαφέροντος και μετατροπή του cell array σε matrix. Κατόπιν επιλέχθηκε τυχαίο υποσύνολο (50% του συνολικού μεγέθους) του διανύσματος των περιγραφητών μέσω της μεθόδου *randperm*.
- Εφαρμογή του αλγορίθμου kmeans για τον υπολογισμό των κέντρων.
- Εύρεση του κέντρου με την ελάχιστη απόσταση για κάθε τοπικό περιγραφητή για το σύνολο των εικόνων train και test.
- Στη συνέχεια παρατίθεται με ψευδοκώδικα μια ειδικότερη απεικόνιση υλοποίησης του αλγορίθμου μας.

```
function[BOF_tr]=myBofVW(data_train)
%%concat features in one vector and kmeans
```

```
train_matrix = cell2mat(rot90(data_train));
use_randperm_to take random elements of matrix
[idx,C]=kmeans(data_matrix_subset, centersnumber);
for every picture
    for every pixel in the picture
        use repmat to avoid 3rd loop
        find_euclid_dist(centers meaning C matrix,pixels)
```

```

H(j) = min_dist_idx;
end
BOF_tr(i, :) = histc(H,1:c_number);
BOF_tr(i, :)=(BOF_tr(i, :)/(norm( BOF_tr(i, min_dist_idx))+eps)));
%histc and normalize with L2 norm avoid dividing by zero
end

```

## Μέρος 3

### Κατασκευή Δενδρογράμματος για τον Διαχωρισμό των Δράσεων

Σκοπός του τρίτου μέρους της άσκησης είναι ουσιαστικά η κατανόηση της κατηγοριοποίησης των 3<sup>ων</sup> βίντεο ανθρώπινων δράσεων με βάση την απόσταση των διανυσμάτων χαρακτηριστικών τους μέσω δενδρογραμμάτων που απεικονίζουν το πόσο ουσιαστικά απέχει η μία κατηγορία από την άλλη και το βαθμό στον οποίο διαχωρίζονται.

#### 3.1 Κατασκευή δενδρογράμματος αποστάσεων από BoVW ιστογράμματα

Στο συγκεκριμένο ερώτημα καλούμαστε να σχεδιάσουμε με βάση τα ιστογράμματα που εξάγαμε στο δεύτερο μέρος τα δενδρογράμματα αποστάσεων με χρήση ως μετρικής της  $\chi^2$  απόστασης η οποία δίνεται από τον τύπο (6).

$$D(H_i, H_j) = \frac{1}{2} \sum_{n=1}^K \frac{(h_{in} - h_{jn})^2}{h_{in} + h_{jn}} \quad (6)$$

Όπου:

$H_a = \{h_{a1}, h_{a1}, \dots, h_{aK}\}$  : ένα τυχαίο ιστόγραμμα που εξάγαμε αποτελούμενο από K σημεία

K: πλήθος των κέντρων του ιστογράμματος

Για την εφαρμογή της  $\chi^2$  απόστασης και το clustering των ιστογραμμάτων χρησιμοποιούμε τη συνάρτηση *linkage* του matlab η οποία κατηγοριοποιεί με βάση την απόσταση τα ιστογράμματα.

Η γραφική απεικόνιση της ικανότητας διαχωρισμού των κατηγοριών επιτυγχάνεται με την εφαρμογή της συνάρτησης *dendrogram* η οποία οπτικοποιεί την απόσταση των διανυσμάτων χαρακτηριστικών.

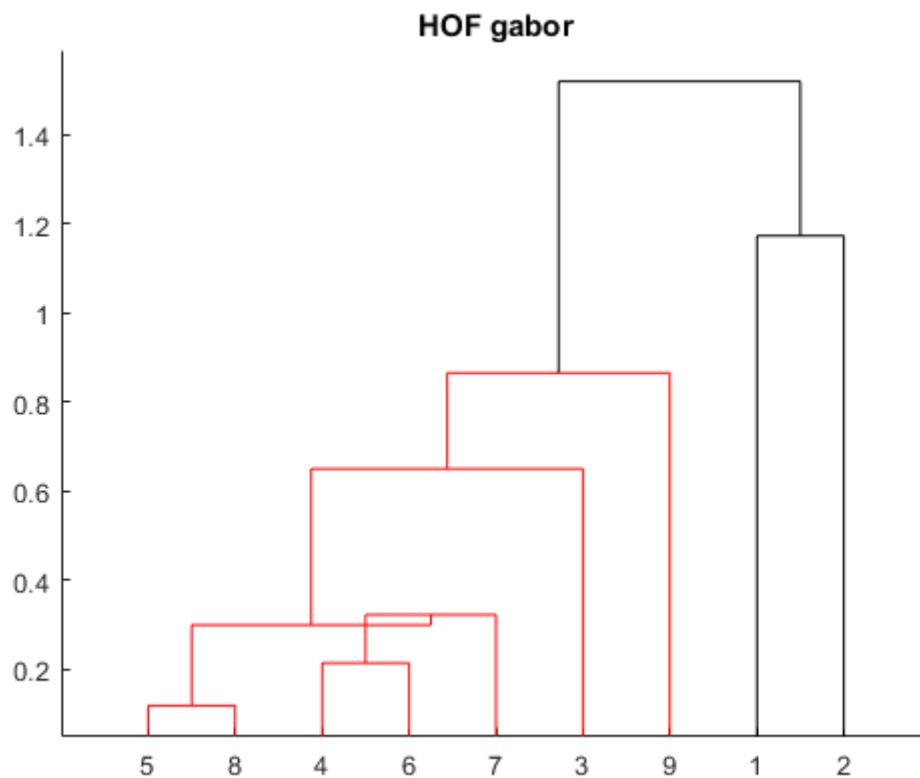
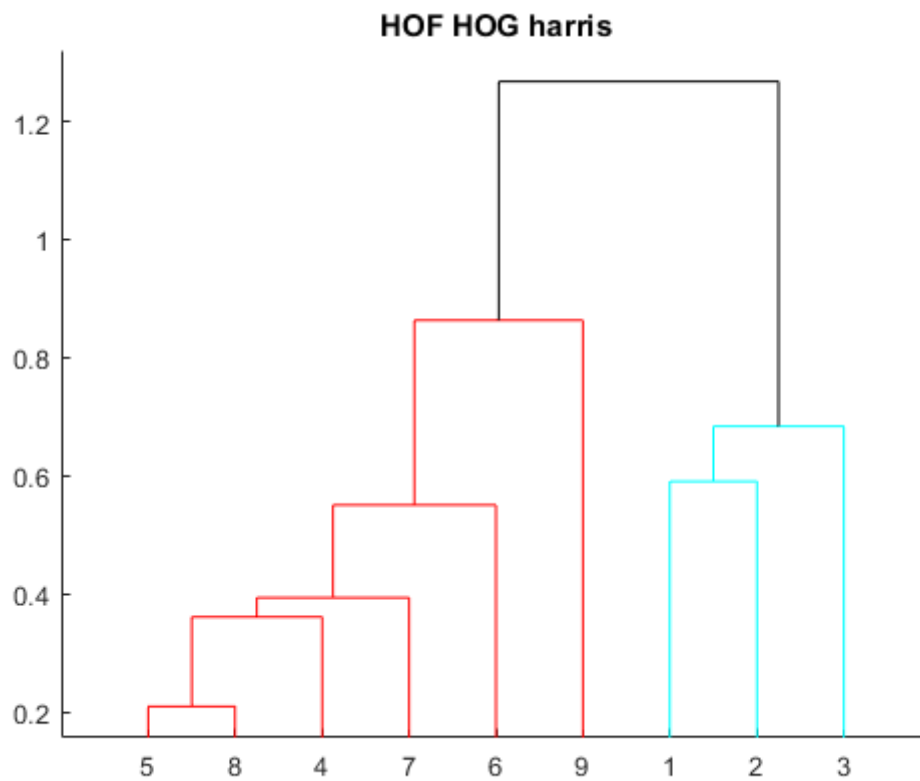
Για την παραγωγή των δενδρογραμμάτων κατασκευάστηκε η συνάρτηση *create\_dendrogram* η οποία επίσης υπολογίζει έναν δείκτη αξιολόγησης της ικανότητας διαχωρισμού των κλάσεων βασιζόμενο στη μέγιστη απόσταση εντός της ίδιας κλάσης των αντικειμένων και την ελάχιστη απόσταση μεταξύ αντικειμένων διαφορετικών κλάσεων. Προφανώς ο λόγος:

$$Rc(i) = \frac{MaxIntraDist}{MinInterDist}$$

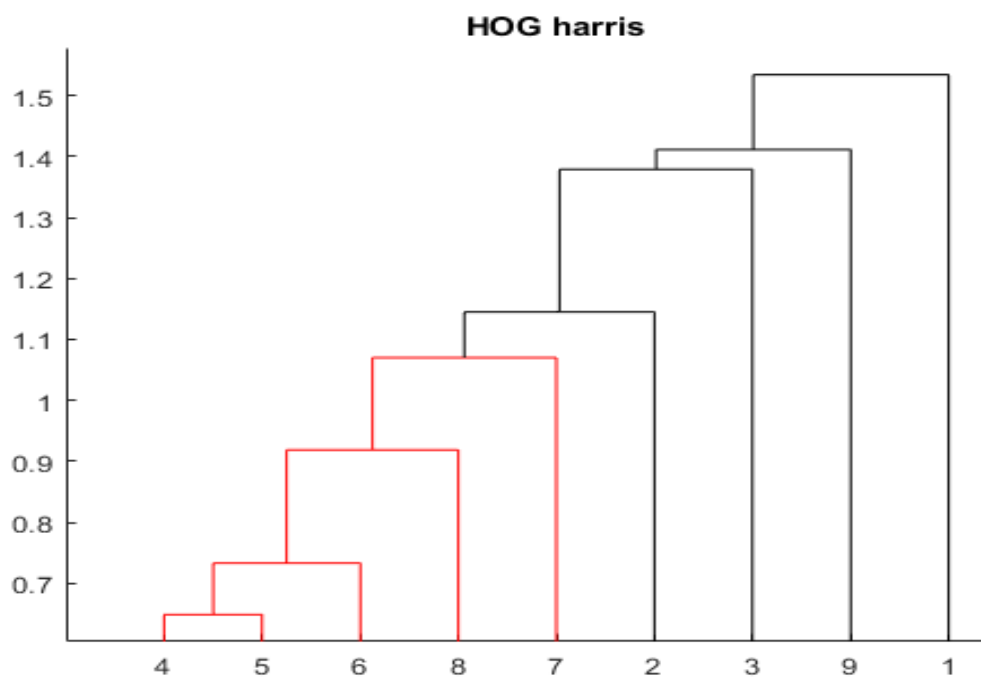
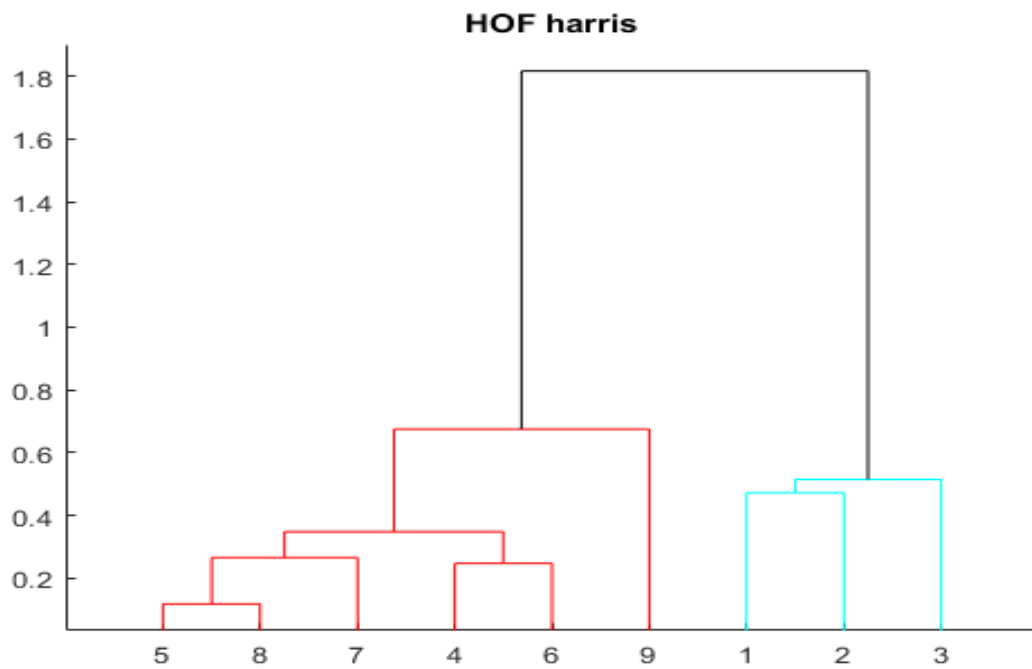
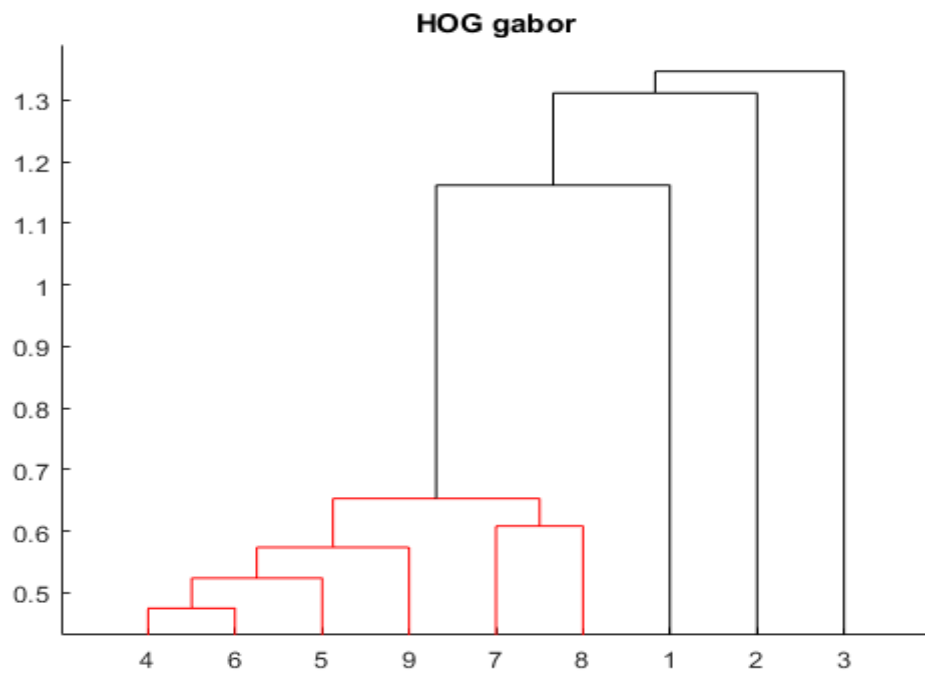
πρέπει να μεγιστοποιείται όταν πραγματοποιούμε «καλή» κατηγοριοποίηση.

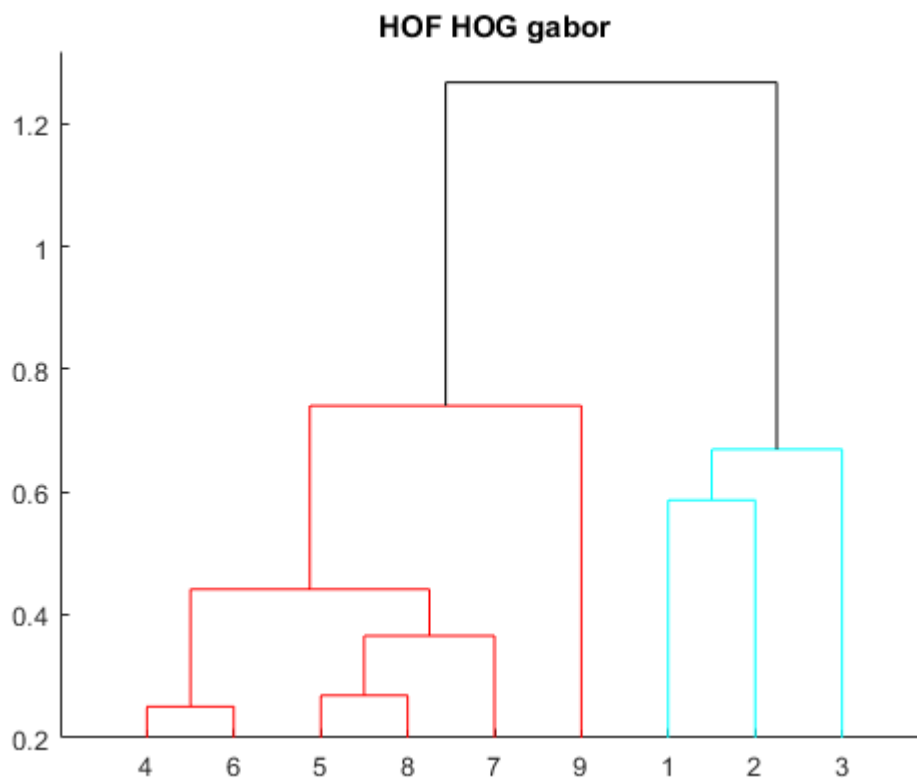
### 3.2 Πειραματισμός κατασκευής δενδρογραμμάτων για διαφορετικούς ανιχνευτές/περιγραφητές

Να τονίσουμε ότι τα video έχουν αντιστοιχεί με τα παρακάτω labels boxing(1-3 ) running(4-6) walking(7-9).









Παρατηρούμε ότι ο hog περιγραφητής σε κάθε περίπτωση δεν μας δίνει τα επιθυμητά αποτελέσματα για παράδειγμα δεν καταφέρνει να διαχωρίσει το boxing από τις άλλες 2 δραστηριότητες σε κανένα. Γεγονός που είναι λογικό αφού πληροφορία που περιέχει είναι για την κίνηση μικρή σε σχέση με τον HOF περιγραφητή και έτσι αδυνατεί να ξεχωρίσει ακόμα και προφανείς περιπτώσεις αν και η αναισθησία του σε μεταβολές περιστροφής και μετατόπισης αλλά και φωτομετρικές μεταβολές βοηθάει αλλά στην περίπτωση μας οι μεταβολές είναι κυρίως γεωμετρικές και μάλιστα μεταφορικές οπότε δεν αρκεί να στηριχτούμε σε αυτόν. Αντιθέτως ο περιγραφητής HOF μας δίνει καλά αποτελέσματα και για τον Gabor και για τον Harris –ιδιαίτερα για τον δεύτερο– σε κάποιες περιπτώσεις δηλαδή καταφέρνει να ξεχωρίσει το boxing από τις άλλες δύο δραστηριότητες είτε πλήρως είτε μερικώς πράγμα λογικό αφού το κύριο συστατικό στοιχείο του video είναι η κίνηση στην οποία δίνει έμφαση λόγω της οπτικής ροής που χρησιμοποιεί. Τα καλύτερα αποτελέσματα τα δίνουν οι HOG/HOF περιγραφητές και για τις δύο μεθόδους (Gabor, Harris) αφού ξεχωρίζουν το boxing από τις δύο άλλες κινήσεις και το walking, running κατηγοριοποιούνται όσο καλύτερα γίνεται. Οι κινήσεις walking και running είναι παραπλήσιες και επομένως δύσκολο να διαχωριστούν και να κατηγοριοποιηθούν για αυτό στις περισσότερες περιπτώσεις κατηγοριοποιούνται επιμέρους βίντεο των κλάσεων δηλαδή π.χ. video με μεγάλο / μικρό βηματισμό μπαίνουν μαζί. Συμπερασματικά, ο συνδυασμός HOG/HOF είναι ο καλύτερος και για τους δύο ανιχνευτές αφού ενσωματώνει και γεωμετρική-φωτομετρική αναισθησία αλλά και πολύ πληροφορία κίνησης εξασφαλίζοντας καλά αποτελέσματα για κάθε περίπτωση όχι μόνο για τα συγκεκριμένα video που επεξεργαστήκαμε εκτός του αν το video περιέχει χαρακτηριστικά που ανιχνεύει ένας από τους δύο αρκετά καλά οπότε για οικονομία πράξεων μετά και από τις απαραίτητες συγκρίσεις μπορούμε να εμπιστευτούμε αυτόν (πχ HOG μηδαμινή κίνηση-φωτομετρικές/γεωμετρικές μεταβολές μόνο, HOF κύριο μας ενδιαφέρον η κίνηση). Η αλλαγή μεθόδου επιλογής αποστάσεων μεταξύ των clusters κατά την εφαρμογή της συνάρτησης linkage στο matlab δεν οδήγησε σε σημαντικές βελτιώσεις.

