ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΌΡΑΣΗ ΥΠΟΛΟΓΙΣΤΩΝ

ΑΚΑΔΗΜΑΪΚΟ ΈΤΟΣ 2017-2018

ΑΝΑΦΟΡΑ 3ης ΕΡΓΑΣΤΗΡΙΑΚΗΣ ΆΣΚΗΣΗΣ

Δημητριάδης Νικόλαος
03114016
8ο Εξάμηνο

Θανάς Λεράι
03114156
8ο Εξάμηνο

Αθήνα, Ιούλιος 2018

Περιεχόμενα

1 Χωρο-χρονικά Σημεία Ενδιαφέροντος

Επεξήγηση Υλοποίησης

Η υλοποίηση αυτού του μέρους πραγματοποιείται στο αρχείο `computeDetector.m` και στα αντίστοιχα call functions.
Σημείωση για υλοποίηση: Πρόκειται για μία συνάρτηση στην οποία χρησιμοποιείται Input Parser για το πέρασμα των αντίστοιχων παραμέτρων. Απαιτούνται ο προσδιορισμός του video και της μεθόδου εύρεσης των χωρο-χρονικών σημείων ενδιαφέροντος. Οι υπόλοιπες παράμετροι δεν είναι απαραίτητες, καθώς έχουν οριστεί default τιμές. Ωστόσο, με τρόπο όμοιο με τον προσδιορισμό του περιγραφητή δύναται ο καθορισμός τους. Οι default τιμές των optional parameters είναι:

parameter	default value
sigma	1.5
tau	3
integSigma	1.5
integTau	3
thetaHarris	0.005
thetaGabor	0.01
k	0.0005

Χρησιμοποιούνται διαφορετικά θ_{corn} για κάθε μέθοδο λόγω των διαφορετικών default τιμών που επιλέχθηκαν.

1.1 Harris Detector

computeDetector(video,'Harris',···)
Στην ενότητα αυτή υλοποιήσαμε το χωροχρονικό ανιχνευτή σημείων ενδιαφέροντος Harris. Ο κώδικας για τον αλγόριθμο αυτό βρίσκεται στο αρχείο `Harris_Detector_3D.m`.
Αρχικά δημιουργήσαμε ένα 3D Γκαουσιανό φίλτρο $G_{\sigma,\tau}$ συνελίσσοντας μια 2D χωρική Γκαουσιανή με κλίμακα σ και μια 1D χρονική Γκαουσιανή με κλίμακα τ. Στη συνέχεια, φιλτράραμε το video σε 3 διαστάσεις με το προαναφερθέν φίλτρο. Υπολογίσαμε τις 2 χωρικές και τη χρονική παράγωγο του φιλτραρισμένου video χρησιμοποιώντας για την παραγώγιση τον πυρήνα κεντρικών διαφορών

$$\begin{bmatrix}-1 & 0 & 1\end{bmatrix}^T.$$

Δημιουργήσαμε επίσης ένα 3D Γκαουσιανό φίλτρο ολοκλήρωσης $G_{\rho,s}$ συνελίσσοντας μια 2D χωρική Γκαουσιανή με κλίμακα ολοκλήρωσης ρ και μια 1D χρονική Γκαουσιανή με κλίμακα ολοκλήρωσης s. Υπολογίσαμε στη συνέχεια τα στοιχεία του δομικού τανυστή **J** της μεθόδου ανίχνευσης γωνιών Harris-Stephens:

$$\mathbf{J} = \begin{bmatrix} J_{xx}(x,y,t) & J_{xy}(x,y,t) & J_{xt}(x,y,t) \\ J_{xy}(x,y,t) & J_{yy}(x,y,t) & J_{yt}(x,y,t) \\ J_{xt}(x,y,t) & J_{yt}(x,y,t) & J_{tt}(x,y,t) \end{bmatrix}$$

τα οποία προκύπτουν ως εξής:

$$J_{ij}(x,y,t) = G_{\rho,s} * \left(\frac{\partial I_{\sigma,\tau}}{\partial i} \cdot \frac{\partial I_{\sigma,\tau}}{\partial j} \right)(x,y,t)$$

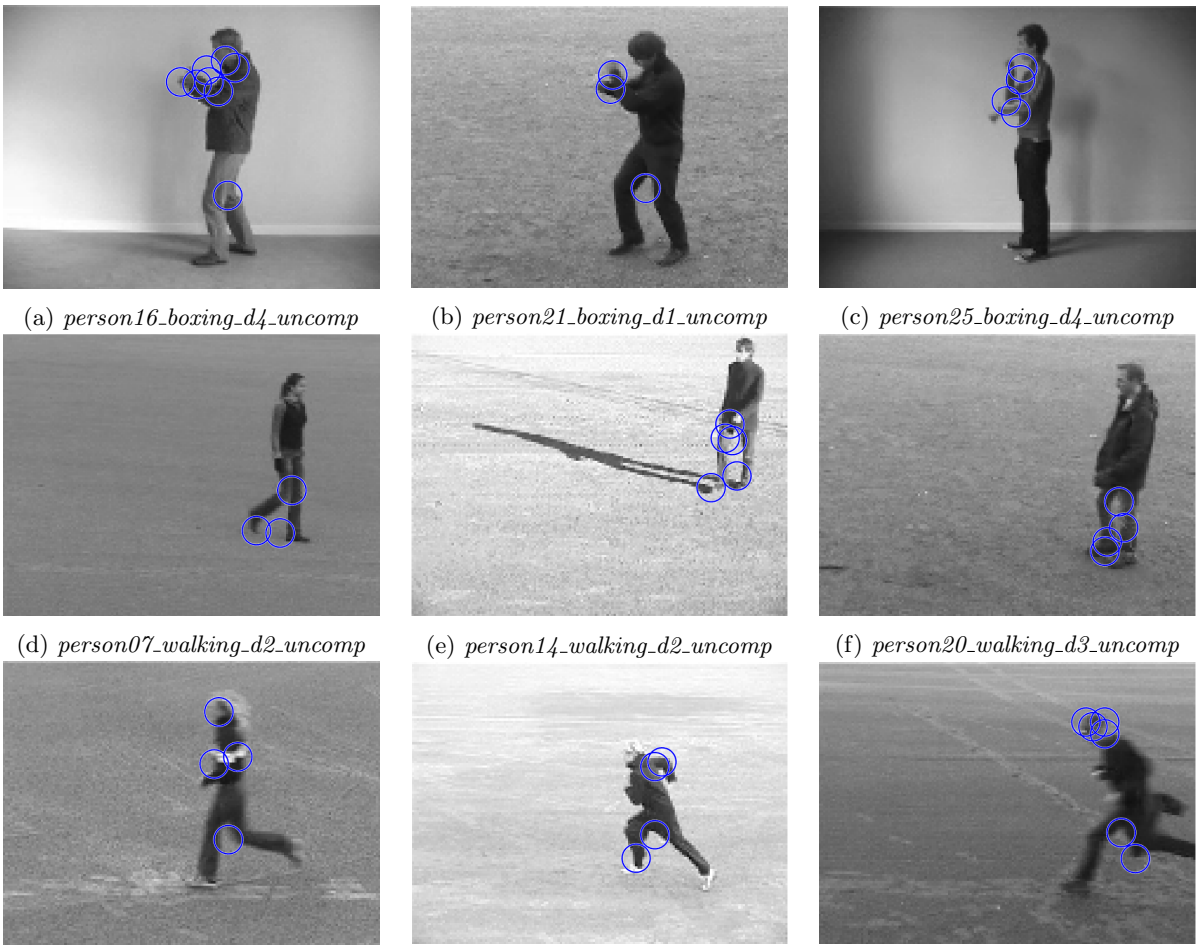
όπου

$$I_{\sigma,\tau} = G_{\rho,s} * I$$

Όλα τα παραπάνω φιλτραρίσματα $(*)$ πραγματοποιήθηκαν με τη συνάρτηση `imfilter`. Επίσης, οι συνελίξεις των 2D και 1D Γκαουσιανών για τη σύνθεση των 3D Γκαουσιανών έγινε με τη χρήση της συνάρτησης `convn`. Στη συνέχεια υπολογίζουμε το κριτήριο γωνιότητας με βάση τη σχέση

$$H(x,y,t) = det(\mathbf{J}(x,y,t)) - k \cdot trace^3(\mathbf{J}(x,y,t))$$

Ως σημεία ενδιαφέροντος επιλέχθηκαν τα σημεία της συνάρτησης $H(x,y,t)$ που είναι σημεία τοπικού μεγίστου και ικανοποιούν και τη σχέση $H(x,y,t) > \theta_{corn} H_{max}$ για κάποιο επιλεγμένο κατώφλι θ_{corn}.
Τα τοπικά μέγιστα της H βρέθηκαν με τη χρήση της συνάρτησης `imregionalmax`. Η απεικόνιση των σημείων ενδιαφέροντος πραγματοποιήθηκε με τη βοήθεια της συνάρτησης `showDetection.m`. Σημειώνουμε ότι απορρίφθηκαν ως σημεία ενδιαφέροντος εκείνα που ανιχνεύθηκαν πάνω στο περίγραμματου κάθε frame.
Παραθέτουμε κάποια αποτελέσματα του ανιχνευτή Harris 3 διαστάσεων:



(a) *person16_boxing_d4_uncomp* (b) *person21_boxing_d1_uncomp* (c) *person25_boxing_d4_uncomp*

(d) *person07_walking_d2_uncomp* (e) *person14_walking_d2_uncomp* (f) *person20_walking_d3_uncomp*

(g) *person09_running_d1_uncomp* (h) *person15_running_d1_uncomp* (i) *person23_running_d3_uncomp*

Εικόνα 1: *Σημεία ενδιαφέροντος με βάση τον ανιχνευτή Harris για όλα τα video των κατηγοριών boxing, walking, running για τιμές παραμέτρων* $\sigma = 2, \tau = 3, \rho = 2, s = 3, k = 0.005, \theta_{corn} = 0.005$

1.2 Gabor Detector

computeDetector(video,'Gabor',···)
Στην ενότητα αυτή υλοποιήσαμε τον χωροχρονικό ανιχνευτή σημείων ενδιαφέροντος Harris. Ο κώδικας για τον αλγόριθμο αυτό βρίσκεται στο αρχείο `Gabor_Detector_3D.m`.
Αρχικά δημιουργούμε ένα 2D χωρικό Γκαουσιανό φίλτρο με τυπική απόκλιση σ με τη χρήση της συνάρτησης `fspecial`. Στη συνέχεια δημιουργήσαμε ένα χρονικό παράθυρο $[-2\tau, 2\tau]$ στο οποίο ορίσαμε το ζεύγος Gabor φίλτρων h_{ev} και h_{od} ως εξής:

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

Το ζεύγος αυτό υπέστη κανονικοποίηση με την L1 νόρμα, δηλαδή οι παραπάνω συναρτήσεις διαιρέθηκαν με το άθροισμα των απολύτων τιμών τους στο (διακριτό) διάστημα $[-2\tau, 2\tau]$ οπότε προέκυψαν το Gabor ζεύγος φίλτρων:

$$\begin{cases} g_{ev}(t;\tau,\omega) = \dfrac{h_{ev}(t;\tau,\omega)}{\|h_{ev}\|_1} \\ \\ g_{od}(t;\tau,\omega) = \dfrac{h_{od}(t;\tau,\omega)}{\|h_{od}\|_1} \end{cases}$$

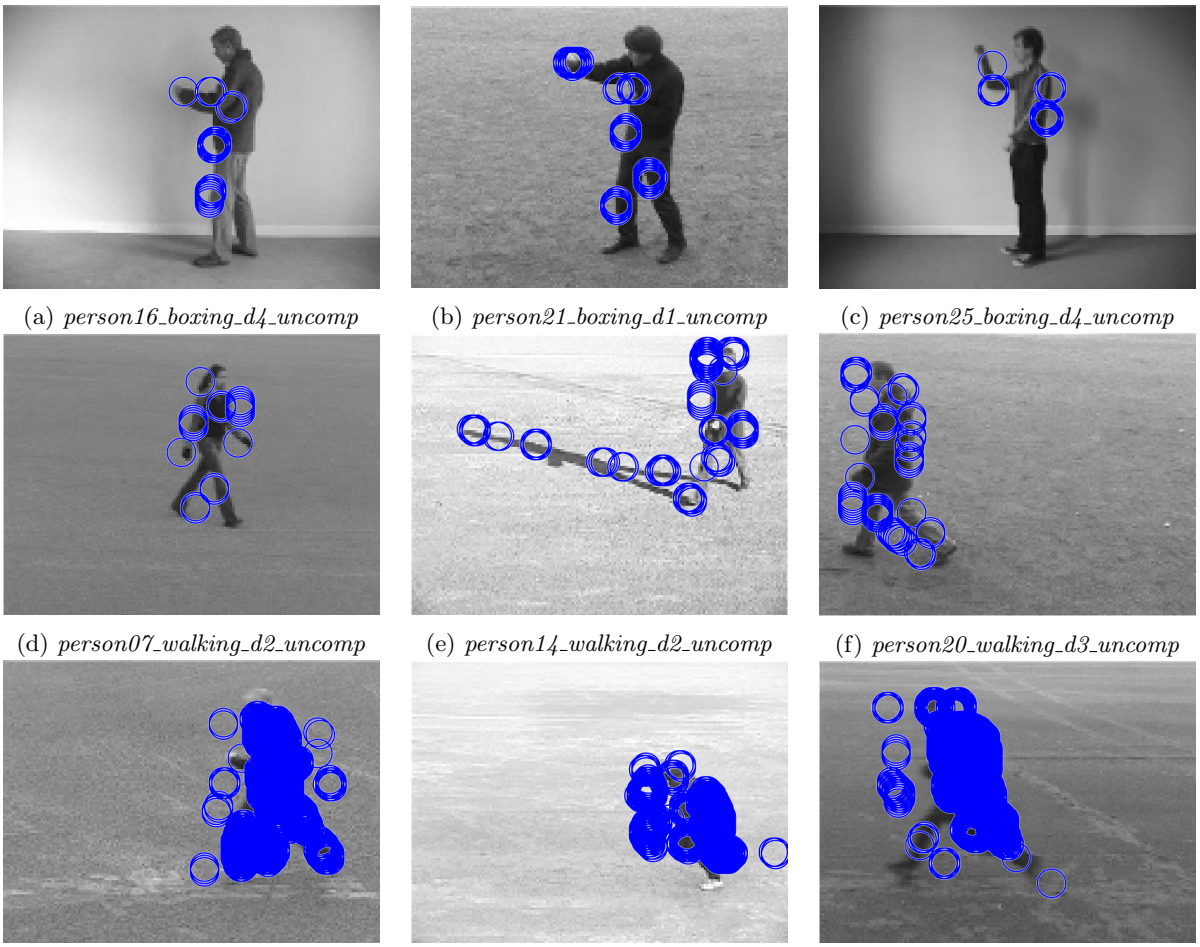
Πραγματοποιήθηκε στη συνέχεια, μέσω της συνάρτησης `convn`, η συνέλιξη των παραπάνω κανονικοποιημένων φίλτρων με το προαναφερθέν 2D χωρικό Γκαουσιανό φίλτρο για να προκύψουν το ομαλοποιημένο πλεον ζεύγος Gabor φίλτρων σύμφωνα με τις σχέσεις:

$$\begin{cases} g_{ev,normalized} = g_{ev} * G_\sigma \\ \\ g_{od,normalized} = g_{od} * G_\sigma \end{cases}$$

Φιλτράραμε το video με κάθε ένα από τα παραπάνω ομαλοποιημένα φίλτρα Gabor και προέκυψε το κριτήριο σημαντικότητας με βάση την παρακάτω σχέση:

$$H(x,y,t) = \left(I(x,y,t) * g_{ev,normalized}\right)^2 + \left(I(x,y,t) * g_{od,normalized}\right)^2$$

Τα σημεία που επιλέχθηκαν ως σημεία ενδιαφέροντος ικανοποιούν τις ίδιες προδιαγραφές όπως και στον ανιχνευτή Gabor. Παραθέτουμε κάποια αποτελέσματα του ανιχνευτή Gabor 3 διαστάσεων:



(a) *person16_boxing_d4_uncomp* (b) *person21_boxing_d1_uncomp* (c) *person25_boxing_d4_uncomp*

(d) *person07_walking_d2_uncomp* (e) *person14_walking_d2_uncomp* (f) *person20_walking_d3_uncomp*

(g) *person09_running_d1_uncomp* (h) *person15_running_d1_uncomp* (i) *person23_running_d3_uncomp*

Εικόνα 2: *Σημεία ενδιαφέροντος με βάση τον ανιχνευτή Gabor για όλα τα video των κατηγοριών boxing, walking, running για τιμές παραμέτρων* $\sigma = 2, \tau = 3, \rho = 2, s = 3, \theta_{harris} = 0.005, \theta_{gabor} = 0.01$

1.3 Παρατηρήσεις

Παρατηρούμε ότι γενικά ο ανιχνευτής Harris βρίσκει πολύ λιγότερα σημεία ενδιαφέροντος από τον ανιχνευτή Gabor. Ο λόγος για τον οποίο συμβαίνει αυτό είναι ότι σε όλα τα video samples που μας δόθηκαν οι κινήσεις διέπονται από περιοδικότητα, επομένως υπάρχει περιοδική μεταβολή του intensity των frames των video λόγω της επαναληπτικότητας κάποιων κινήσεων, γεγονός που καθιστά τον ανιχνευτή Gabor πιο ευαίσθητο στην ανίχνευση τέτοιων σημείων ενδιαφέροντος. Επιπλέον, καθώς αυξάνεται η ταχύτητα των κινήσεων, ο ανιχνευτής Gabor βρίσκει περισσότερα σημεία ενδιαφέρονος. Αυτό συμβαίνει, διότι ο εν λόγω ανιχνευτής βασίζεται στην ενέργεια των κινήσεων, η οποία είναι άμεση συνδεδεμένη με την ταχύτητα.

2 Χωρο-χρονικοί Ιστογραφικοί Περιγραφητές

Στο μέρος αυτό υπολογίζονται οι χωρο-χρονικοί ιστογραφικοί περιγραφητές. Πιο συγκεκριμένα, χρησιμοποιούνται οι περιγραφητές HOG και HOF καθώς και ο συνδυασμός τους HOG/HOF.

Επεξήγηση Υλοποίησης

Η υλοποίηση αυτού του μέρους πραγματοποιείται στο αρχείο `computeDescriptor.m` και στα αντίστοιχα call functions.
Σημείωση για υλοποίηση: Πρόκειται για μία συνάρτηση στην οποία χρησιμοποιείται Input Parser για το πέρασμα των αντίστοιχων παραμέτρων. Απαιτούνται ο προσδιορισμός του video, των σημείων ενδιαφέροντος και της μεθόδου υλοποίησης χωρο-χρονικού ιστογραφικού περιγραφητή. Οι υπόλοιπες παράμετροι δεν είναι απαραίτητες, καθώς έχουν οριστεί default τιμές. Ωστόσο, με τρόπο όμοιο με τον προσδιορισμό του περιγραφητή δύναται ο καθορισμός τους. Οι default τιμές των optional parameters είναι:

parameter	default value
scale	1.5
nbins	4
n	3
m	3

Για το σύνολο των περιγραφητών, απαιτείται προηγουμένως ο υπολογισμός των σημείων ενδιαφέροντος.

2.1 Περιγραφητές

Ο περιγραφητής υπολογίζεται σε μία γειτονιά γύρω από το αντίστοιχο σημείο ενδιαφέροντος. Το μέγεθος αυτής της γειτονιάς προσδιορίζεται από το Box_size που είναι ίσο με $4 \times scale$ με $scale = 1.5$ default τιμή. Αυτή η γειτονιά εξάγεται με την εντολή `crop`, η οποία καλείται με τρόπο ώστε να δίνεται προσοχή στα όρια της εικόνας.

2.1.1 HOG

computeDescriptor(video, pointsOfInterest,'HOG',···)
Για τον περιγραφητή HOG, υπολογίζεται η κατευθυντική παράγωγος μέσω της εντολής `imgradientxy` και παράγονται τα διανυσματικά πεδία G_x, G_y. Στη συνέχεια, υπολογίζεται η ιστογραμματική περιγραφή χρησιμοποιώντας τη συνάρτηση `OrientationHistogram.p`.

2.1.2 HOF

computeDescriptor(video, pointsOfInterest, 'HOF',···)
Για τον περιγραφητή HOF, υπολογίζεται η οπτική ροή (optical flow) με τη μέθοδο Lucas-Kanade του 2ου εργαστηρίου και παράγονται τα διανυσματικά πεδία G_x, G_y. Στη συνέχεια, υπολογίζεται η ιστογραμματική περιγραφή χρησιμοποιώντας τη συνάρτηση `OrientationHistogram.p`.

2.1.3 HOG/HOF

computeDescriptor(video, pointsOfInterest, 'HOGHOF',···)
Για τον περιγραφητή HOG/HOF, υπολογίζονται τα ιστογράμματα των μεθόδων HOG και HOF σύμφωνα με τα παραπάνω και στη συνέχεια συνενώνονται σε ένα με την εντολή `horzcat`.

2.2 Bag of Visual Words

Υπολογίζεται η τελική αναπαράσταση global representation για κάθε βίντεο. Εφόσον σε αυτή την περίπτωση δεν έχουμε train data, τα κέντρα των clusters εξάγονται από τα δεδομένα μας. Πιο συγκεκριμένα, για κάθε ιστογραμματική περιγραφή υπολογίζονται τα clusters με τη συνάρτηση `kmeans` επιλέγοντας $k = 60$. Στη συνέχεια, συνενώνονται όλα τα clusters (από όλα τα videos δηλαδή) σε ένα, κάτι που είναι απαραίτητο για το μέρος 3.
Σημειώνεται ότι λόγω του αλγορίθμου *kmeans* που υλοποιείται by default στο `MATLAB` [1], ο οποίος επιλέγει τυχαία δείγματα και centroids, τα αποτελέσματα διαφέρουν κάθε φορά που εκτελείται το `lab3.m`. Αυτό είναι εμφανές στα παραγόμενα δενδροδιαγράμματα του μέρους 3.

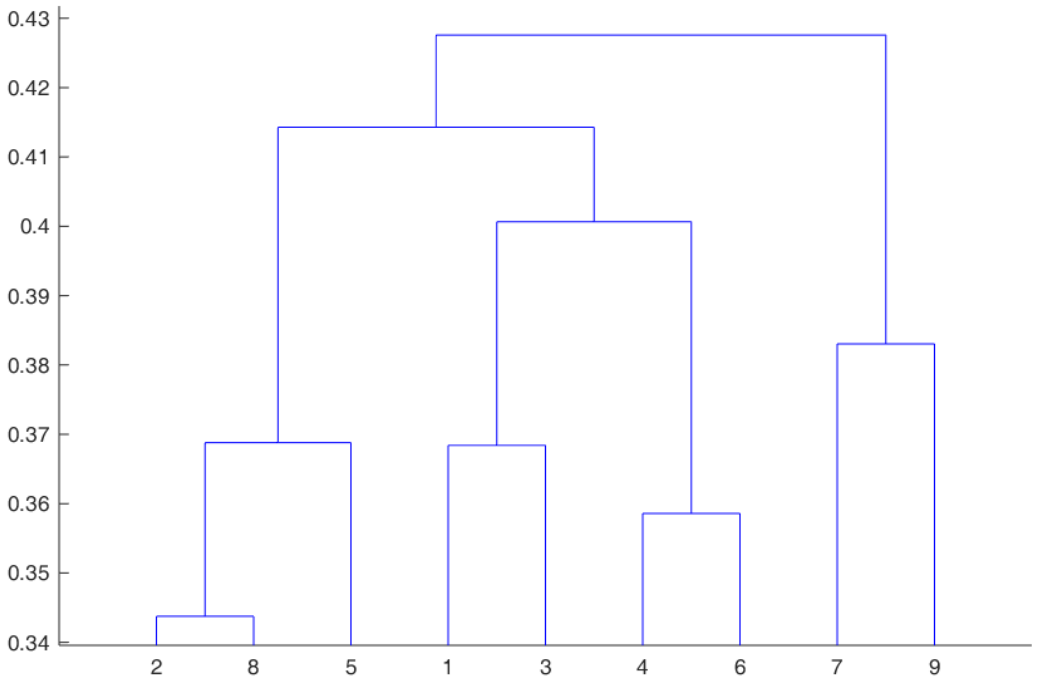
[1] https://www.mathworks.com/help/stats/kmeans.html#bueftl4-1 με υλοποίηση σύμφωνα με Arthur, David, and Sergi Vassilvitskii. "K-means++: The Advantages of Careful Seeding." SODA '07: Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms. 2007, pp. 1027–1035.

3 Κατασκευή Δενδρογράμματος για τον Διαχωρισμό των Δράσεων

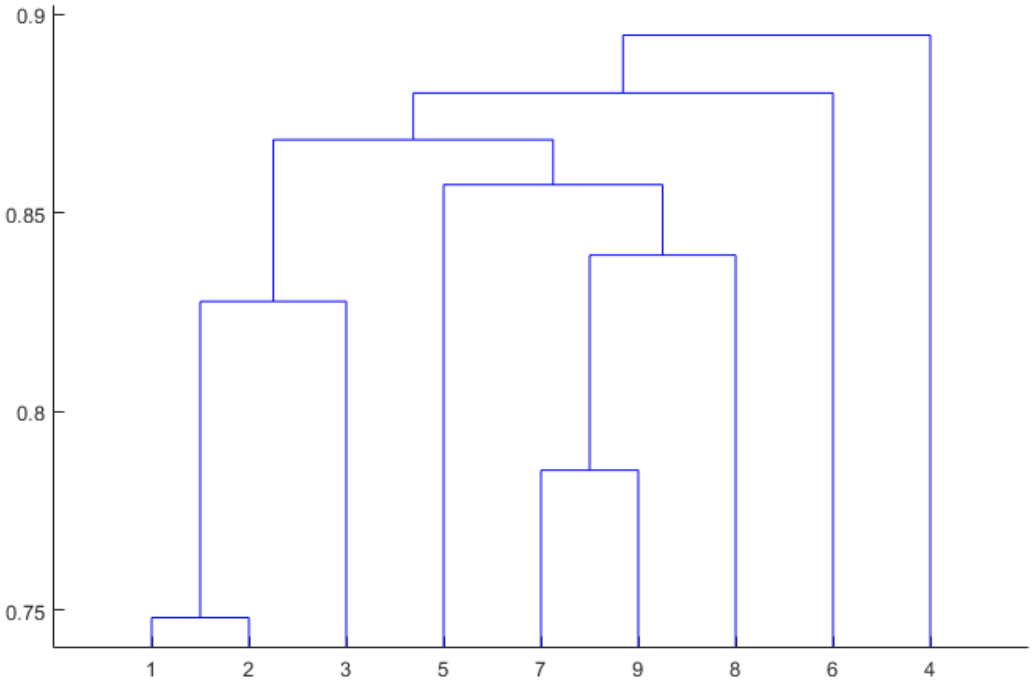
Στην ενότητα αυτή επιχειρούμε την κατηγοριοποίηση των 9 video που μας δόθηκαν σε 3 επιμέρους κατηγορίες (walking, running, boxing) χρησιμοποιώντας τις BoWV αναπαραστάσεις που βασίζονται σε HOG / HOF χαρακτηριστικά του προηγούμενου ερωτήματος. Η κατηγοριοποίηση θα επιτευχθεί ποιοτικά με την οπτικοποίηση της απόστασης των διανυσμάτων χαρακτηριστικών μέσω της κατασκευής ενός δενδροδιαγράμματος αποστάσεων που αντιπροσωπεύει την ικανότητα διαχωρισμού των 3 διαφορετικών δράσεων.
Για τον υπολογισμό των αποστάσεων ανάμεσα σε κάθε ζεύγος ιστογραμμάτων χρησιμοποιήθηκε η συνάρτηση `euclidist.m` και ως μετρική της απόστασης χρησιμοποιήθηκε η έτοιμη συνάρτηση `distChiSq.m`. Στη συνέχεια, χρησιμοποιήσαμε τη συνάρτηση `linkage` για τη σύνδεση σε ζεύγη των αντικειμένων που βρίσκονται πιο κοντά σε δυαδικούς clusters. Η συνάρτηση αυτή χρησιμοποιεί τις αποστάσεις που δημιουργήθηκαν προηγουμένως. Σημειώνουμε ότι η συνάρτηση κλήθηκε ως:

```
linkage(BoVW,'average',@distChiSq)
```

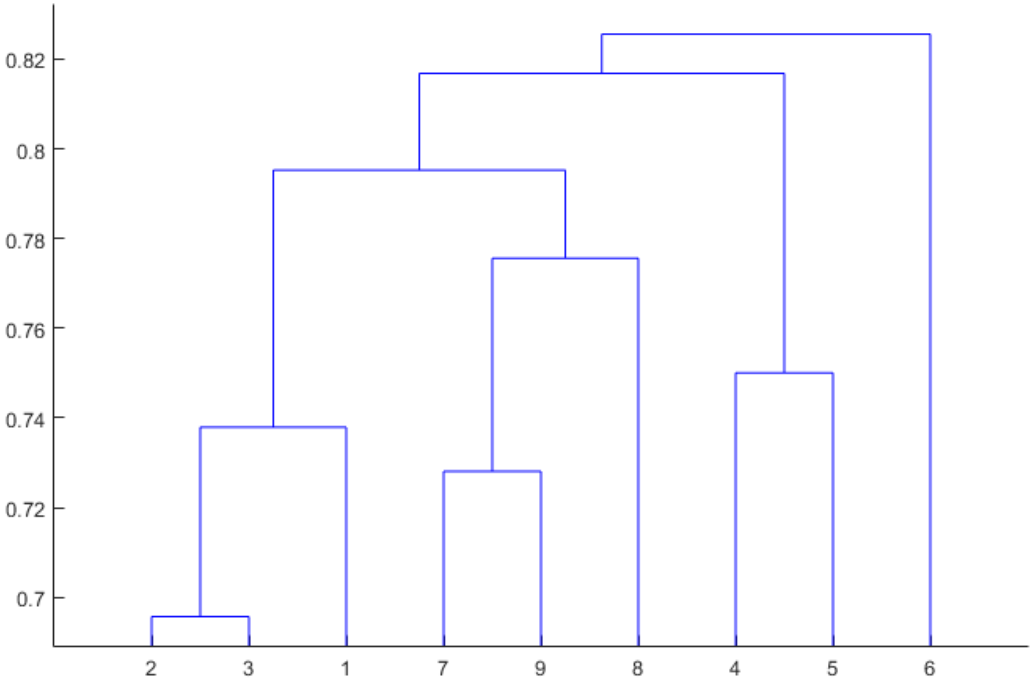
Για την οπτικοποίηση του δέντρου που σχημάτισε η `linkage` χρησιμοποιούμε τη συνάρτηση `dendrogram`. Παρακάτω παρουσιάζουμε τα δενδροδιαγράμματα που προέκυψαν από όλους του συνδυασμούς ανιχνευτών / περιγραφητών για συγκεκριμένες τιμές των παραμέτρων. Υπενθυμίζεται ότι τα αντικείμενα 1, 2, 3 αντιστοιχούν στα video boxing, τα 4, 5, 6 στα video walking και τα 7, 8, 9 στα video running.



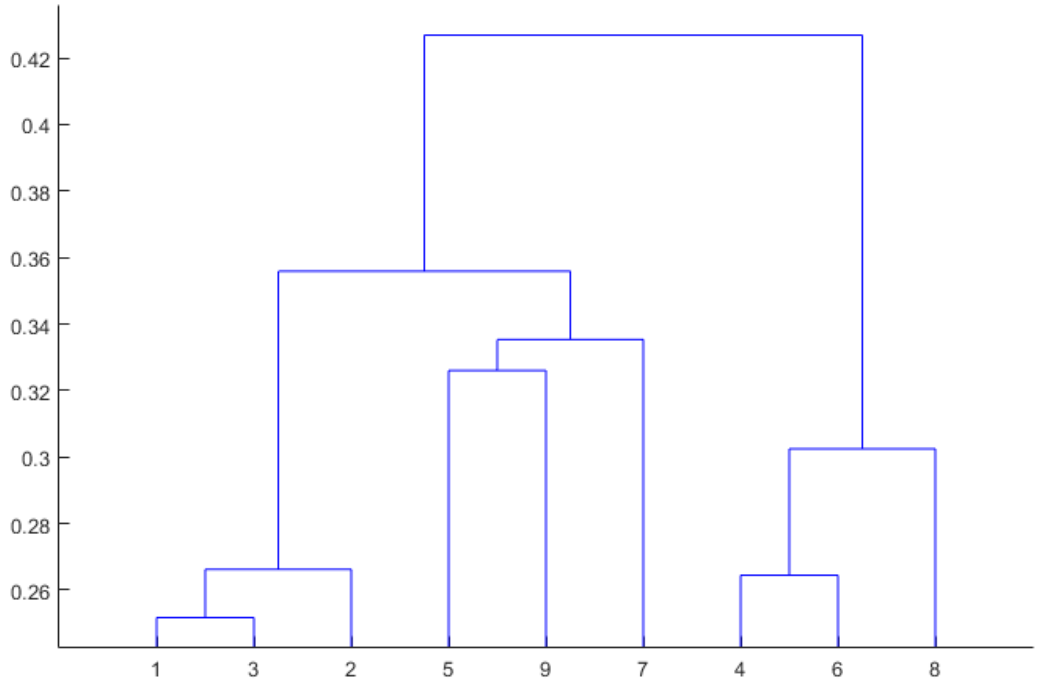
Harris - HOF



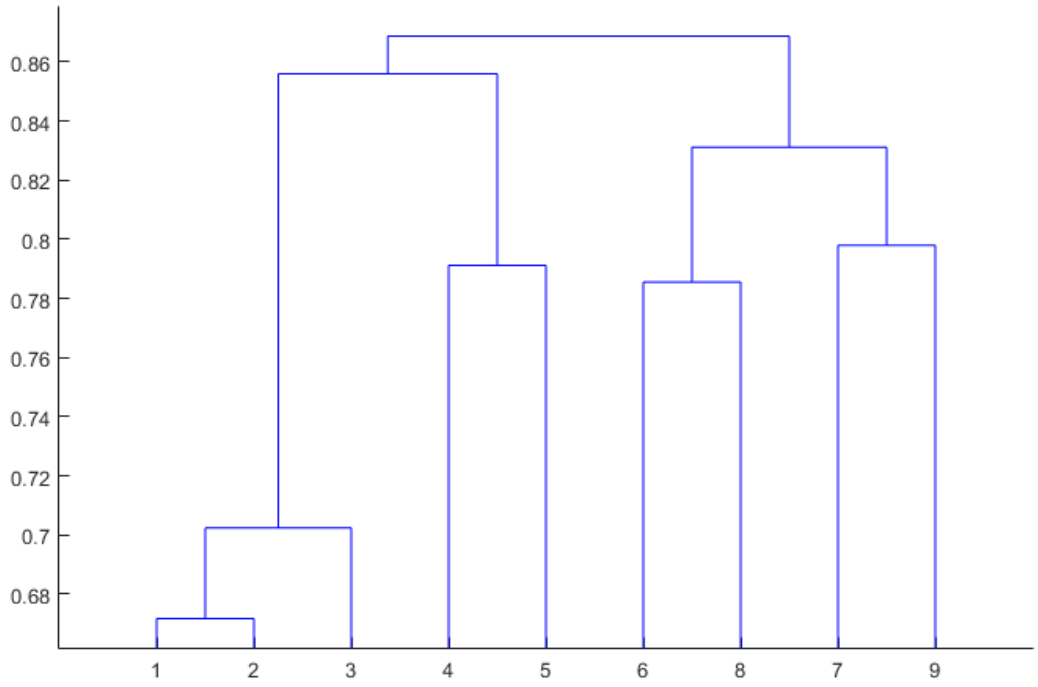
Harris - HOG



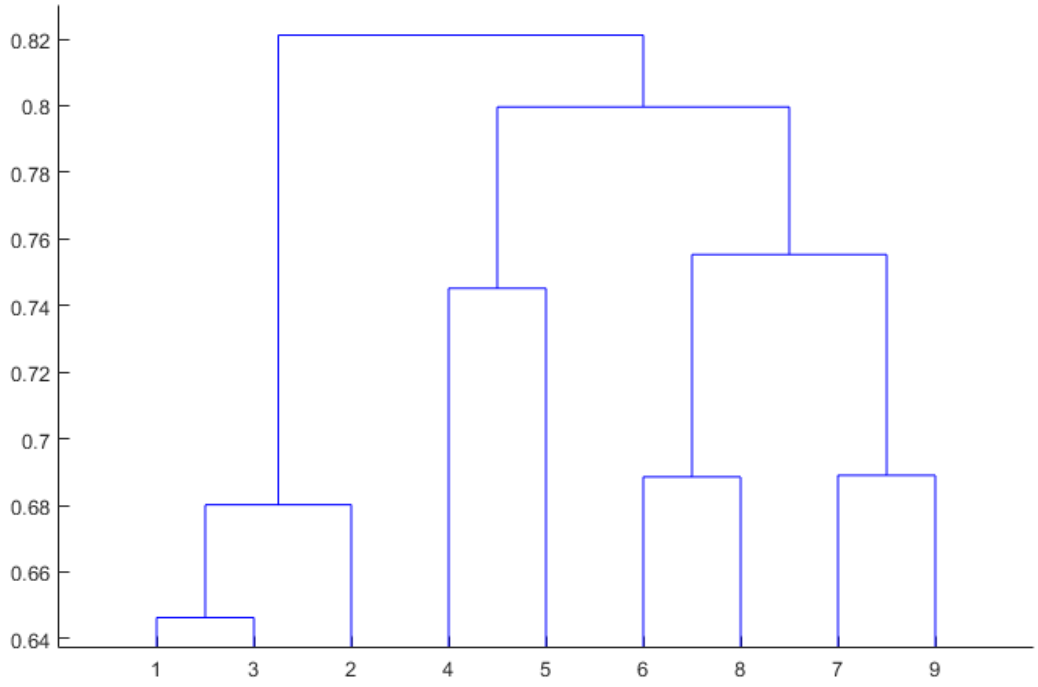
Harris - HOG/HOF



Gabor - HOF



Gabor - HOG



Gabor - HOG/HOF

Εικόνα 3: *Δενδροδιαγράμματα για όλους τους συνδυασμούς ανιχνευτών / περιγραφητών για* $\sigma = 1.5, \tau = 3, \rho = 1.5, s = 3, k = 0.005, \theta_{harris} = 0.005, \theta_{gabor} = 0.01, k\_clusters = 60, box\_size = 4\sigma,$

Παρατηρούμε ότι το καλύτερο αποτέλεσμα προκύπτει από τον συνδυασμό Harris - HOG/HOF καθώς με αυτόν τον συνδυασμό γίνεται πλήρης κατηγοριοποίηση των video σε 3 διαφορετικές δράσεις. Γενικότερα, βλέπουμε ότι τα αντικείμενα 1, 2, 3, που αντιστοιχούν στο boxing, ομαδοποιούνται σχεδόν σε όλους τους συνδυασμούς εκτός του Harris - HOF. Οι 2 άλλες κατηγορίες (walking και running) σε όλους τους συνδυασμούς εκτός του Harris - HOG/HOF συγχέονται λόγω της ομοιότητας των 2 δράσεων.