**ΕΜΠ**

**ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ**

**ΟΡΑΣΗ ΥΠΟΛΟΓΙΣΤΩΝ**

**2η Εργαστηριακή άσκηση**

**«Εντοπισμός Χωρο-χρονικών σημείων ενδιαφέροντος και εξαγωγή χαρακτηριστικών σε Βίντεο ανθρωπίνων δράσεων»**

8ο Εξάμηνο

ΑΚ.ΕΤΟΣ 2015-2016

**Ονοματεπώνυμα:**

Καλαβρυτινός Χαράλαμπος ΑΜ: 03112024

Χάτζος Μηνάς ΑΜ: 03112127

**ΕΙΣΑΓΩΓΗ**  
Σκοπός της 3ης εργαστηριακής άσκησης είναι η εξαγωγή χωρο-χρονικών χαρακτηριστικών ( *Spatio-temporal interest points*) με σκοπό **την κατηγοριοποιήση βίντεο** που περιέχουν ανθρώπινες δράσεις. Αρχικά εξάγουμε τοπικά χαρακτηριστικά σε ακολουθίες εικόνων βίντεο και έπειτα δημιουργούμε **τοπικούς περιγραφητές** που υπολογίζονται σε γειτονιές ανιχθευθέντων σημείων ενδιαφέροντος. Τελικά, χρησιμοποιώντας την τεχνική **Βag of Visual Words** δημιουργούμε μία τελική αναπαράσταση global representation με βάση την οποία γίνεται η τελική κατηγοριοποίηση.

**ΜΕΡΟΣ 1: Χωροχρονικά σημεία ενδιαφέροντος**

Στο πρώτο μέρος της άσκησης ζητείται η εξαγωγή χωρο-χρονικών σημείων ενδιαφέροντος από βίντεο. Αρχικά *διαβάζουμε 200 frames* από τα βίντεο που περιέχονται στο φάκελο samples και περιέχουν ανθρώπινες δράσεις **boxing, walking και running**. Για την ανάγνωση χρησιμοποιείται η έτοιμη συνάρτηση readVideo() που μας δίνεται. Για την εξαγωγή χρησιμοποιούνται συνολικά 2 διαφορετικοί ανιχνευτές, ο ανιχνευτής γωνιών **Harris σε 3 Διαστάσεις**, και ένας ανιχνευτής βασισμένος σε **χρονικά φίλτρα gabor**.

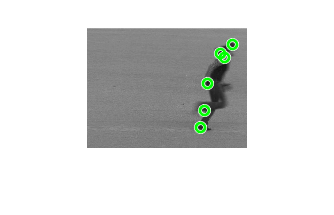
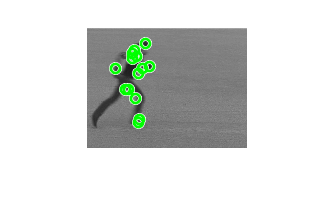
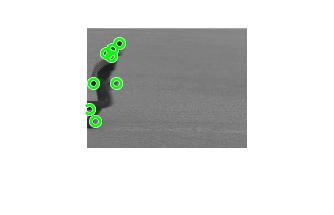
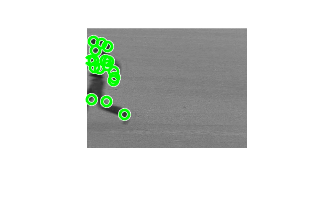
**Ανιχνευτής Harris σε 3D:** Ο ανιχνευτής γωνιών Harris-Stephens χρησιμοποιήθηκε στην πρώτη εργαστηριακή άσκηση για εξαγωγή γωνιών σε δισδιάστατες εικόνες. Αργότερα οι *Laptev* και *Linderberg* [1], επέκτειναν αυτόν τον ανιχνευτή στις τρεις διαστάσεις για εξαγωγή χωρο-χρονικών χαρακτηριστικών μέσω ενός πίνακα:

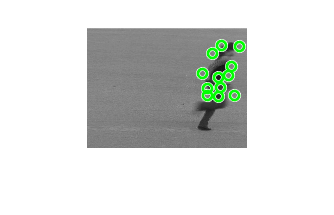
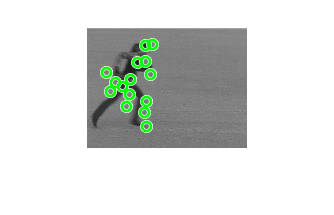
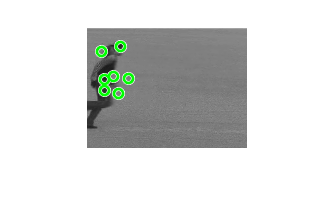
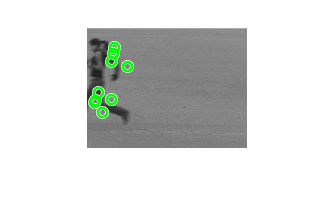
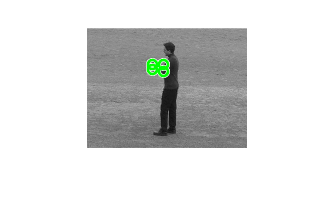
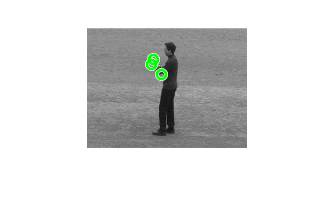
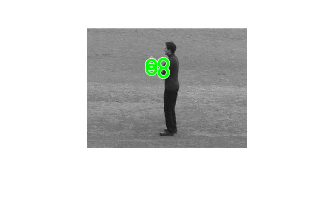
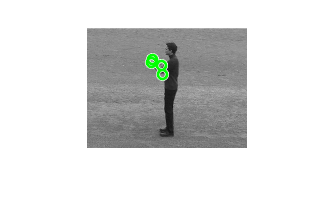
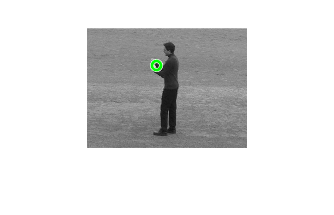
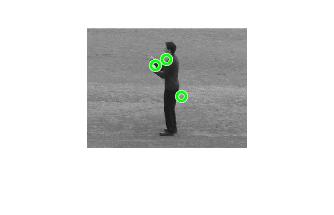
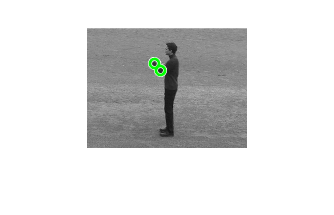
όπου χρησιμοποιούνται ανεξάρτητα η χωρική και χρονική κλίμακα σ και τ αντίστοιχα, ένας τρισδιάστατος gaussian πυρήνας ομαλοποίησης g και οι χωροχρονικές παράγωγοι. Τελικά το 3D κριτήριο γωνιότητας προκύπτει ως .

Οι θέσεις των σημείων ενδιαφέροντος προκύπτουν ως τα τοπικά ελάχιστα του παραπάνω κριτηρίου, τα οποία είναι μεγαλύτερα από μία σταθερά γωνιότητας. Στην παρούσα εργασία χρησιμοποιήθηκαν παράμετροι k=0.0005 και διάφορες κλίμακες σ=1.5,2,3,4, τ=3,4,6,8.

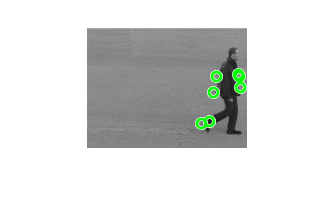
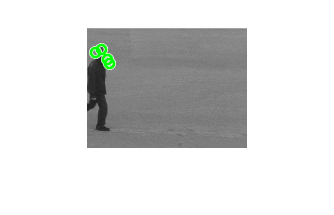
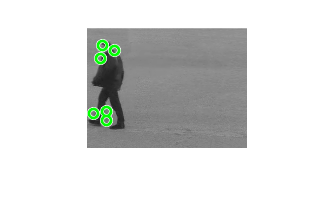
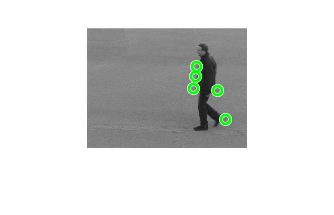
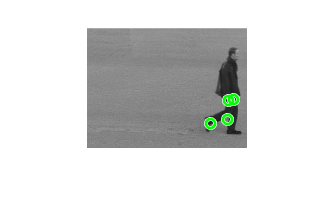
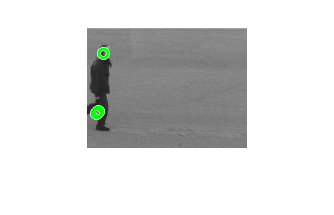
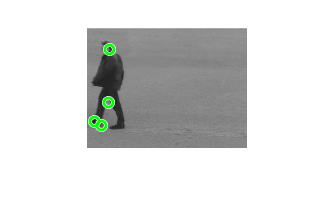
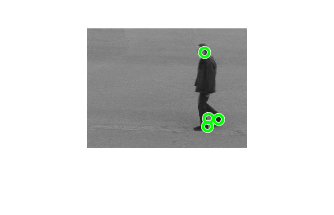
**Ανιχνευτής με φίλτρα Gabor:** αυτός ο ανιχνευτής βασίζεται σε χρονικά 1Δ φίλτρα gabor και προτάθηκε από τον Dollar [2] . Το κριτήριο σημαντικότητας προκύπτει από την τετραγωνική ενέργεια εξόδου για το ζεύγος των Gabor φίλτρων ως:

όπου g ένας 2Δ χωρικός πυρήνας ομαλοποίησης, και , ένα quadrature pair από 1Δ Gabor φίλτρα εφαρμοσμένα στο χρόνο. Τα τελικά σημεία ενδιαφέρονοτας προκύπτουν και πάλι ως τα τοπικά μέγιστα του κριτηρίου H(x,y,t). Παρακάτω παρουσιάζονται ενδεικτικά αποτελέσματα από τις δύο μεθόδους.





Εικόνα 1.2. Στιγμιότυπα ανίχνευσης σημείων ενδιαφέροντος STIP με χρήση του ανιχνευτή Harris 3D ( πάνω) και Gabor 3D (κάτω) για τα αντίστοιχα frames 22, 88, 91 και 101 (από αριστερά προς δεξιά) στο 3ο running video για sigma=1.5, tau=3.



Εικόνα 1.3. Στιγμιότυπα ανίχνευσης σημείων ενδιαφέροντος STIP με χρήση του ανιχνευτή Harris 3D ( πάνω) και Gabor 3D (κάτω) για τα αντίστοιχα frames 24, 54, 115 και 153 (από αριστερά προς δεξιά) στο 2ο walking video για sigma=1.5, tau=3.

Εικόνα 1.2. Στιγμιότυπα ανίχνευσης σημείων ενδιαφέροντος STIP με χρήση του ανιχνευτή Harris 3D ( πάνω σειρά) και Gabor 3D (κάτω σειρά) για τα frames 22, 43, 134 και 169 (από αριστερά προς δεξιά) στο 1ο boxing video για sigma=1.5, tau=3.

Γενικά παρατηρούμε ότι ο ανιχνευτής **Harris** βρίσκει κάθε φορά λιγότερα σημεία σε σχέση με τον ανιχνευτή gabor. Ενδεικτικά, στα αποτελέσματα που παρουσιάζονται παραπάνω συγκρίνουμε τους δύο ανιχνευτές στα ίδια ακριβώς τυχαία στιγμιότυπα ενός βίντεο.

Οι ανθρώπινες συμπεριφορές όπως **boxing, τρέξιμο, περπάτημα**, κά, χαρακτηρίζονται εύκολα από την αντιστρεψιμότητα της κίνησης των άκρων. Οπότε αυτές οι συμπεριφορές **περιέχουν αρκετές χωροχρονικές γωνίες** και άρα ο ανιχνευτής Harris επιτυγχάνει. Σε άλλες περιπτώσεις όμως όπου οι χωροχρονικές γωνίες είναι σπάνιες ο ανιχνευτής Ηarris δεν φέρει σωστά αποτελέσματα. Επιπλέον, οι γωνίες αυτές δεν είναι πάντα αρκετές για την κατηγοριοποιήση μίας συμπεριφοράς. Αντίθετα, ο **ανιχνευτής gabor** ανιχνεύει σημεία που υπάρχουν περιοδικές μεταβολές στο intensity των εικόνων, δηλαδή βρίσκει σημεία ενδιαφέροντος με βάση και τις γωνίες σε χωροχρονικό επίπεδο αλλα και την επαναληψιμότητα κάποιων κινήσεων (*κινήσεις που συμβαίνουν περιοδικά).*

**ΜΕΡΟΣ 2: Χωρο-χρονικοί Ιστογραφικοί περιγραφητές**

Στο δεύτερο μέρος της άσκησης υλοποιούμε χωρο-χρονικούς περιγραφητές με βάση των υπολογισμό ιστογραμμάτων **κατευθυντικής παραγώγου (HOG) και οπτικής ροής (HOF)** γύρω από τα σημεία ενδιαφέροντος που υπολογίστηκαν στο πρώτο μέρος.

Για κάθε εικόνα της ακολουθίας βίντεο υπολογίζουμε το διάνυσμα κλίσης και το διάνυσμα οπτικής ροής με τη την πολυκλιμακωτό αλγόριθμο **Lucas-Kanade** όπως στην προηγούμενη εργαστηριακή άσκηση. Χρησιμοποιώντας τη συνάρτηση OrientationHistogram.p() που μας δίνεται υπολογίζουμε τους δύο ιστογραφικούς περιγραφητές για κάθε σημείο ενδιαφέροντος. Εναλλακτικά χρησιμοποιούμε τη συνάρτηση Orientation\_Histogram() την οποία φτιάξαμε εμείς. Αρχικά εξάγουμε τα διανυσματικά πεδία σε μία τετραγωνική περιοχή γύρω από κάθε σημείο ενδιαφέροντος μεγέθους 4∙scale και έπειτα υπολογίζουμε τα ιστογράμματα. Αφού υπολογίσουμε τα ιστογράμματα HOG και HOF συνενώνουμε τους δύο περιγραφητές για τη δημιουργία του HOG/HOF περιγραφητή.

Εφόσον βρούμε όλα τα ιστογράμματα, καλούμαστε να υπολογίσουμε την τελική αναπαράσταση global representation για κάθε ένα από τα 9 βίντεο υλοποιώντας την τεχνική **BoVW** (Bag of Visual Words). Χρησιμοποιούμε τη συνάρτηση kmeans() και βρίσκουμε από τα ιστογράμματα **τις visual words** οι οποίες θα χρησιμοποιηθούν αργότερα για την κατηγοριοποιήση των βίντεο.

Επίσης, αναφέρουμε ότι κατασκευάσαμε μια δικιά μας υλοποίηση της συνάρτησης OrientationHistogram.p()η οποία τελικά μας έδωσε όμοια αποτελέσματα με την δοσμένη. Στο Directory που βρίσκονται τα script, υπάρχει η συναρτήση Orientation\_Histogram() καθώς και μία επιπλέον, η Point\_Orientation\_Histogram() η οποία καλείται αποκλειστικά απο την πρώτη, και υπολογίζει το ιστόγραμμα για μέγεθoς grid 1 x 1.  
Τέλος, τα ορίσματα της συνάρτησης είναι ίδια με αυτά που δίναμε στην δοσμένη συνάρτηση και χρησιμοποιήθηκε 50 % επικάλυψη παραθύρων και για τις δύο διευθύνσεις.

**ΜΕΡΟΣ 3: Κατασκευή Δενδρογράμματος για Διαχωρισμό των Δράσεων**

Στο τελευταίο μέρος της εργαστηριακής άσκησης ζητείται να γίνει η κατηγοριοποίηση των 9 βίντεο που αναλύθηκαν νωρίτερα σε 3 κατηγορίες. Η κατηγοριοποιήση γίνεται χρησιμοποιώντας τις **BoVW αναπαραστάσεις**, βασισμένες σε **HOG/HOF** χαρακτηριστικά που βρέθηκαν στο προηγούμενο ερώτημα. Για την κατηγοριοποίηση χρησιμοποιούμε την οπτικοποίηση της απόστασης των διανυσμάτων των χαρακτηριστικών μέσω της κατασκευής ενός δενδρογράμματος αποστάσεων (*ιεαραρχική κατηγοριοποίηση*). **Για τη δημιουργία των δενδρογραμμάτων:**

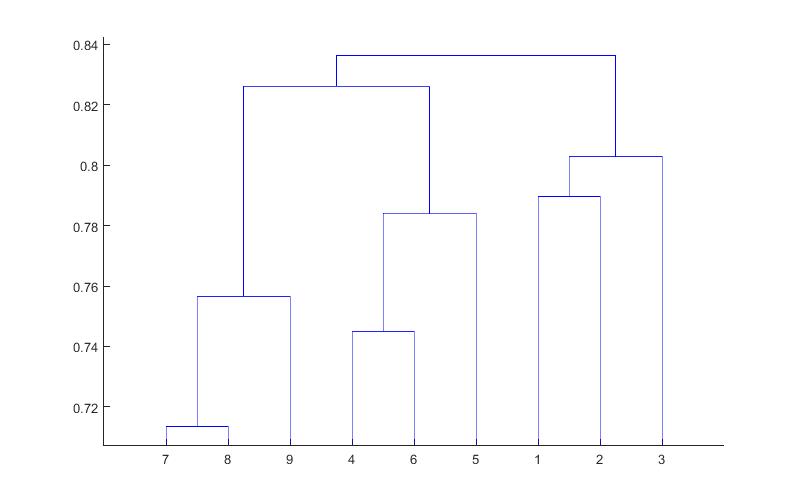
Αρχικά χρησιμοποιείται η pdist() για τον υπολογισμό των αποστάσεων ανάμεσα σε κάθε ζεύγος ιστογραμμάτων ( το καθένα αντιστoιχεί σε ένα βίντεο). Ως μετρική της απόστασης χρησιμοποιούμε τη δοθείσα συνάρτηση distChiSq(). Αφού υπολογιστούν οι αποστάσεις, χρησιμοποιείται η συνάρτηση linkage() , η οποία χρησιμοποιεί τις αποστάσεις που δημιουργήθηκαν από την προηγούμενη συνάρτηση και **συνδέει σε ζεύγη τα αντικείμενα** (κάθε αντικείμενο αντιστοιχεί σε ένα βίντεο) που βρίσκονται πιο κοντά σε *δυαδικούς clusters.* Η συνάρτηση συνεχίζει να συνδέει τους clusters που δημιουργεί μεταξύ τους ώσπου να συνδεθούν όλα τα δεδομένα δημιουργώντας ένα ιεραρχικό δένδρο. Τέλος, η συνάχρτηση dendrogram() χρησιμοποιείται για την *οπτικοποίηση του δένδρου* που δημιουργήθηκε.

Παρακάτω φαίνονται ενδεικτικά τα καλύτερα δυνατά αποτελέσματα δενδρογραμμάτων για διάφορους συνδυασμούς *περιγραφητών/ ανιχνευτών/ τιμών παραμέτρων* (π.χ.αριθμός των λέξεων). Σημειώνουμε ότι τα **1,2,3 αντικείμενα** αντιστοιχούν στα βίντεο **boxing**, τα **4,5,6** αντιστοιχούν στα video **running** ενώ **τα 7,8,9 σε video walking**. Ο καλύτερος δυνατός συνδυασμός επιτεύχθηκε με χρήση του **ανιχνευτή Gabor 3D και του συνδυασμένου περιγραφητή HOG/HOF**.

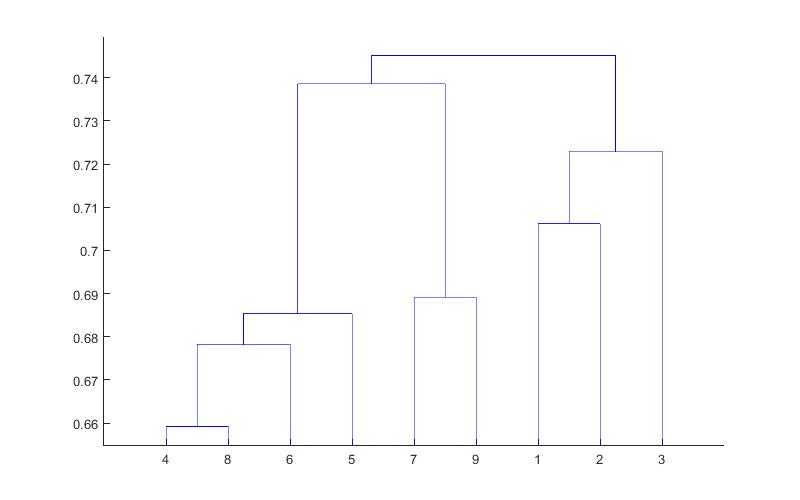
*ΣΗΜΕΙΩΣΗ:* Επειδή σε κάποιες εκδόσεις του Matlab υπάρχει η πιθανότητα η συνάρτηση Point\_Orientation\_Histogram να σκάει, θα πρέπει να προστεθεί το εξής στο τέλος της συνάρτησης:

if (sum(point\_desc) ~= 0) point\_desc = point\_desc / norm(point\_desc,2) ;

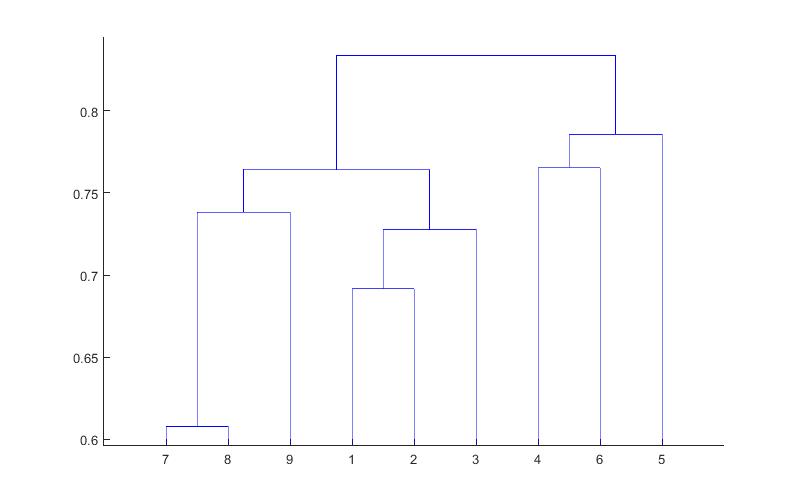
end



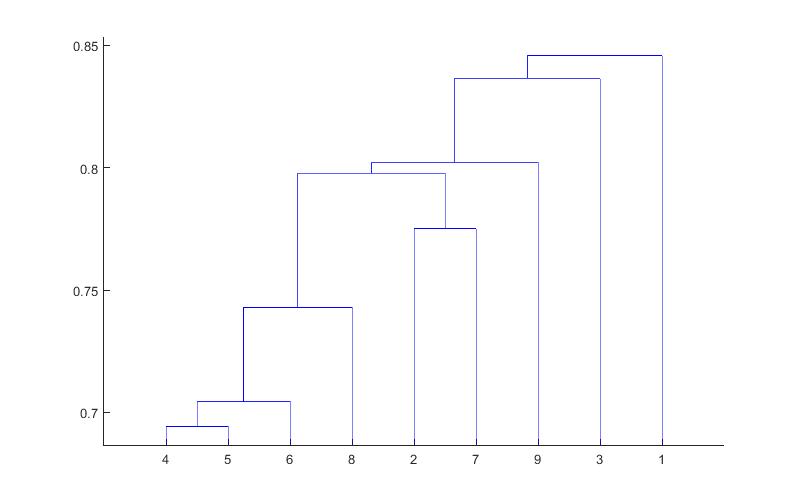
Εικόνα 3.1. Δενδρόγραμμα για ανιχνευτή Gabor 3D, και παρακολουθητή HOG/HOF με κλίμακες sigma=1.5, tau=3, nwords=60 λέξεις, No=600 σημεία ενδιαφέροντος. Παρατηρούμε ότι μπορεί να ομαδοποιήσει επιτυχώς και τα 9 βίντεο στις σωστές κατηγορίες



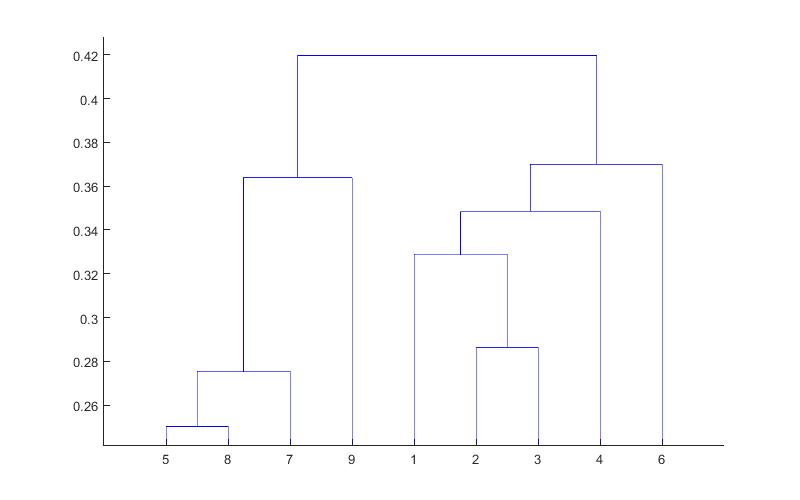
Εικόνα 3.2. Δενδρόγραμμα για ανιχνευτή Harris 3D, και παρακολουθητή HOG/HOF με κλίμακες sigma=1.5, tau=3, nwords=60 λέξεις, No=600 σημεία ενδιαφέροντος. Παρατηρούμε ότι τα boxing videos εχουν ομαδοποιηθεί σωστά, όπως και δύο από τα walking, αλλά υπάρχει σφάλμα στο βίντεο που κατηγοριοποιηθηκε μαζί με τα running. Συνολικά έχουν βρεθεί στη σωστή κατηγορία 8 από τα 9 βίντεο



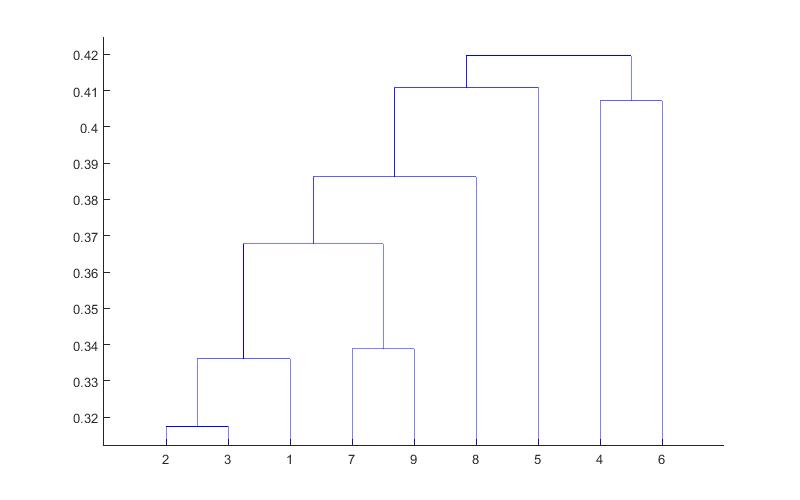
Εικόνα 3.3. Δενδρόγραμμα για ανιχνευτή Gabor 3D, και παρακολουθητή HOG με κλίμακες sigma=1.5, tau=3, nwords=60 λέξεις, No=600 σημεία ενδιαφέροντος. Και σε αυτή την περίπτωσηη κατηγοριοποίηση στις τρεις κλάσεις πέτυχε.



Εικόνα 3.4. Δενδρόγραμμα για ανιχνευτή Harris 3D, και παρακολουθητή HOG με κλίμακες sigma=1.5, tau=3, nwords=60 λέξεις, No=600 σημεία ενδιαφέροντος. Παρατηρούμε ότι η κατηγοριοποιήση είναι επιτυχής μόνο στις 3 περιπτώσεις



Εικόνα 3.5. Δενδρόγραμμα για ανιχνευτή Gabor 3D, και παρακολουθητή HOF με κλίμακες sigma=1.5, tau=3, nwords=60 λέξεις, No=600 σημεία ενδιαφέροντος.



Εικόνα 3.6. Δενδρόγραμμα για ανιχνευτή Harris 3D, και παρακολουθητή HOF με κλίμακες sigma=1.5, tau=3, nwords=60 λέξεις, No=600 σημεία ενδιαφέροντος.

Από τα παραπάνω συμπεραίνουμε ότι ο ανιχνευτής Gabor 3D δίνει στη γενική περίπτωση καλύτερα αποτελέσματα, και πως ο **συνδυασμός HOG/HOF** κατηγοριοποιεί καλύτερα τα βίντεο σε σχέση με τους επιμέρους περιγραφητές HOG ή HOF. Σε όλες σχεδόν τις περιπτώσεις τα βίντεο boxing κατηγοριοποιούνται εύκολα μαζί, ενώ τα βίντεο με *running και walking* πολλές φορές μπερδεύονται μεταξύ τους, μιας και η συμπεριφορά είναι παρόμοια.

Έπειτα από αρκετές δοκιμές για διαφορετικές κλίμακες χώρου και χρόνου, καταλήξαμε ότι τα βέλτιστα αποτελέσματα δίνονται με τους παραπάνω συνδυασμούς. Σε διάφορα papers προτείνονται τιμές *sigma=2 και tau=4* για τη μέθοδο **Gabor**, και *sigma=2, tau=3* με διπλάσιες ολοκληρωτικές κλίμακες sigma=4 και tau=6 για τη μέθοδο **Harris,** παρολαυτά παρατηρήσαμε ότι με τις τιμές 1.5 και 3 εξάγαμε καλύτερα αποτελέσματα.

REFERENCES:

[1] I. Laptev and T. Lindeberg. Space-time interest points. In ICCV , 2003

[2] P. Dollar, V. Rabaud, G. Cottrell, and S. Belongie. Behavior recognition via sparse spatio-temporal features. In VS-PETS , 2005.