



## ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ  
ΥΠΟΛΟΓΙΣΤΩΝ

---

### ΔΕΥΤΕΡΗ ΕΡΓΑΣΤΗΡΙΑΚΗ ΑΝΑΦΟΡΑ ΟΡΑΣΗ ΥΠΟΛΟΓΙΣΤΩΝ

---

Γεώργιος Καούκης  
03119006  
geoka01@gmail.com  
Αναστασία Χριστίνα Λίβα  
03119029  
anachriliva@gmail.com

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

Σκοπός του πρώτου Μέρους του εργαστηρίου είναι η υλοποίηση ενός συστήματος παρακολούθησης προσώπου και χεριών (Face and Hands Tracking) σε μια ακολουθία βίντεο νοηματικής γλώσσας. Το σύστημα αρχικά θα ανιχνεύει στο πρώτο πλαίσιο την περιοχή του προσώπου και των χεριών (περιοχές ενδιαφέροντος) με χρήση ενός πιθανοτικού ανιχνευτή ανθρώπινου δέρματος. Στη συνέχεια θα μπορεί να παρακολουθεί τις περιοχές ενδιαφέροντος χρησιμοποιώντας τα εξαγόμενα διανύσματα οπτικής ροής, υπολογισμένα με τη μέθοδο των Lucas-Kanade.

### Ανίχνευση Δέρματος Προσώπου και Χεριών

Στο πρώτο ερώτημα ζητείται 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. Το χρώμα του δέρματος μοντελοποιείται με μια διδιάστατη Γκαουσιανή κατανομή:

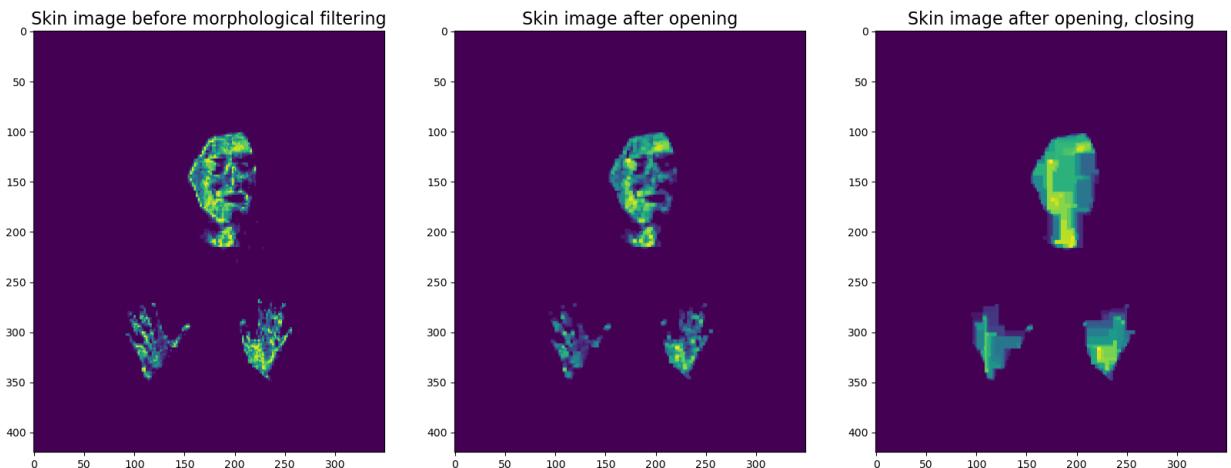
$$P(c = \text{skin}) = \frac{1}{\sqrt{|\Sigma|(2\pi)}} e^{-\frac{1}{2}(c-\mu)^T \Sigma^{-1} (c-\mu)'}$$

όπου

$$\mu = [\mu_{Cb} \quad \mu_{Cr}]^\top$$

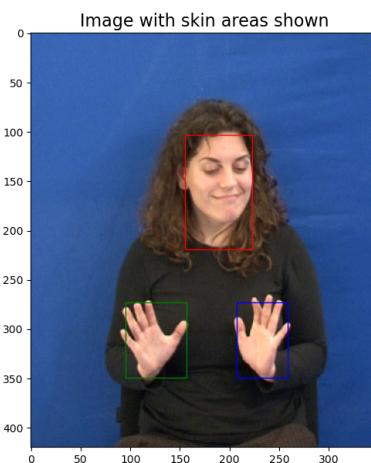
η μέση τιμή και  $\Sigma$  ο  $2 \times 2$  πίνακας συνδιακύμανσης από τα δείγματα δέρματος που λαμβάνουμε από το αρχείο mat που δίνεται.

Η δυαδική εικόνα ανίχνευσης δέρματος προκύπτει από την εικόνα πιθανοτήτων  $P(c(x, y) = \text{skin}), \forall(x, y)$  με κατωφλιοποίηση. Για την τελική ανίχνευση των περιοχών δέρματος του προσώπου και των χεριών πραγματοποιήσαμε μια μορφολογική επεξεργασία της δυαδικής εικόνας δέρματος καλύπτοντας τις τρύπες που εμφανίζονται, εφαρμόζοντας opening με ένα πολύ μικρό δομικό στοιχείο και closing με ένα μεγάλο δομικό στοιχείο. Αυτό είχε σαν αποτέλεσμα να εξαφανισθούν οι μικρές περιοχές και να αποκτήσουν συνοχή οι περιοχές του προσώπου και των χεριών. Αυτό φαίνεται στο παρακάτω σχήμα:



Σχήμα 1: Αποτέλεσματα Μορφολογικής Επεξεργασίας

Τέλος, δημιουργήσαμε τρία ορθογώνια που ύπαρχαν στην περιοχή της πρόσωπος και των χεριών με χρήση της συνάρτησης  $\text{fd}(\mathbf{I}, \mu, \text{cov})$  η οποία δέχεται ως παραμέτρους την πρώτη εικόνα της ακολουθίας βίντεο, τη μέση τιμή μ και τη συνδιακύμανση  $\Sigma$  της Γκαουσιανής κατανομής. Το αποτέλεσμα που προέκυψε είναι το εξής:



Σχήμα 2: Bounding Boxes

## Παρακολούθηση Προσώπου και Χεριών

Η αρχικοποίηση των Bounding Boxes που περιλαμβάνει το πρόσωπο και τα χέρια της νοηματίστριας στη μορφή  $[x, y, \text{width}, \text{height}]$  είναι η εξής: Πρόσωπο: [154, 102,

67, 115], Αριστερό χέρι: [93, 272, 56, 83], Δεξί χέρι: [201, 270, 56, 83]

## Υλοποίηση του Αλγορίθμου των Lucas-Kanade

Ο αλγόριθμος των Lucas-Kanade υπολογίζει την οπτική ροή σε κάθε σημείο της εικόνας  $x$  χρησιμοποιώντας τη μέθοδο των ελάχιστων τετραγώνων, ψεωρώντας ότι το  $d(x) = (d_x, d_y)$  είναι σταθερό σε ένα μικρό παράθυρο γύρω από το σημείο  $x$  και ελαχιστοποιώντας το τετραγωνικό σφάλμα:

$$J_x(d) = \int_{R^2} G_\rho(x - x')[I_n(x') - I_{n-1}(x' + d)] dx'$$

όπου  $I_x$ ,  $I_y$  και  $I_t$  είναι οι παράγωγοι της εικόνας  $I$  σε σχέση με το  $x$ , το  $y$  και το χρόνο  $t$  αντίστοιχα, και  $\rho$  είναι ένα σημείο στο παράθυρο γύρω από το επεξεργαζόμενο σημείο. όπου  $G_\rho(x)$  είναι μια Γκαουσιανή συνάρτηση παραθύρωσης με τυπική απόκλιση  $\rho$ .

Στόχος μας είναι η παρακολούθηση του προσώπου και των χεριών κατά τη διάρκεια του βίντεο. Έτσι ψέλουμε να εκτιμήσουμε την μετατόπιση του αντικειμένου από το  $i-1$  frame στο  $i$  frame, ώστε να έρθουν σε αντιστοιχία δύο διαδοχικές εικόνες

$$I_n(x) \approx I_{n-1}(x + d)$$

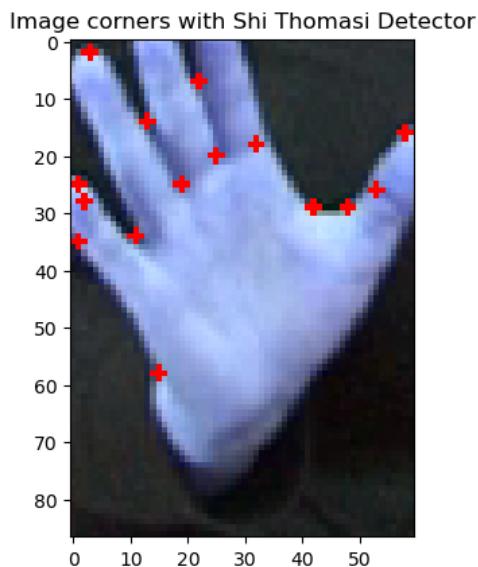
Στον αλγόριθμο αυτό δίνουμε ως είσοδο το bounding box του προσώπου, του αριστερού και του δεξιού χεριού αντίστοιχα, όπως ορίστηκαν παραπάνω, προκειμένου να υπολογιστεί η οπτική ροή. Τονίζουμε όμως ότι πρώτα έχει γίνει εξαγωγή των φεατορες στο αντίστοιχο βουνδινγ βοξ μέσω της συνάρτησης cv2.goodFeaturesToTrack, που υλοποιεί την μέθοδο των Shi και Tomashi, προκειμένου να βρεθεί η οπτική ροή στα σημεία που αντιστοιχούν σε features και όχι σε όλη την εικόνα. Έπειτα ο αλγόριθμος προσπαθεί επαναληπτικά να βρεί μια προσέγγιση του  $d$ , η οποία ελαχιστοποιεί τη διαφορά ανάμεσα στα δύο frames. Σε κάθε επανάληψη του αλγορίθμου προσεγγίζουμε την τιμή  $d_{i+1} = d_i + u$ , όπου με  $i$  συμβολίζουμε το γύρο των επαναλήψεων και  $u$  το διάνυσμα που προκύπτει από την παρακάτω σχέση:

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

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

$$E(x) = I_n(x) - I_{n-1}(x + d_i)$$

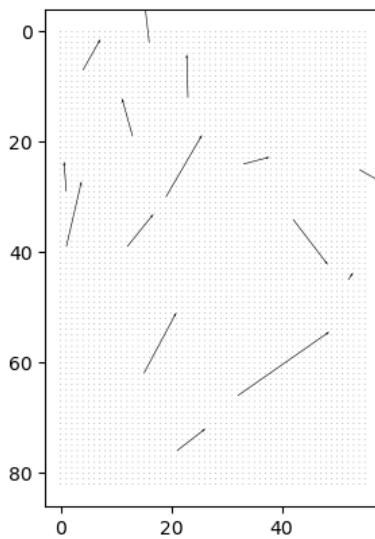
Η συνάρτηση αυτή δέχεται ως εισόδους δύο εικόνες (κομμένα παράθυρα με βάση το bounding box από δύο διαδοχικά πλαίσια του βίντεο), ένα σύνολο από σημεία ενδιαφέροντος (γωνίες) εντός του παραθύρου, το εύρος  $\rho$  του γκαουσιανού παραθύρου, τη θετική σταθερά κανονικοποίησης  $\epsilon$ , και την αρχική εκτίμηση  $d_0$  για το πεδίο οπτικής ροής, και επιστρέφει το διάνυσμα μετατόπισης για την κάθιση σημείου ενδιαφέροντος.



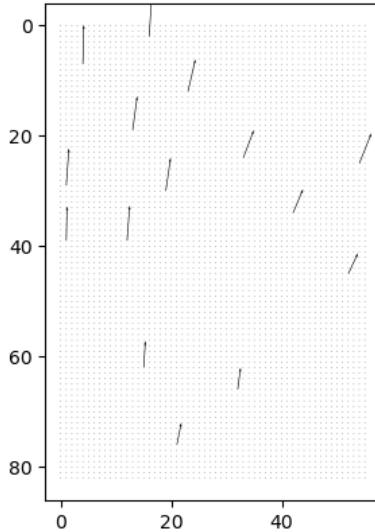
Σχήμα 3: Σημεία Ενδιαφέροντος για το Αριστερό Χέρι της Πρώτης Εικόνας

Στη συνέχεια χρησιμοποιούμε τον αλγόριθμο των Lucas - Kanade για να δούμε την οπτική ροή για το αριστερό χέρι ανάμεσα στο πρώτο και το δεύτερο στιγμιότυπο.

**Σημείωση:** Παίρνουμε σημεία ενδιαφέροντος **πάντα** από το μεταγενέστερο χρονικά στιγμιότυπο



Σχήμα 4: Οπτική Ροή Μεταξύ Πρώτου και Δεύτερου Στιγμιοτύπου για το Αριστερό Χέρι με Χρήση της Δικής μας Lucas Kanade



Σχήμα 5: Οπτική Ροή Μεταξύ Πρώτου και Δεύτερου Στιγμιοτύπου για το Αριστερό Χέρι με Χρήση Έτοιμης Συνάρτησης

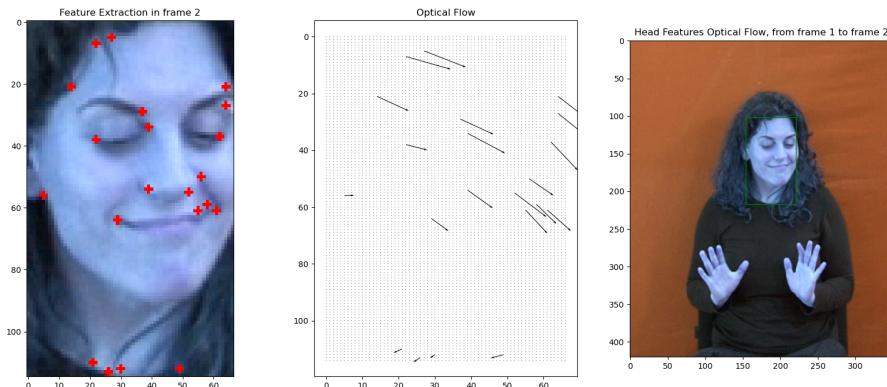
Βλέπουμε πως κατά βάση η οπτική ροή βγαίνει παρόμοια αφού τα διανύσματα έχουν ίδιες αφετηρίες αν και αρκετές φορές διαφέρει η κλίση.

## Υπολογισμός της Μετατόπισης των Παραθύρων από τα Διανύσματα Οπτικής Ροής

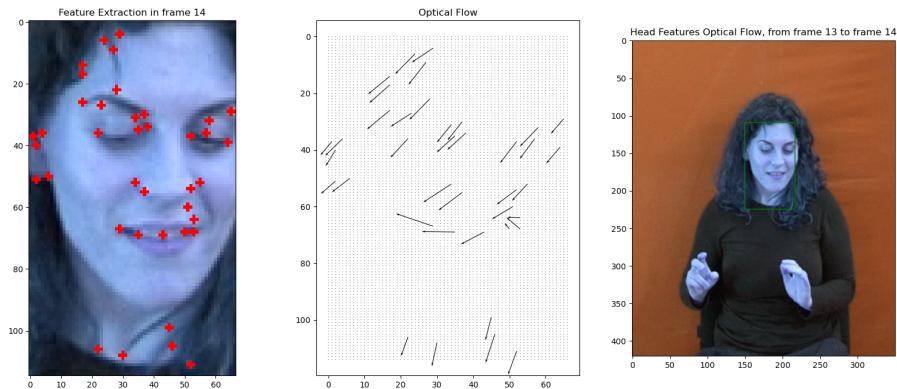
Έχοντας υπολογίσει την οπτική ροή της εικόνας  $I_n$  στα σημεία που ορίζουν τα σημεία ενδιαφέροντος εντός του bounding box της εικόνας  $I_{n1}$ , απομένει να βρούμε το συνολικό διάνυσμα μετατόπισης του bounding box ορθογωνίου, με όσο το δυνατόν μεγαλύτερη ακρίβεια. Γνωρίζουμε ότι τα διανύσματα οπτικής ροής έχουν κατά κανόνα μεγαλύτερο μήκος σε σημεία που ανήκουν σε περιοχές με έντονη πληροφορία υφής (π.χ. ακμές, κορυφές) και σχεδόν μηδενικό μήκος σε σημεία που ανήκουν σε περιοχές με ομοιόμορφη και επίπεδη υφή. Έτσι, καθώς η πλειονότητα των σημείων ενδιαφέροντος (γωνίες) βρίσκονται σε σημεία με έντονη υφή, θα μπορούσαμε απλώς να χρησιμοποιήσουμε τη μέση τιμή των διανυσμάτων μετατόπισης. Όμως, γιανα πετύχουμε καλύτερη ακρίβεια ή για να απορρίψουμε outliers, θα μπορούσαμε να εφαρμόσουμε εναλλακτικά κριτήρια, όπως για παράδειγμα να υπολογίσουμε τη μέση τιμή των διανυσμάτων μετατόπισης που έχουν ενέργεια μεγαλύτερη από μια τιμή κατωφλίου. Ως ενέργεια διανύσματος ταχύτητας ορίζουμε  $\|d\|^2 = d_x^2 + d_y^2$ .

Για τον υπολογισμό του  $d$  ορίσαμε τη συνάρτηση  $displ$  όπως περιγράφεται στην εξφώνηση.

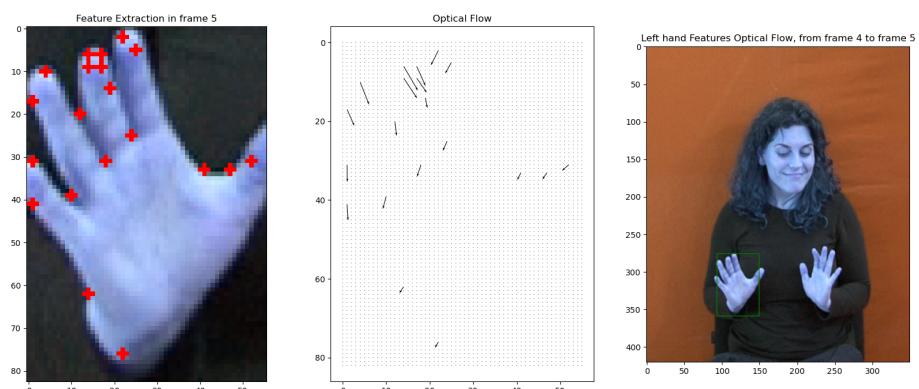
Ακολουθούν ενδεικτικά κάποιες φωτογραφίες από διάφορα στιγμιότυπα:



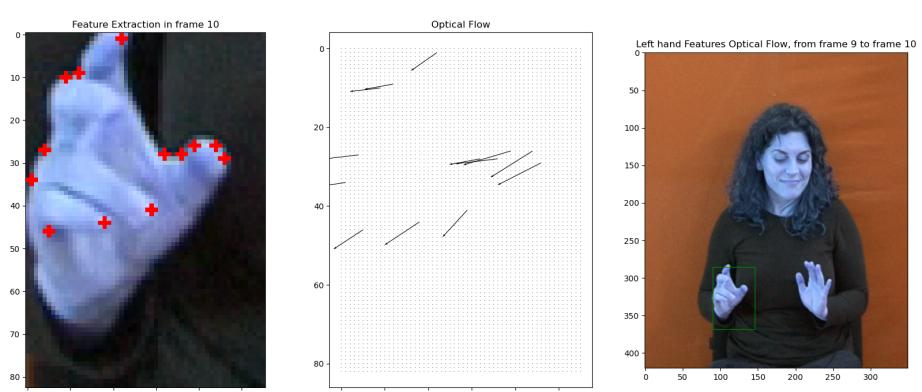
Σχήμα 6: Οπτική Ροή Μεταξύ Πρώτου και Δεύτερου Στιγμιούπου για το Κεφάλι



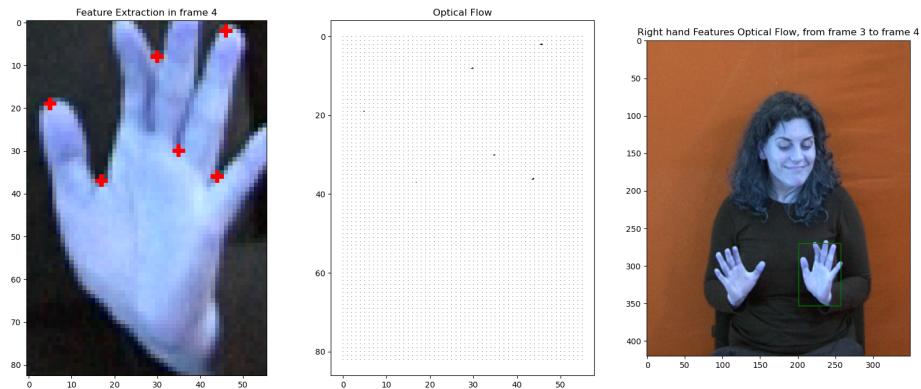
Σχήμα 7: Οπτική Ροή Μεταξύ 13ου και 14ου Στιγμιούπου για το Κεφάλι



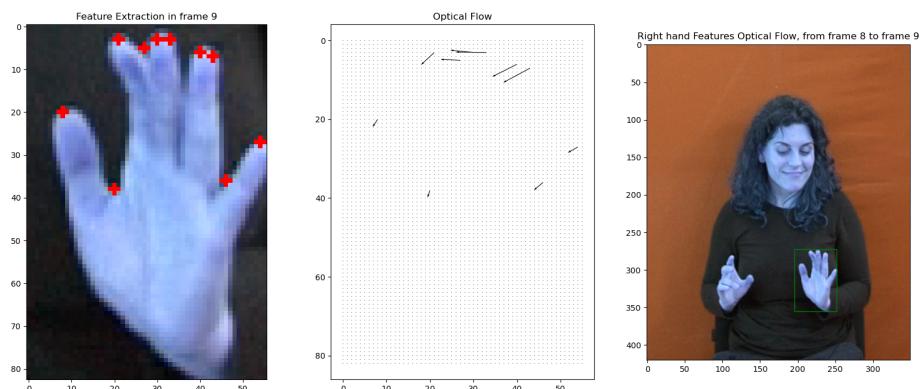
Σχήμα 8: Οπτική Ροή για το αριστερό χέρι



Σχήμα 9: Οπτική Ροή για το αριστερό χέρι

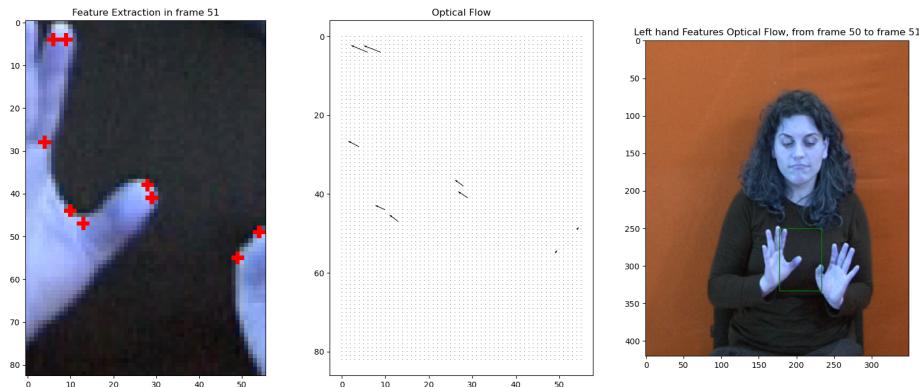


Σχήμα 10: Οπτική Ροή για το δεξιό χέρι



Σχήμα 11: Οπτική Ροή για το δεξιό χέρι

**Παρατηρήσεις:** Για κάποια frames βλέπουμε πως το bounding-box δεν ταυτίζεται με τα σημεία που θα έπρεπε να ακολουθεί, πράγμα το οποίο είναι αναμενόμενο και οφείλεται στην αδυναμία του αλγόριθμου των Lucas-Kanade να παρακολουθήσει μεγάλες και απότομες κινήσεις. Η ασυναμία αυτή παρατηρείται χυρώς σε κινήσεις του αριστερού χεριού καθώς αυτό είναι που κάνει τις πιο απότομες κινήσεις για το sequence του βίντεο που μας δόθηκε σε frames. Το δεξιό χέρι και το πρόσωπο εκτελούν μικρές κινήσεις και συνεπώς δεν παρατηρούμε περίεργη συμπεριφορά από τα bounding boxes.



Σχήμα 12: Εσφαλμένο Bounding Box για αριστερό χέρι

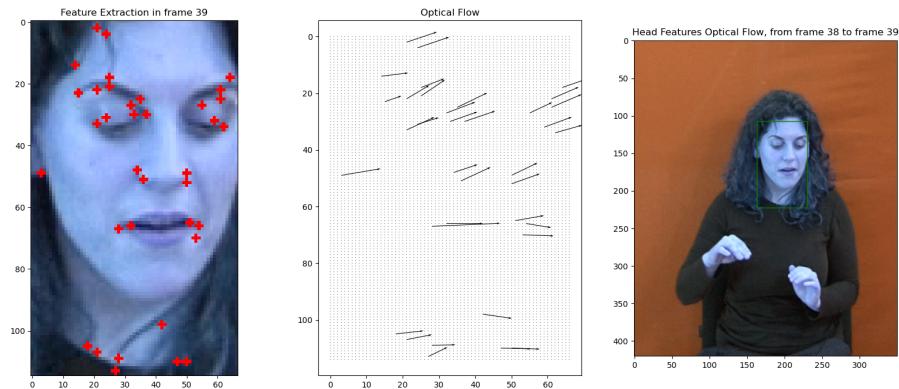
## Πολύ-κλιμακωτός Υπολογισμός Οπτικής Ροής

Η ιδέα του αλγόριθμου ιας κλίμακας των Lucas-Kanade είναι αποδοτική για μικρές κινήσεις. Ωστόσο, η ανάγκη για υπολογισμό της οπτικής ροής σε μεγαλύτερες κινήσεις που υπερβαίνουν κατά πολύ το 1 pixel είναι μεγάλη και οι παραδοχές που είχαν γίνει στον αλγόριθμο μιας κλίμακας για τους όρους πρώτης τάξης δεν ισχύουν. Η απάντηση στο πρόβλημα αυτό έρχεται από τον πολυκλιμακωτό αλγόριθμό των Lucas-Kanade.

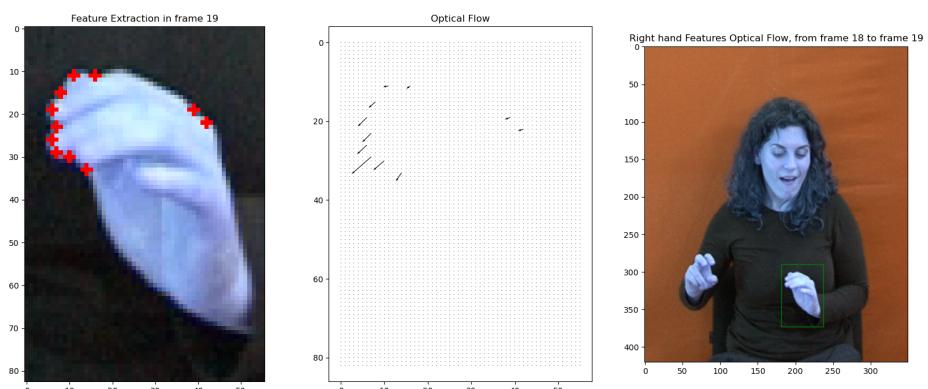
Η νέα εκδοχή του αλγόριθμου χρησιμοποιεί γκαουσιανές πυραμίδες και υπολογίζει την οπτική ροή από τις πιο μεγάλες σε κλίμακα (μικρές σε ανάλυση) εικόνες προς τις πιο μικρές σε κλίμακα (μεγάλες σε ανάλυση εικόνες). Αρχικά υπολογίζει την οπτική ροή στην κορυφαία κλίμακα χρησιμοποιώντας τον κλασικό αλγόριθμο Lucas-Kanade και το αποτέλεσμα του το χρησιμοποιούμε (έπειτα από bilinear interpolation) ως αρχική συνθήκη στον υπολογισμό της οπτικής ροής της επόμενης κλίμακας και συνεχίζεται η διαδικασία έως ότου υπολογίσουμε την οπτική ροήδ για την κανονικής κλίμακας εικόνα.

**Σημείωση:** Οι γκαουσιανές πυραμίδες προκύπτουν από την υποδειγματοληψία της εικόνας μειώνοντας την ανάλυση κάθε φορά στο μισό. Ωστόσο για να μετριαστεί το aliasing της εικόνας πριν την υποδειγματοληψία φιλτράρουμε την εικόνα με ένα βαθυπερατό φίλτρο, το οποίο στην περίπτωση μας είναι η γκαουσιανή με τυπική απόκλιση 3 πιξελ.

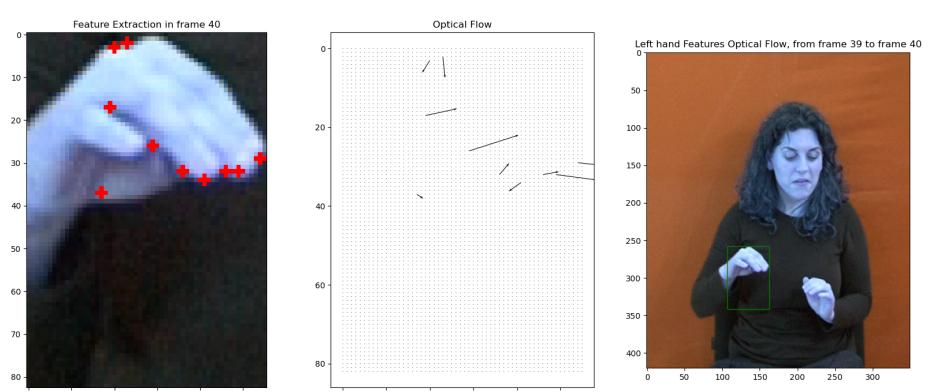
Ακολουθούν ενδεικτικά στιγμιότυπα



Σχήμα 13: Οπτική Ροή για το Κεφάλι



Σχήμα 14: Οπτική Ροή για το Δεξί Χέρι

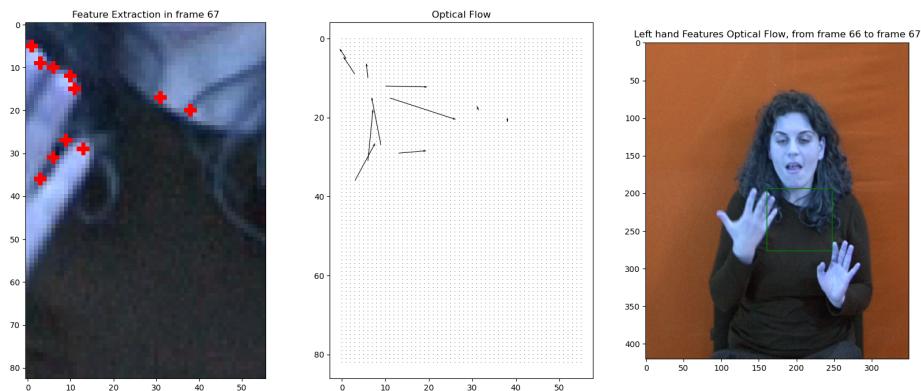


Σχήμα 15: Οπτική Ροή για το Αριστερό Χέρι

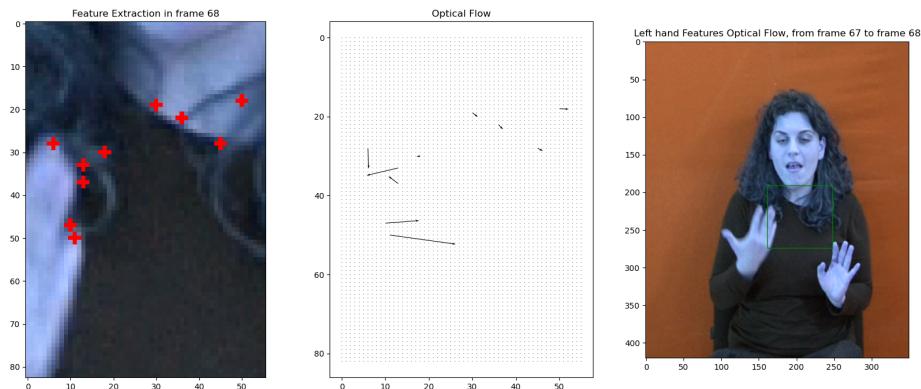
Όπως ποραναφέρθηκε, δεν έχουμε μεγάλες κινήσεις στο πρόσωπο και το δεξί χέρι, συνεπώς είναι άσκοπο να συγχρίνουμε τον μονοκλιμακωτό και τον πολυκλιμακωτό αλγόριθμο για εκείνα τα σημεία.

Ακολουθούν μερικά στιγμιότυπα για το αριστερό χέρι χρησιμοποιώντας μονοκλιμακωτό και πολυκλιμακωτό αλγόριθμο.

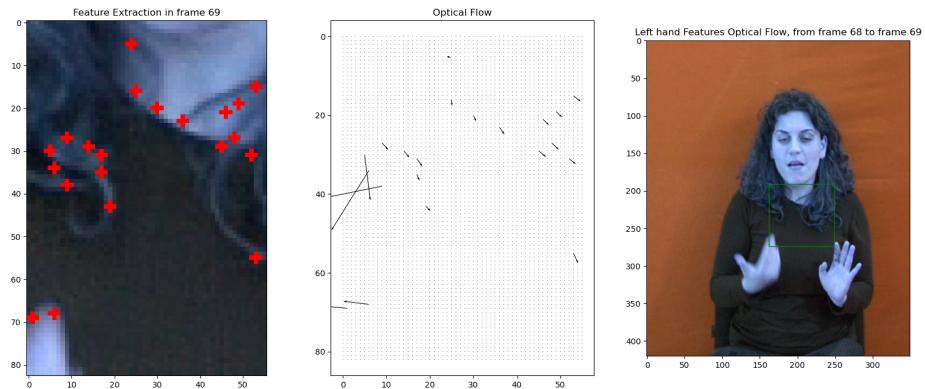
### Μονοκλιμακωτός:



Σχήμα 16: Μονοκλιμακωτή Οπτική Ροή για το Αριστερό Χέρι

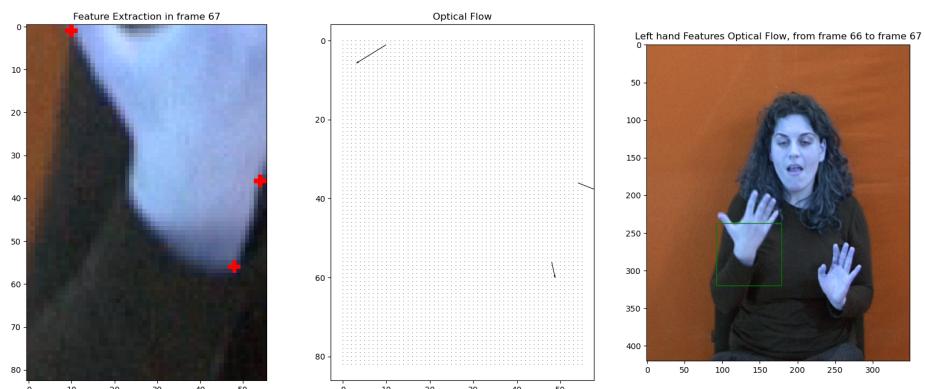


Σχήμα 17: Μονοκλιμακωτή Οπτική Ροή για το Αριστερό Χέρι

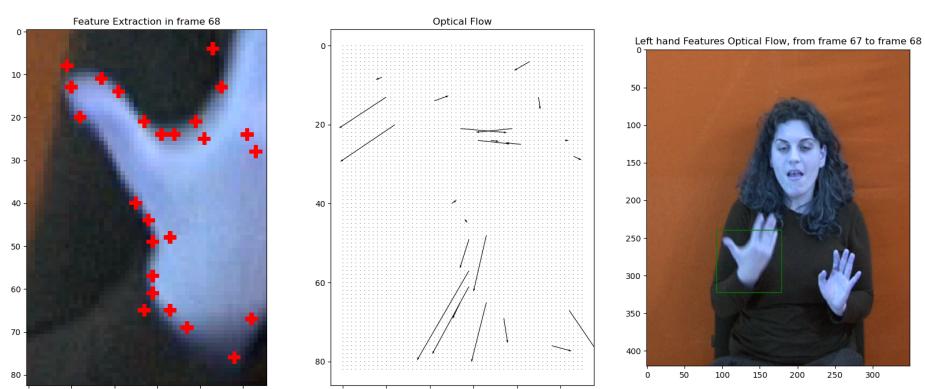


Σχήμα 18: Μονοκλιμακωτή Οπτική Ροή για το Αριστερό Χέρι

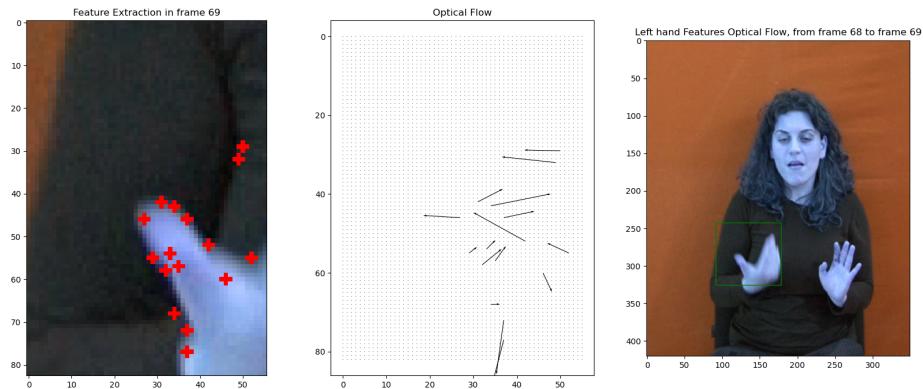
#### Πολυυκλιμακωτός:



Σχήμα 19: Πολυυκλιμακωτή Οπτική Ροή για το Αριστερό Χέρι



Σχήμα 20: Πολυυκλιμακωτή Οπτική Ροή για το Αριστερό Χέρι

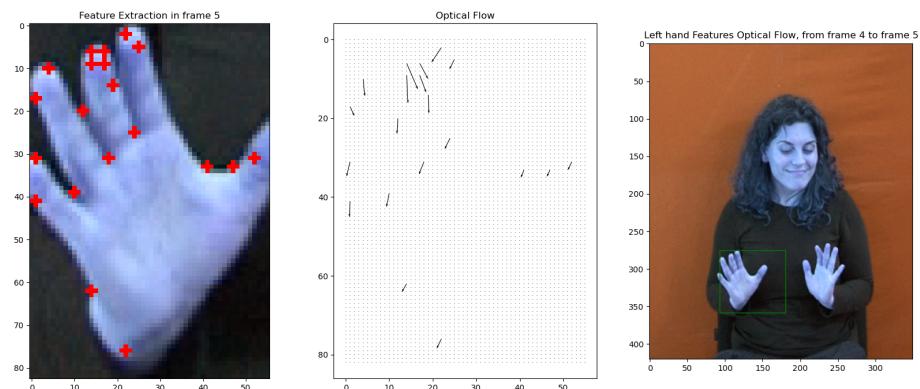


Σχήμα 21: Πολυκλιμακωτή Οπτική Ροή για το Αριστερό Χέρι

Η βελτίωση είναι εμφανής για την πολυκλιμακωτή εκδοχή του αλγορίθμου, εάν και όχι βέλτιστη.

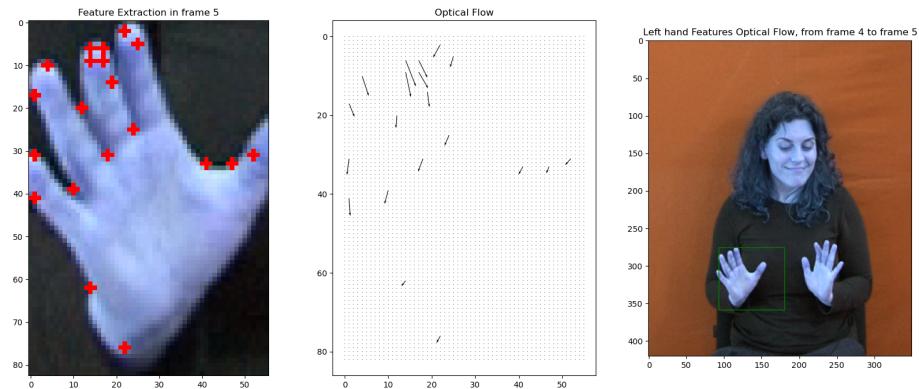
Παρακάτω θα μελετήσουμε πως άλλαζουν τα αποτελέσματα εάν καλέσουμε τη συνάρτηση lk με διαφορετικές παραμέτρους.

$$\text{Για } \rho = 1 \text{ epsilon} = 0.05 \text{ dx0} = 0, \text{ dy0} = 0$$



Σχήμα 22: Πολυκλιμακωτή Οπτική Ροή για το Αριστερό Χέρι

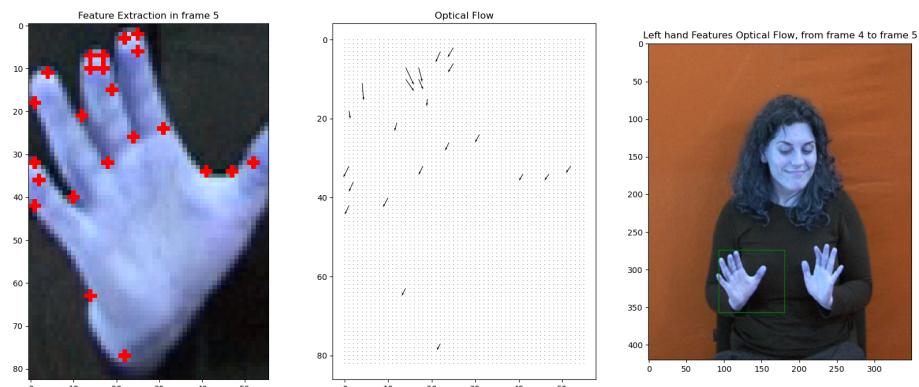
$$\text{Για } \rho = 2 \text{ epsilon} = 0.05 \text{ dx0} = 0, \text{ dy0} = 0$$



Σχήμα 23: Πολυκλιμακωτή Οπτική Ροή για το Αριστερό Χέρι

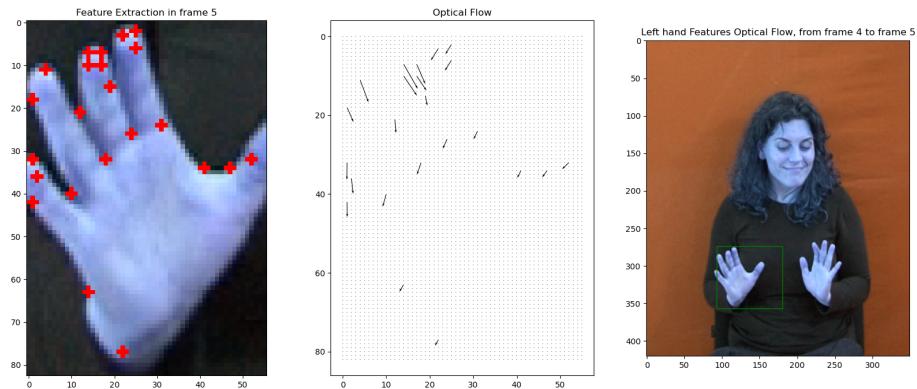
Παρατηρούμε πως με τη μείωση του  $\rho$  μεγαλώνει το bounding box και παρατηρούμε μικρότερη ευχρίνεια στην οπτική ροή.

Για  $\rho = 4$   $\text{epsilon} = 1$   $\text{dx0} = 0$ ,  $\text{dy0} = 0$



Σχήμα 24: Πολυκλιμακωτή Οπτική Ροή για το Αριστερό Χέρι

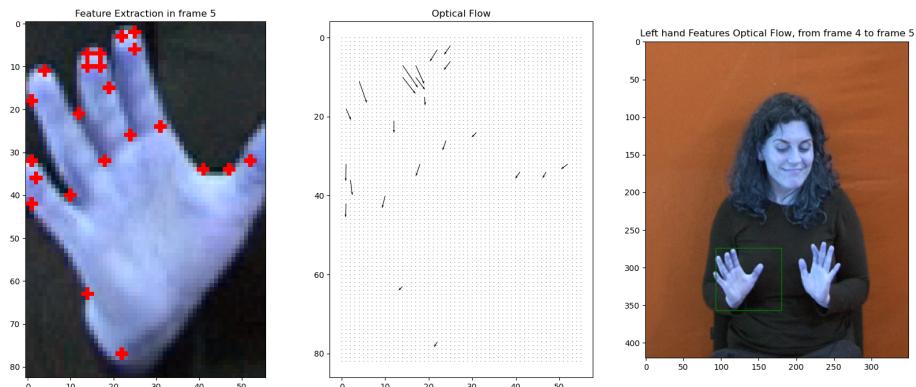
Για  $\rho = 4$   $\text{epsilon} = 0.075$   $\text{dx0} = 0$ ,  $\text{dy0} = 0$



Σχήμα 25: Πολυκλιμακωτή Οπτική Ροή για το Αριστερό Χέρι

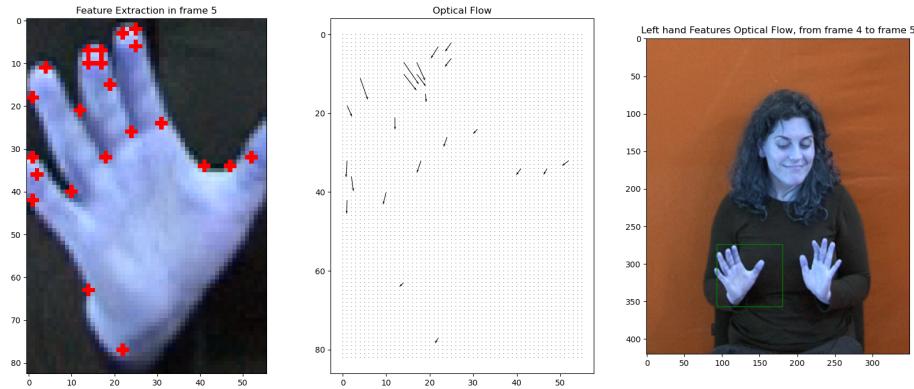
Η εν λόγω σταθερά έχει σκοπό να βελτιώσει το αποτέλεσμα σε επίπεδες περιοχές με μειωμένη υφή και άρα μειωμένη πληροφορία για τον υπολογισμό της οπτικής ροής. Άρα, αναμένουμε ότι αύξηση της σταθεράς θα οδηγεί σε πιο “ευσταθή” αποτελέσματα. τραςχ(‘Λεφτ’ ηανδ’, Φαλσε, 4, 0.05, 0, 0, 5)

Για  $\rho = 4$  epsilon = 0.05 dx0 = 0, dy0 = 0 και 5 κλίμακες



Σχήμα 26: Πολυκλιμακωτή Οπτική Ροή για το Αριστερό Χέρι

Για  $\rho = 4$  epsilon = 0.05 dx0 = 0, dy0 = 0 και 6 κλίμακες



Σχήμα 27: Πολυκλιμακωτή Οπτική Ροή για το Αριστερό Χέρι

Οι διαφορές είναι αμελητέες καθώς έχουμε μικρές κινήσεις

## Χωροχρονικά Σημεία Ενδιαφέροντος

Ο ανιχνευτής Harris αποτελεί μία προέκταση σε 3 διαστάσεις του ανιχνευτή γωνιών Harris-Stephens που υλοποιήσαμε στην 1η εργαστηριακή άσκηση προσθέτοντας έναν  $2\Delta$  τανυστή και τη χρονική παράγωγο. Αρχικά για κάθε voxel του βίντεο υπολογίζουμε τον ακόλουθο πίνακα

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

Στην ουσία ομαλοποιούμε χωροχρονικά με γκαουσιανά φιλτραρίσματα (χωρική κλίμακα σ και χρονική κλίμακα τ) το βίντεο. Τις παραγώγους τις υπολογίζουμε εφαρμόζοντας συνέλιξη με τον πυρήνα κεντρικών διαφορών  $[-101]^T$  προσαρμοσμένο κάθε φορά στην κατάλληλη διάσταση.

Τέλος, χρησιμοποιήθηκε το εξής κριτήριο γωνιότητας:

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

Ο ανιχνευτής Gabor βασίζεται στο χρονικό φιλτράρισμα του βίντεο με ένα ζεύγος Gabor φίλτρων αφού πρώτα αυτό έχει υποστεί εξομάλυνση στις χωρικές διαστάσεις μέσω ενός  $2\Delta$  γκαουσιανού πυρήνα  $g(x, y; \sigma)$  με τυπική απόκλιση  $\sigma$ . Τα /τεξτλατιν-Γαβορ φίλτρα ορίζονται ως:

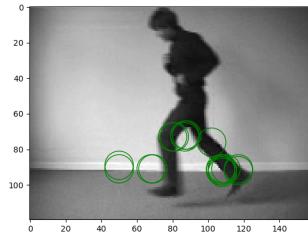
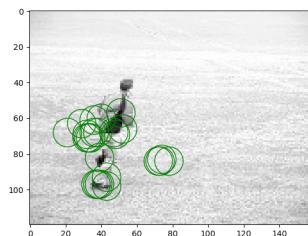
$$h_{ev}(t; \tau, \omega) = \cos(2\pi t\omega) \exp\left(-\frac{t^2}{2\tau^2}\right)$$

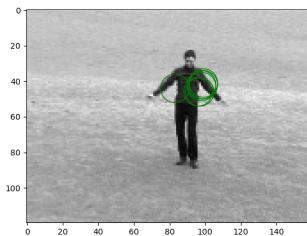
$$h_{od}(t; \tau, \omega) = \sin(2\pi t\omega) \exp\left(-\frac{t^2}{2\tau^2}\right)$$

Η συχνότητα ω του Gabor φίλτρου συνδέεται με την χρονική αλίμακα  $\tau$  (απόκλιση της γκαουσιανής συνιστώσας του) μέσω της σχέσης:  $\omega = 4/\tau$ . Το κριτήριο σημαντικότητας προκύπτει παίρνοντας την τετραγωνική ενέργεια της εξόδου για το ζεύγος Gabor φίλτρων:

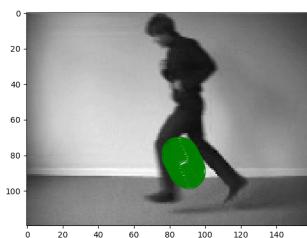
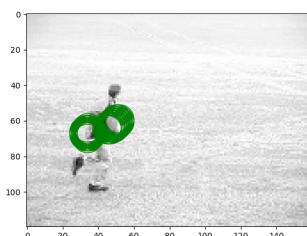
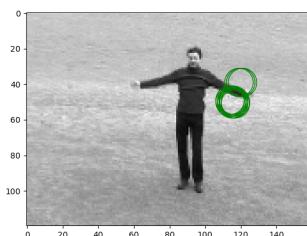
$$H(x, y, t) = (I(x, y, t) * g * h_{ev})^2 + (I(x, y, t) * g * h_{od})^2$$

Παρατηρώντας όλα τα εξαγόμενα βίντεο με τα εντοπισμένα σημεία ενδιαφέροντος βλέπουμε πως ο ανιχνευτής Harris εντοπίζει σχεδόν κάθε φορά την κίνηση στα βίντεο ενώ ο ανιχνευτής Gabor εντοπίζει ένα μικρό μέρος των κινήσεων. Ωστόσο ο Gabor ανιχνευτής είναι πιο εύστοχος από τον Harris. Αυτό συμβαίνει καθώς ο ανιχνευτής H-αρρις εντοπίζει κυρίως γωνίες στα frame των βίντεο. Χαρακτηριστικό παράδειγμα είναι πως εντοπίζει τη γωνία εντός των ποδιών στο τρέξιμο και το περπάτημα και την άκρη του χεριού στο wave. Αντίθετα στον Gabor ανιχνευτή τα σημεία που εντοπίζονται είναι πιο στοχευμένα και αυτό φαίνεται και από τα παραπάνω παραδείγματα όπου οι πράσινοι κύκλοι είναι πιο έντονοι, δηλαδή πολλά εντοπίζονται πολλά σημεία στην γειτονιά που όντως υπάρχει κίνηση. **Harris**





Gabor:



## 0.1 Χωρο-χρονικοί Ιστογραφικοί Περιγραφητές

Αρχικά υπολογίζουμε το διάνυσμα κλίσης και την TLV1 οπτική ροή σε κάθε pixel που θα χρειαστούν για τους περιγραφητές. Οι περιγραφητές υπολογίζονται με την

δοθείσα συνάρτηση orientation\_histogram η οποία δέχεται ως είσοδο τα Gx, Gy που ορίζουν ένα διανυσματικό πεδίο. Για τον HOG τα Gx, Gy υπολογίζονται από το διάνυσμα κλίσης αλλά για τον HOF απαιτείται ένα πυκνό πεδίο οπτικής ροής και για αυτό πάφρονται τα Gx, Gy από την TLV1 οπτική ροή. Ο περιγραφητής συνδυασμού των HOG/HOF προκύπτει με συνένωση των δύο.

## Κατασκευή Bag of Visual Words και χρήση Support Vector Machines για ταξινόμηση δράσεων

Για τον διαχωρισμό των αρχείων σε δύο σύνολα (εκπαίδευσης και δοκιμής) με αντίστοιχα λαβελς χρησιμοποιήσαμε κώδικα που παίρνει ως είσοδο ένα αρχείο κειμένου ώστε εύκολα κανείς να μπορεί να αλλάξει τον διαμοιρασμό

ακολουθούν για κάθε συνδυασμό ανιχνευτών/περιγραφητών η αρρίβεια και αναλυτικοί πίνακας που δείχνουν ως τι αναγνωρίστηκε κάθε δράσης.

---

<b>detector:</b>	<b>Harris</b>																																																																																	
<b>descriptor:</b>	<b>HOG</b>																																																																																	
The accuracy for the combination is 26.087%																																																																																		
Here is a table with detailed results:																																																																																		
<table border="0"> <thead> <tr> <th>Video Name</th> <th>Correct</th> <th>Prediction</th> </tr> </thead> <tbody> <tr><td>person07_running_d4_uncomp.avi</td><td>walk</td><td>run      False</td></tr> <tr><td>person12_running_d2_uncomp.avi</td><td>walk</td><td>run      False</td></tr> <tr><td>person11_running_d1_uncomp.avi</td><td>walk</td><td>run      False</td></tr> <tr><td>person14_running_d4_uncomp.avi</td><td>walk</td><td>run      False</td></tr> <tr><td>person18_running_d4_uncomp.avi</td><td>walk</td><td>run      False</td></tr> <tr><td>person15_running_d1_uncomp.avi</td><td>walk</td><td>run      False</td></tr> <tr><td>person20_running_d3_uncomp.avi</td><td>walk</td><td>run      False</td></tr> <tr><td>person25_running_d4_uncomp.avi</td><td>walk</td><td>run      False</td></tr> <tr><td>person16_running_d2_uncomp.avi</td><td>walk</td><td>run      False</td></tr> <tr><td>person19_running_d1_uncomp.avi</td><td>walk</td><td>run      False</td></tr> <tr><td>person17_running_d3_uncomp.avi</td><td>walk</td><td>run      False</td></tr> <tr><td>person13_running_d3_uncomp.avi</td><td>walk</td><td>run      False</td></tr> <tr><td>person05_running_d2_uncomp.avi</td><td>walk</td><td>run      False</td></tr> <tr><td>person04_running_d1_uncomp.avi</td><td>walk</td><td>run      False</td></tr> <tr><td>person08_running_d2_uncomp.avi</td><td>walk</td><td>run      False</td></tr> <tr><td>person07_handwaving_d3_uncomp.avi</td><td></td><td>handwaving      handwaving</td></tr> <tr><td>True</td><td></td><td></td></tr> <tr><td>person11_handwaving_d2_uncomp.avi</td><td></td><td>handwaving      handwaving</td></tr> <tr><td>True</td><td></td><td></td></tr> <tr><td>person06_handwaving_d1_uncomp.avi</td><td></td><td>handwaving      handwaving</td></tr> <tr><td>True</td><td></td><td></td></tr> <tr><td>person15_handwaving_d4_uncomp.avi</td><td></td><td>handwaving      run      False</td></tr> <tr><td>person25_running_d4_uncomp.avi</td><td>run</td><td>run      True</td></tr> <tr><td>person24_running_d3_uncomp.avi</td><td>run</td><td>run      True</td></tr> <tr><td>person10_running_d2_uncomp.avi</td><td>run</td><td>handwaving      False</td></tr> <tr><td>person20_running_d1_uncomp.avi</td><td>run</td><td>run      True</td></tr> </tbody> </table>		Video Name	Correct	Prediction	person07_running_d4_uncomp.avi	walk	run      False	person12_running_d2_uncomp.avi	walk	run      False	person11_running_d1_uncomp.avi	walk	run      False	person14_running_d4_uncomp.avi	walk	run      False	person18_running_d4_uncomp.avi	walk	run      False	person15_running_d1_uncomp.avi	walk	run      False	person20_running_d3_uncomp.avi	walk	run      False	person25_running_d4_uncomp.avi	walk	run      False	person16_running_d2_uncomp.avi	walk	run      False	person19_running_d1_uncomp.avi	walk	run      False	person17_running_d3_uncomp.avi	walk	run      False	person13_running_d3_uncomp.avi	walk	run      False	person05_running_d2_uncomp.avi	walk	run      False	person04_running_d1_uncomp.avi	walk	run      False	person08_running_d2_uncomp.avi	walk	run      False	person07_handwaving_d3_uncomp.avi		handwaving      handwaving	True			person11_handwaving_d2_uncomp.avi		handwaving      handwaving	True			person06_handwaving_d1_uncomp.avi		handwaving      handwaving	True			person15_handwaving_d4_uncomp.avi		handwaving      run      False	person25_running_d4_uncomp.avi	run	run      True	person24_running_d3_uncomp.avi	run	run      True	person10_running_d2_uncomp.avi	run	handwaving      False	person20_running_d1_uncomp.avi	run	run      True
Video Name	Correct	Prediction																																																																																
person07_running_d4_uncomp.avi	walk	run      False																																																																																
person12_running_d2_uncomp.avi	walk	run      False																																																																																
person11_running_d1_uncomp.avi	walk	run      False																																																																																
person14_running_d4_uncomp.avi	walk	run      False																																																																																
person18_running_d4_uncomp.avi	walk	run      False																																																																																
person15_running_d1_uncomp.avi	walk	run      False																																																																																
person20_running_d3_uncomp.avi	walk	run      False																																																																																
person25_running_d4_uncomp.avi	walk	run      False																																																																																
person16_running_d2_uncomp.avi	walk	run      False																																																																																
person19_running_d1_uncomp.avi	walk	run      False																																																																																
person17_running_d3_uncomp.avi	walk	run      False																																																																																
person13_running_d3_uncomp.avi	walk	run      False																																																																																
person05_running_d2_uncomp.avi	walk	run      False																																																																																
person04_running_d1_uncomp.avi	walk	run      False																																																																																
person08_running_d2_uncomp.avi	walk	run      False																																																																																
person07_handwaving_d3_uncomp.avi		handwaving      handwaving																																																																																
True																																																																																		
person11_handwaving_d2_uncomp.avi		handwaving      handwaving																																																																																
True																																																																																		
person06_handwaving_d1_uncomp.avi		handwaving      handwaving																																																																																
True																																																																																		
person15_handwaving_d4_uncomp.avi		handwaving      run      False																																																																																
person25_running_d4_uncomp.avi	run	run      True																																																																																
person24_running_d3_uncomp.avi	run	run      True																																																																																
person10_running_d2_uncomp.avi	run	handwaving      False																																																																																
person20_running_d1_uncomp.avi	run	run      True																																																																																

```

detector: Harris
descriptor: HOF

The accuracy for the combination is 17.391%

Here is a table with detailed results:

  Video Name          Correct Prediction
person07_running_d4_uncomp.avi   walk   handwaving   False
person12_running_d2_uncomp.avi   walk   handwaving   False
person11_running_d1_uncomp.avi   walk   handwaving   False
person14_running_d4_uncomp.avi   walk   handwaving   False
person18_running_d4_uncomp.avi   walk   handwaving   False
person15_running_d1_uncomp.avi   walk   handwaving   False
person20_running_d3_uncomp.avi   walk   handwaving   False
person25_running_d4_uncomp.avi   walk   handwaving   False
person16_running_d2_uncomp.avi   walk   handwaving   False
person19_running_d1_uncomp.avi   walk   handwaving   False
person17_running_d3_uncomp.avi   walk   handwaving   False
person13_running_d3_uncomp.avi   walk   handwaving   False
person05_running_d2_uncomp.avi   walk   handwaving   False
person04_running_d1_uncomp.avi   walk   handwaving   False
person08_running_d2_uncomp.avi   walk   handwaving   False
person07_handwaving_d3_uncomp.avi   handwaving   handwaving
True                           handwaving   handwaving
person11_handwaving_d2_uncomp.avi   handwaving   handwaving
True                           handwaving   handwaving
person06_handwaving_d1_uncomp.avi   handwaving   handwaving
True                           handwaving   handwaving
person15_handwaving_d4_uncomp.avi   handwaving   handwaving
True                           handwaving   handwaving
person25_running_d4_uncomp.avi   run    handwaving   False
person24_running_d3_uncomp.avi   run    handwaving   False
person10_running_d2_uncomp.avi   run    handwaving   False
person20_running_d1_uncomp.avi   run    handwaving   False

```

```

detector:      Harris
descriptor:    HOF and HOG

The accuracy for the combination is 30.435%

Here is a table with detailed results:

  Video Name          Correct Prediction
person07_running_d4_uncomp.avi   walk   run   False
person12_running_d2_uncomp.avi   walk   run   False
person11_running_d1_uncomp.avi   walk   run   False
person14_running_d4_uncomp.avi   walk   run   False
person18_running_d4_uncomp.avi   walk   run   False
person15_running_d1_uncomp.avi   walk   run   False
person20_running_d3_uncomp.avi   walk   run   False
person25_running_d4_uncomp.avi   walk   run   False
person16_running_d2_uncomp.avi   walk   run   False
person19_running_d1_uncomp.avi   walk   run   False
person17_running_d3_uncomp.avi   walk   run   False
person13_running_d3_uncomp.avi   walk   run   False
person05_running_d2_uncomp.avi   walk   run   False
person04_running_d1_uncomp.avi   walk   run   False
person08_running_d2_uncomp.avi   walk   run   False
person07_handwaving_d3_uncomp.avi handwaving   handwaving
True                           handwaving   handwaving
person11_handwaving_d2_uncomp.avi handwaving   handwaving
True                           handwaving   handwaving
person06_handwaving_d1_uncomp.avi handwaving   handwaving
True                           handwaving   handwaving
person15_handwaving_d4_uncomp.avi handwaving   handwaving
True                           run     run     True
person25_running_d4_uncomp.avi   run     run     True
person24_running_d3_uncomp.avi   run     run     True
person10_running_d2_uncomp.avi   run     handwaving   False
person20_running_d1_uncomp.avi   run     run     True

```

```

detector:      Gabor
descriptor:    HOG

The accuracy for the combination is 26.087%

Here is a table with detailed results:

Video Name      Correct Prediction
person07_running_d4_uncomp.avi   walk   run   False
person12_running_d2_uncomp.avi   walk   run   False
person11_running_d1_uncomp.avi   walk   run   False
person14_running_d4_uncomp.avi   walk   run   False
person18_running_d4_uncomp.avi   walk   run   False
person15_running_d1_uncomp.avi   walk   run   False
person20_running_d3_uncomp.avi   walk   run   False
person25_running_d4_uncomp.avi   walk   run   False
person16_running_d2_uncomp.avi   walk   run   False
person19_running_d1_uncomp.avi   walk   handwaving  False
person17_running_d3_uncomp.avi   walk   run   False
person13_running_d3_uncomp.avi   walk   handwaving  False
person05_running_d2_uncomp.avi   walk   run   False
person04_running_d1_uncomp.avi   walk   run   False
person08_running_d2_uncomp.avi   walk   run   False
person07_handwaving_d3_uncomp.avi   handwaving  handwaving
True
person11_handwaving_d2_uncomp.avi   handwaving  handwaving
True
person06_handwaving_d1_uncomp.avi   handwaving  handwaving
True
person15_handwaving_d4_uncomp.avi   handwaving  handwaving
True
person25_running_d4_uncomp.avi   run   run   True
person24_running_d3_uncomp.avi   run   handwaving  False
person10_running_d2_uncomp.avi   run   handwaving  False
person20_running_d1_uncomp.avi   run   run   True

```

**detector:** Gabor  
**descriptor:** HOF

The accuracy for the combination is 17.391%

Here is a table with detailed results:

Video Name	Correct Prediction
person07_running_d4_uncomp.avi	walk run False
person12_running_d2_uncomp.avi	walk run False
person11_running_d1_uncomp.avi	walk run False
person14_running_d4_uncomp.avi	walk run False
person18_running_d4_uncomp.avi	walk run False
person15_running_d1_uncomp.avi	walk run False
person20_running_d3_uncomp.avi	walk run False
person25_running_d4_uncomp.avi	walk run False
person16_running_d2_uncomp.avi	walk run False
person19_running_d1_uncomp.avi	walk run False
person17_running_d3_uncomp.avi	walk run False
person13_running_d3_uncomp.avi	walk run False
person05_running_d2_uncomp.avi	walk run False
person04_running_d1_uncomp.avi	walk run False
person08_running_d2_uncomp.avi	walk run False
person07_handwaving_d3_uncomp.avi	handwaving run False
person11_handwaving_d2_uncomp.avi	handwaving run False
person06_handwaving_d1_uncomp.avi	handwaving run False
person15_handwaving_d4_uncomp.avi	handwaving run False
person25_running_d4_uncomp.avi	run run True
person24_running_d3_uncomp.avi	run run True
person10_running_d2_uncomp.avi	run run True
person20_running_d1_uncomp.avi	run run True

```
detector:      Gabor
descriptor:    HOF and HOG
```

The accuracy for the combination is 26.08%

Here is a table with detailed results:

Video Name	Correct	Prediction	
person07_running_d4_uncomp.avi	walk	run	False
person12_running_d2_uncomp.avi	walk	run	False
person11_running_d1_uncomp.avi	walk	run	False
person14_running_d4_uncomp.avi	walk	run	False
person18_running_d4_uncomp.avi	walk	run	False
person15_running_d1_uncomp.avi	walk	run	False
person20_running_d3_uncomp.avi	walk	run	False
person25_running_d4_uncomp.avi	walk	run	False
person16_running_d2_uncomp.avi	walk	run	False
person19_running_d1_uncomp.avi	walk	handwaving	False
person17_running_d3_uncomp.avi	walk	handwaving	False
person13_running_d3_uncomp.avi	walk	handwaving	False
person05_running_d2_uncomp.avi	walk	run	False
person04_running_d1_uncomp.avi	walk	run	False
person08_running_d2_uncomp.avi	walk	run	False
person07_handwaving_d3_uncomp.avi		handwaving	handwaving
True			
person11_handwaving_d2_uncomp.avi		handwaving	handwaving
True			
person06_handwaving_d1_uncomp.avi		handwaving	handwaving
True			
person15_handwaving_d4_uncomp.avi		handwaving	handwaving
True			
person25_running_d4_uncomp.avi	run	run	True
person24_running_d3_uncomp.avi	run	run	True
person10_running_d2_uncomp.avi	run	handwaving	False
person20_running_d1_uncomp.avi	run	handwaving	False

Παρατηρούμε λοιπόν πως γενικά οι περιγραφητές μόνοι τους δεν πετυχαίνουν ιδιαίτερα υψηλά ποσοστά αλλά ο συνδυασμός των δύο βγάζει ψηλότερο γενικά accuracy

Τα αποτελέσματα αυτά δεν είναι λογικά, θα έπρεπε προφανώς να είναι υψηλότερα ωστόσο ο υπολογιστής μου αποφάσισε να φύγει από το μάταιο τούτο κόσμο τις τελευταίες δύο μέρες συνεπώς δε μπορώ να παραθέσω εικόνες. Σε κάθε περίπτωση από προηγούμενο τρέξιμο το οποίο για άγνωστο λόγο δε μπήκα στον κόπο να σώσω αν όμως μαζι μαζι καλά είχαμε γύρω στο 0.75 για το Harris και 0.58 για το Gabor.