

Εθνικό Μετσόβιο Πολυτεχνείο

Σχολή Ηλεκτρολόγων Μηχανικών & Μηχανικών Υπολογιστών

Ακαδημαϊκό Έτος: 2019-2020



ΟΡΑΣΗ ΥΠΟΛΟΓΙΣΤΩΝ

2η Εργαστηριακή Άσκηση

Θέμα: Εκτίμηση Οπτικής Ροής (Optical Flow)

Τζε Χριστίνα-Ουρανία | 03116079
Ψαρουδάκης Ανδρέας | 03116001

27 Μαΐου 2020

Εισαγωγή

Σκοπός του πρώτου μέρους της δεύτερης εργαστηριακής άσκησης είναι η υλοποίηση ενός συστήματος παρακολούθησης προσώπου και χεριών (Face and Hands Tracking) σε μια ακολουθία βίντεο νοηματικής γλώσσας. Αρχικά, γίνεται εντοπισμός των περιοχών ενδιαφέροντος (περιοχές προσώπου και χεριών) στο πρώτο πλαίσιο της ακολουθίας, με την χρήση ενός πιθανοτικού ανιχνευτή ανθρώπινου δέρματος. Στη συνέχεια, εισάγεται η έννοια του διανυσματικού πεδίου της οπτικής ροής, το οποίο υπολογίζεται με χρήση του αλγορίθμου των Lucas-Kanade και χρησιμοποιείται για την παρακολούθηση των περιοχών αυτών μεταξύ διαδοχικών εικόνων (καρέ) του βίντεο.

Μέρος 1: Παρακολούθηση Προσώπου και Χεριών με Χρήση του Πεδίου Οπτικής Ροής με τη μέθοδο Lukas-Kanade

1.1 Ανίχνευση Δέρματος Προσώπου και Χεριών (Bonus Ερώτημα)

Απαραίτητη προϋπόθεση για την παρακολούθηση του προσώπου και των χεριών αποτελεί ο εντοπισμός τους στο πρώτο πλαίσιο της δοσμένης ακολουθίας εικόνων βίντεο. Αυτό γίνεται με την απομόνωση των περιοχών εκείνων που έχουν το χρώμα του δέρματος. Για τον σκοπό αυτό χρησιμοποιείται ο χρωματικός χώρος YCbCr και συγκεκριμένα τα κανάλια Cb, Cr καθώς το Y κανάλι σχετίζεται με την φωτεινότητα και όχι με την ταυτότητα του χρώματος και για τον λόγο αυτό δεν χρησιμοποιείται στην συγκεκριμένη εφαρμογή. Η μοντελοποίηση του χρώματος του δέρματος γίνεται με την ακόλουθη δισδιάστατη Γκαουσιανή κατανομή:

$$P(\mathbf{c} = skin) = \frac{1}{\sqrt{|\Sigma|} (2\pi)^2} e^{-\frac{1}{2}(\mathbf{c}-\mu)\Sigma^{-1}(\mathbf{c}-\mu)'} \quad (1)$$

την οποία εκπαιδεύουμε αρχικά με τα δείγματα δέρματος του συμπληρωματικού αρχείου skinSamplesRGB.mat, αφού γίνει μετατροπή τους με χρήση της συνάρτησης **cvtColor** της cv2 στον κατάλληλο χρωματικό χώρο. Όπως φαίνεται και στην (1) για την εκπαίδευση της Γκαουσιανής απαιτείται ο υπολογισμός του 1×2 διανύσματος \mathbf{c} των τιμών Cb και Cr σε κάθε σημείο της εικόνας καθώς επίσης και του ιδίων διαστάσεων διανύσματος μ των μέσων τιμών των καναλιών Cb και Cr για τα δείγματα δέρματος. Για την εξαγωγή των μέσων τιμών χρησιμοποιούμε την έτοιμη συνάρτηση **mean** της βιβλιοθήκης numpy. Επιπλέον, απαιτείται η εύρεση του 2×2 πίνακα συνδιακύμανσης Σ των δύο καναλιών, τον οποίο υπολογίζουμε με χρήση της συνάρτησης **cov** της numpy. Ο υπολογισμός της ορίζουσας του Σ καθώς και του αντιστρόφου του, Σ^{-1} , γίνεται με τις συναρτήσεις **det** και **inv** αντίστοιχα του πακέτου numpy.linalg. Στην συνέχεια κανονικοποιούμε την $P(\mathbf{c}=skin)$ διαιρώντας με την μέγιστη τιμή της ώστε να λαμβάνει τιμές πιθανότητας στο διάστημα $[0,1]$. Η δυαδική εικόνα ανίχνευσης δέρματος προκύπτει τελικά από την εικόνα πιθανοτήτων P με κατάλληλη κατωφλιοποίηση.

Πειραματιζόμενοι με τις ενδεικτικές τιμές κατωφλίου στο διάστημα $[0.05, 0.25]$ καταλήγουμε τελικά στα καλύτερα αποτελέσματα ανίχνευσης για τιμή ίση με 0.06. Το πρώτο πλαίσιο της ακολουθίας εικόνων καθώς επίσης και η δυαδική εικόνα ανίχνευσης δέρματος κατά τα διάφορα στάδια της επεξεργασίας της φαίνονται στα ακόλουθα σχήματα:



(α') Αρχικό πλαίσιο



(β') Δυαδική εικόνα δέρματος πριν τη μορφολογική επεξεργασία



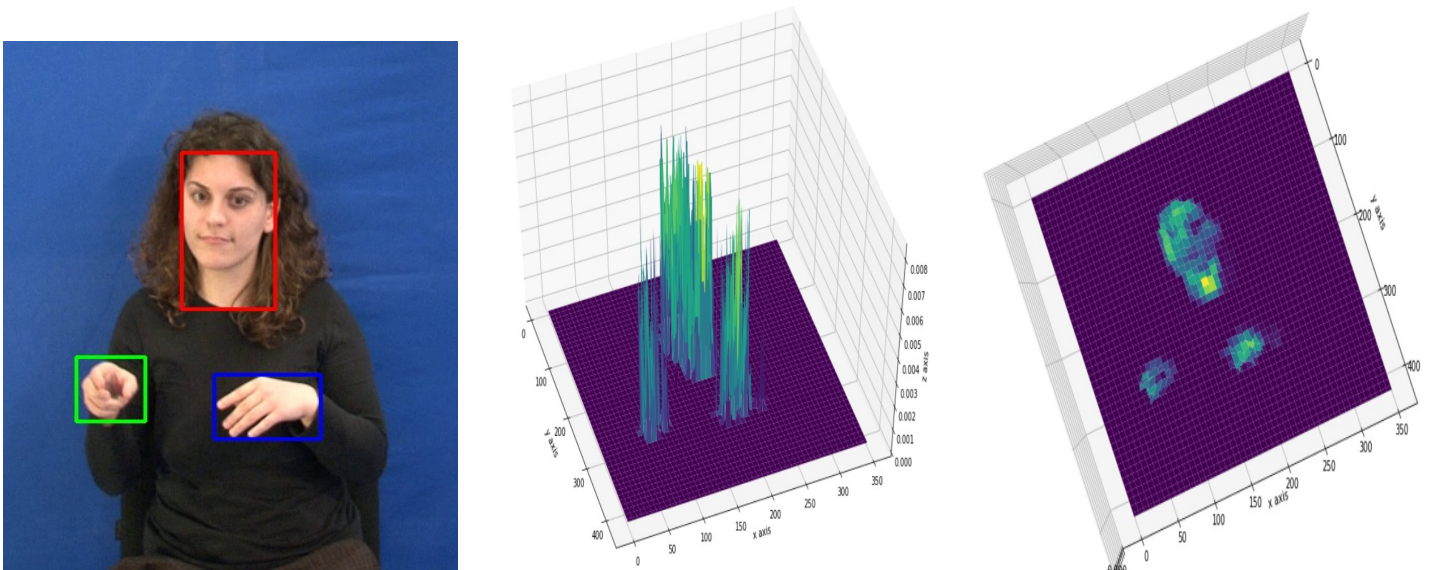
(γ') Δυαδική εικόνα δέρματος μετά το opening



(δ') Τελική ανίχνευση προσώπου και χεριών

Παρατηρούμε πως πριν το μορφολογικό φιλτράρισμα το αποτέλεσμα που λαμβάνουμε είναι ικανοποιητικό αλλά απαιτείται μία επιπλέον επεξεργασία της δυαδικής εικόνας για την τελική ανίχνευση των περιοχών δέρματος του προσώπου και των χεριών. Συγκεκριμένα, χρειάζεται να γίνει μία κάλυψη των τρυπών που εμφανίζονται. Για να το πετύχουμε αυτό εφαρμόζουμε αρχικά opening με ένα πολύ μικρό δομικό στοιχείο (δίσκο μεγέθους 1) και στη συνέχεια closing με ένα μεγάλο δομικό στοιχείο (δίσκο μεγέθους 10). Τα μορφολογικά φιλτραρίσματα γίνονται με χρήση της συνάρτησης **morphologyEx** της cv2 ενώ για την δημιουργία του δομικού στοιχείου χρησιμοποιούμε την έτοιμη συνάρτηση **disk_strel** που δινόταν στο συμπληρωματικό υλικό της πρώτης εργαστηριακής άσκησης.

Τελικό βήμα αποτελεί η δημιουργία τριών ορθογώνιων περιοχών (bounding boxes) τα οποία περιβάλλουν τις περιοχές ενδιαφέροντος-δέρματος και χρησιμοποιούνται στο επόμενο μέρος για τον υπολογισμό του πεδίου Οπτικής Ροής και την τελική παρακολούθηση προσώπου και χεριών. Για τον σκοπό αυτό χρησιμοποιούμε αρχικά την έτοιμη συνάρτηση **label** της scipy.ndimage η οποία πραγματοποιεί την κατάτμηση του πλαισίου σε τρία διαφορετικά μέρη ενώ αποδίδει και διαφορετικά labels στο καθένα από αυτά. Τα labels που αποδίδει για τις περιοχές του προσώπου, του αριστερού και του δεξιού χεριού είναι αντίστοιχα 1,2 και 3. Στη συνέχεια βρίσκουμε για κάθε μία από τις τρεις διαφορετικές περιοχές τα όριά της και προσαρμόζουμε σε αυτά το αντίστοιχο bounding box. Σημειώνουμε πως στην περίπτωση των δύο χεριών (αριστερού και δεξιού) αφού ορίζουμε τα όρια του bounding box, το επεκτείνουμε κατά 6 pixel ανά κατεύθυνση για το αριστερό χέρι και κατά 4 για το δεξί προκειμένου να του δώσουμε ένα περιθώριο κίνησης. Για την αναπαράσταση των bounding boxes πάνω στα διαδοχικά πλαίσια της ακολουθίας εικόνων του βίντεο χρησιμοποιούμε τη συνάρτηση **rectangle** της cv2. Ενδεικτικά παρουσιάζουμε παρακάτω την τοποθέτηση των τριών bounding boxes γύρω από τις περιοχές του προσώπου και των χεριών για το πρώτο πλαίσιο της ακολουθίας ενώ απεικονίζουμε και την αντίστοιχη γκαουσιανή κατανομή πιθανότητας από δύο διαφορετικές όψεις:



Η υλοποίηση όλων των παραπάνω πραγματοποιείται εντός της συνάρτησης **boundingBox** η οποία δέχεται για ορίσματα το πρώτο πλαίσιο, τη μέση τιμή μ και την συνδιακύμανση Σ της Γκαουσιανής κατανομής. Αυτή επιστρέφει τρεις λίστες όπου η κάθε μία αφορά το πλαίσιο οριοθέτησης μιας εκ των τριών περιοχών ενδιαφέροντος και είναι της μορφής $[x, y, \text{width}, \text{height}]$, όπου x, y οι συντεταγμένες του πάνω αριστερά σημείου, width το πλάτος και height το ύψος του bounding box.

1.2 Παρακολούθηση Προσώπου και Χεριών

Έχοντας πλέον εντοπίσει τις περιοχές ενδιαφέροντος, δηλαδή τις περιοχές του προσώπου και των χεριών, υλοποιούμε τον αλγόριθμο των Lucas-Kanade για την εξαγωγή του διανυσματικού πεδίου οπτικής ροής μεταξύ διαδοχικών frames. Έτσι, επιτυγχάνουμε την παρακολούθηση των περιοχών αυτών καθώς το βίντεο εξελίσσεται.

1.2.1 Υλοποίηση του Αλγόριθμου των Lucas-Kanade

Θεωρητικό υπόβαθρο αλγορίθμου

Σε μια ακολουθία εικόνων N frames $I_n(x)$, όπου $n = 1, 2, \dots, N$ και $x = (x, y)$, το πεδίο οπτικής ροής $-d$, όπου $d(x) = (d_x, d_y)$, φέρνει σε αντιστοιχία δύο διαδοχικές εικόνες, έτσι ώστε:

$$I_n(x) \approx I_{n-1}(x + d) \quad (2)$$

Ο υπολογισμός της οπτικής ροής σε κάθε σημείο της εικόνας x με τον αλγόριθμο των Lucas-Kanade γίνεται με βάση τη μέθοδο των ελάχιστων τετραγώνων. Θεωρούμε ότι το d είναι σταθερό σε ένα μικρό παράθυρο γύρω από το σημείο και ελαχιστοποιούμε το τετραγωνικό σφάλμα

$$J_x(d) = \int_{x' \in R^2} G_p(x - x') [I_n(x') - I_{n-1}(x' + d)]^2 dx' \quad (3)$$

όπου G_p είναι μια συνάρτηση παραθύρωσης (εμείς χρησιμοποιούμε μια Γκαουσιανή με τυπική απόκλιση ρ)

Θεωρούμε ότι έχουμε μια εκτίμηση d_i για το d και προσπαθούμε να τη βελτιώσουμε κατά u , δηλαδή $d_{i+1} = d_i + u$. Αναπτύσσοντας κατά Taylor την έκφραση

$I_{n-1}(x + d) = I_{n-1}(x + d_i + u)$ γύρω από το σημείο $x + d_i$, προκύπτει ότι:

$$I_{n-1}(x + d) \approx I_{n-1}(x + d_i) + \nabla I_{n-1}(x + d_i)^T u \quad (4)$$

Αντικαθιστώντας τώρα την παραπάνω έκφραση στην Εξίσωση (3) αποδεικνύεται ότι η λύση ελάχιστων τετραγώνων για τη βελτίωση του πεδίου οπτικής ροής σε κάθε σημείο είναι

$$u(x) = \begin{bmatrix} (G_\rho * A_1^2)(x) + \epsilon & (G_\rho * (A_1 A_2))(x) + \epsilon \\ (G_\rho * (A_1 A_2))(x) + \epsilon & (G_\rho * A_2^2)(x) + \epsilon \end{bmatrix}^{-1} \cdot \begin{bmatrix} (G_\rho * A_1 E)(x) \\ (G_\rho * A_2 E)(x) \end{bmatrix} \quad (5)$$

όπου

$$A(x) = \begin{bmatrix} A_1(x) & A_2(x) \end{bmatrix} = \begin{bmatrix} \frac{\partial I_{n-1}(x+d_i)}{\partial x} & \frac{\partial I_{n-1}(x+d_i)}{\partial y} \end{bmatrix} \quad (6)$$

$$E(x) = I_n(x) - I_{n-1}(x + d_i) \quad (7)$$

και το ‘*’ δηλώνει συνέλιξη. Η μικρή θετική σταθερά ϵ βελτιώνει το αποτέλεσμα σε επίπεδες περιοχές με μειωμένη υφή και άρα μειωμένη πληροφορία για τον υπολογισμό του πεδίου ροής. Η ανανέωση του πεδίου οπτικής ροής $d_{i+1} = d_i + u$, με το u να υπολογίζεται από την Εξίσωση (5), επαναλαμβάνεται αρκετές φορές ως τη σύγκλιση. Ένας τυπικός αριθμός επαναλήψεων είναι 5-10. Επιλέγουμε να εκτελούμε 10 επαναλήψεις προκειμένου να συγκλίνουμε σε όσο το δυνατόν καλύτερη εκτίμηση του πεδίου.

Υλοποίηση αλγορίθμου

Με βάση το παραπάνω θεωρητικό υπόβαθρο υλοποιούμε τώρα τον αλγόριθμο **Lucas-Kanade** ως μια αυτόνομη συνάρτηση, που δέχεται ως εισόδους δύο εικόνες (κομμένα παράθυρα από δυο διαδοχικά πλαίσια του βίντεο με βάση το bounding box), το εύρος ρ του γκαουσιανού παραθύρου, την θετική σταθερά κανονικοποίησης ϵ , και την αρχική εκτίμηση \mathbf{d}_0 για το πεδίο οπτικής ροής ενώ επιστρέφει το τελικό διανυσματικό πεδίο οπτικής ροής \mathbf{d} .

Για την υλοποίηση της 2Δ Gaussian $n \times n$ εκμεταλλευόμαστε την ιδιότητα της διαχωρισιμότητας, σύμφωνα με την οποία αυτή μπορεί να προκύψει άμεσα ως γινόμενο δύο μονοδιάστατων Gaussian μεγέθους n , τις οποίες και ορίζουμε με χρήση της συνάρτησης **getGaussianKernel** της βιβλιοθήκης `cv2`. Για τον ορισμό κάθε μονοδιάστατης Gaussian με τυπική απόκλιση ρ θεωρούμε πυρήνα μεγέθους ίσο με $n \times n$ με $n = \text{ceil}(3 \cdot \rho) \cdot 2 + 1$.

Για τον υπολογισμό της $I_{n-1}(x + d_i)$ και των $\frac{\partial I_{n-1}(x+d_i)}{\partial x}$, $\frac{\partial I_{n-1}(x+d_i)}{\partial y}$ χρειάζεται να πάρουμε τιμές της I_{n-1} και των μερικών παραγώγων της σε ενδιάμεσα σημεία του πλέγματος.

Αρχικά υπολογίζουμε τις μερικές παραγώγους της cropped εικόνας, όπως αυτή καθορίζεται από τα όρια του bounding box, μέσω της συνάρτησης **gradient** της βιβλιοθήκης `numpy`.

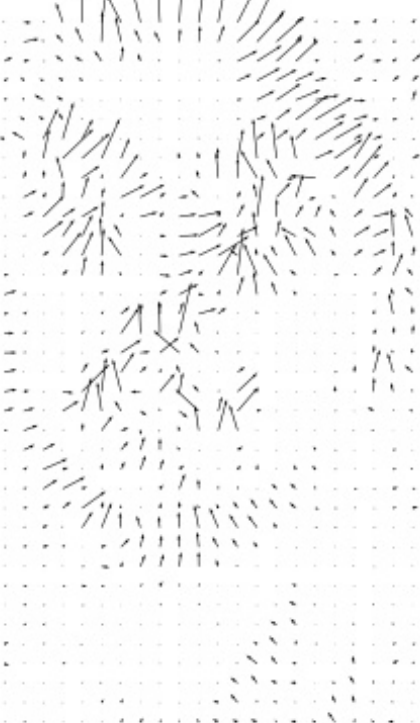
```
[Iprev_y, Iprev_x] = np.gradient(Iprev)
```

Στη συνέχεια υπολογίζουμε τις τιμές των μερικών παραγώγων στα σημεία $x + d_i$ μέσω της συνάρτησης **map_coordinates** της βιβλιοθήκης `numpy`. Πρωτίστως έχουμε ορίσει τα x_0, y_0 μέσω της συνάρτησης **np.meshgrid**, καλώντας την συνάρτηση με τρόπο τέτοιο ώστε να καλύπτονται όλες οι συντεταγμένες του Bounding Box (αυτό επιτυγχάνεται μέσω του `range`). Τέλος, να σημειωθεί πως για την κλήση της συνάρτησης **map_coordinates** χρειάζεται και η συνάρτηση **ravel** της βιβλιοθήκης `numpy` έτσι ώστε οι 2Δ πίνακες $x_0 + d_i$ και $y_0 + d_i$ να μετατραπούν σε διανύσματα ενώ απαραίτητο είναι και ένα **reshape** του τελικού αποτελέσματος που επιστρέφει η **map_coordinates** έτσι ώστε αυτό να λάβει τις διαστάσεις του bounding box.

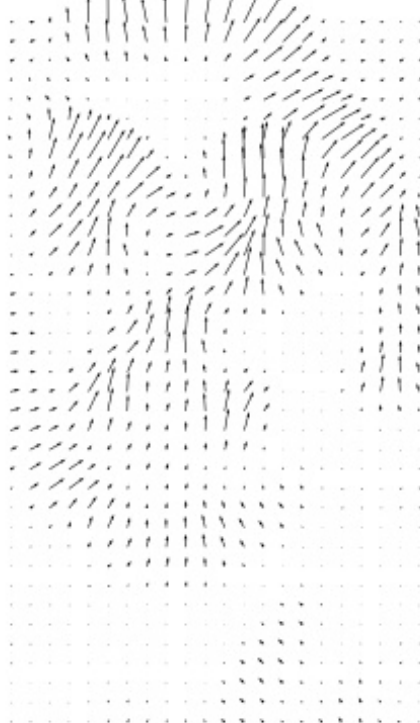
```
x_0, y_0 = np.meshgrid(range(Iprev.shape[1]), range(Iprev.shape[0]))  
I_interpolated = map_coordinates(Iprev, [np.ravel(y_0+d_y), np.ravel(x_0+d_x)],  
                                order=1).reshape((Iprev.shape[0], Iprev.shape[1]))
```

Για τον ορισμό του πίνακα $u(x)$ απαιτούνται συνελίξεις με την δισδιάστατη γκαουσιανή που έχουμε ορίσει. Οι συνελίξεις αυτές πραγματοποιούνται με χρήση της συνανάρτησης **filter2D** της βιβλιοθήκης `cv2`. Πειραματιζόμαστε τώρα με τις τιμές των παραμέτρων ρ και ϵ . Τα αποτελέσματα που προκύπτουν για το διανυσματικό πεδίο του προσώπου φαίνονται παρακάτω:

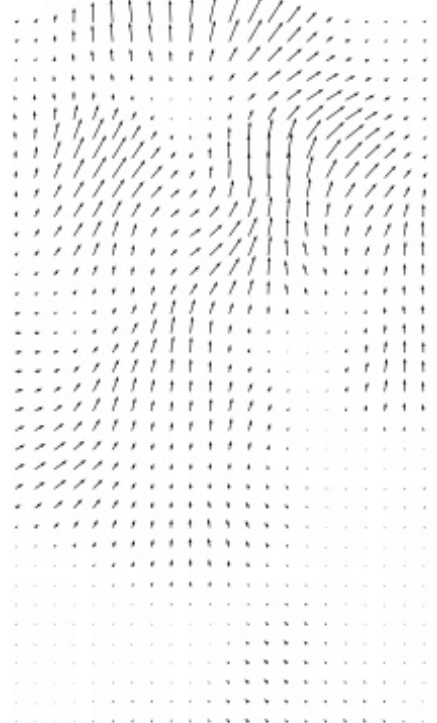
Optical flow for $\rho = 1$ and $\epsilon = 0.01$



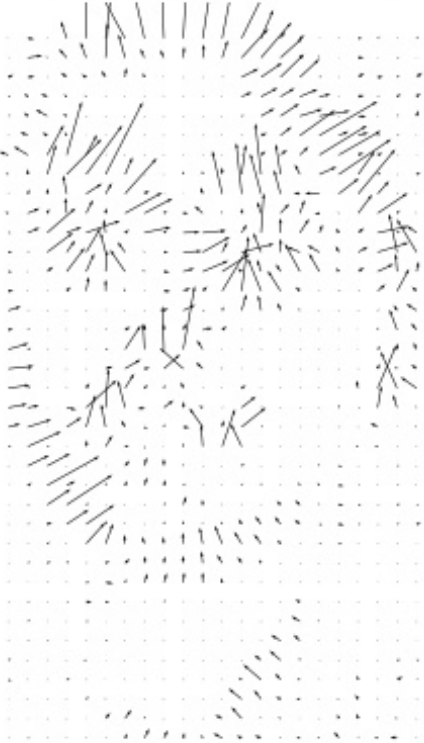
Optical flow for $\rho = 3$ and $\epsilon = 0.01$



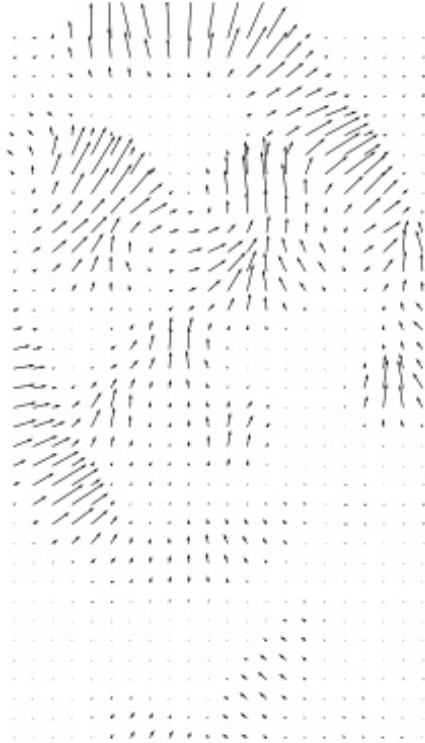
Optical flow for $\rho = 5$ and $\epsilon = 0.01$



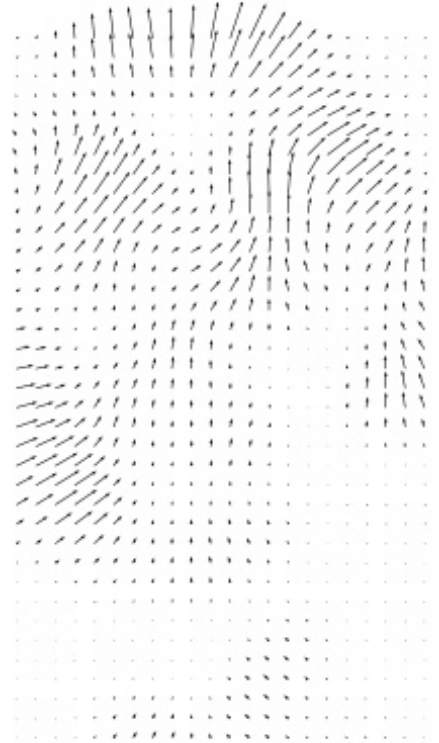
Optical flow for $\rho = 1$ and $\epsilon = 0.05$



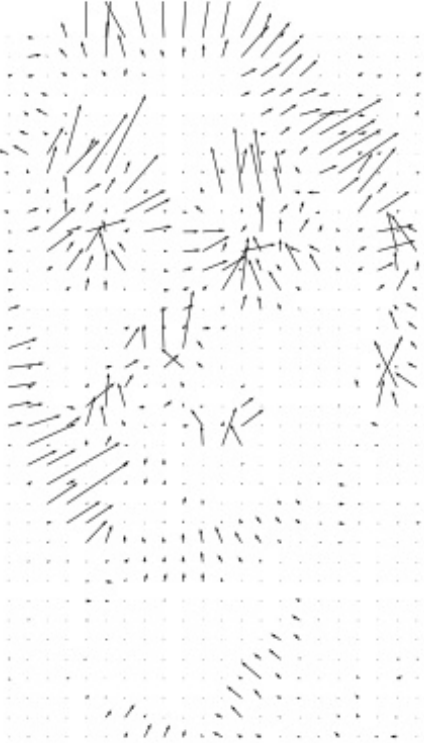
Optical flow for $\rho = 3$ and $\epsilon = 0.05$



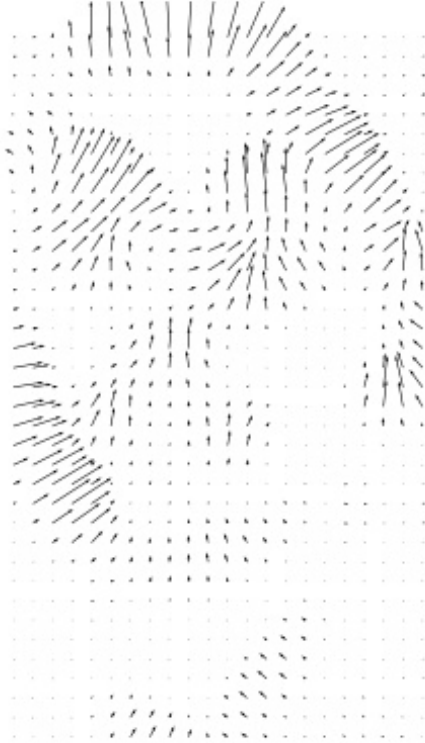
Optical flow for $\rho = 5$ and $\epsilon = 0.05$



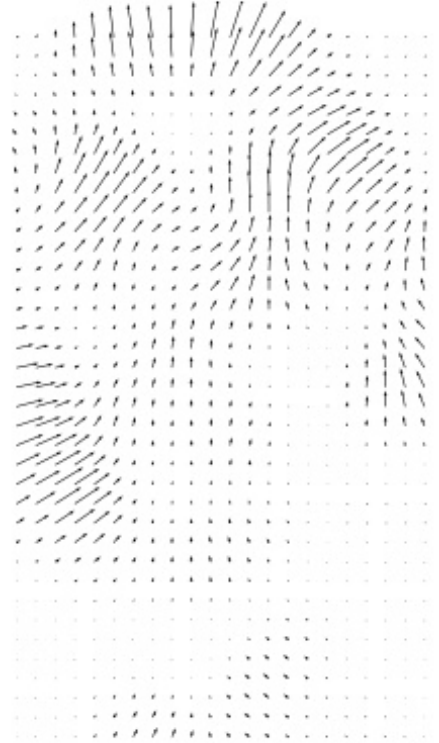
Optical flow for $\rho = 1$ and $\epsilon = 0.1$



Optical flow for $\rho = 3$ and $\epsilon = 0.1$



Optical flow for $\rho = 5$ and $\epsilon = 0.1$



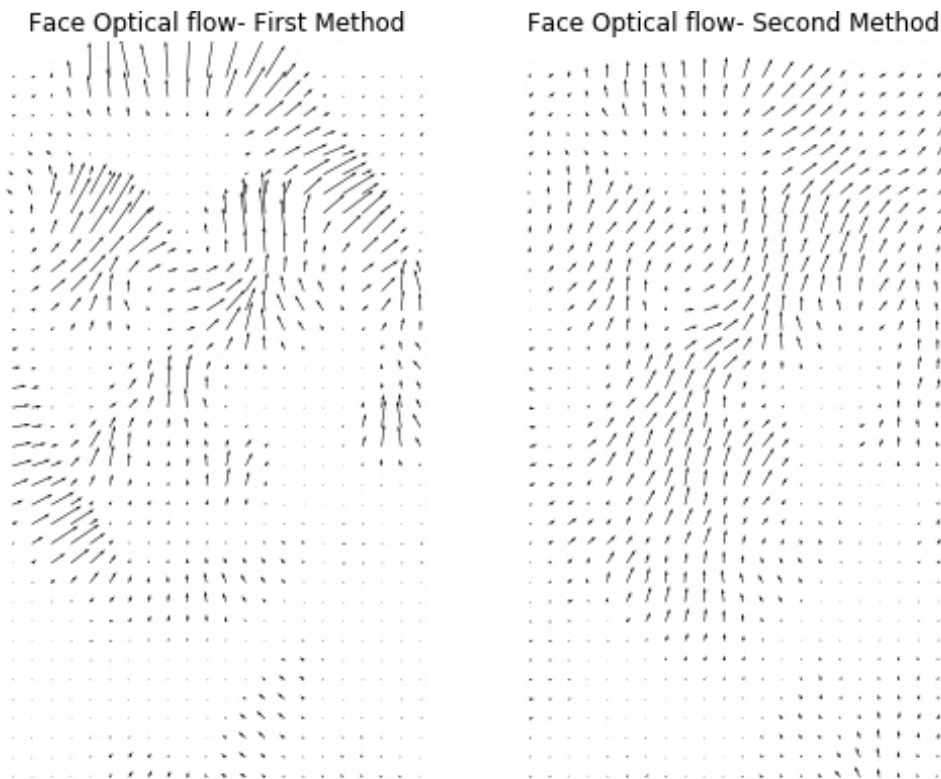
Παρατηρούμε ότι η μικρή θετική σταθερά ϵ βελτιώνει το αποτέλεσμα σε επίπεδες περιοχές με μειωμένη υφή και άρα μειωμένη πληροφορία για τον υπολογισμό του πεδίου ροής. Σχετικά με την παράμετρο ρ ισχύει ότι για σταθερό ϵ καθώς αυτή αυξάνεται το διανυσματικό πεδίο γίνεται

πιο ομαλό. Το αποτέλεσμα αυτό είναι αναμενόμενο καθώς η σταθερά ρ αντιστοιχεί στην τυπική απόκλιση της γκαουσιανής με την οποία συνελίσσουμε και επομένως όσο μεγαλύτερη είναι η τιμή της τόσο μεγαλύτερη ομαλοποίηση επιτυγχάνουμε. Τελικά, επιλέγουμε τις τιμές των παραμέτρων ρ και ϵ ίσες με 3 και 0.07 αντίστοιχα.

Μέχρι στιγμής θεωρούμε ότι το διάνυσμα \mathbf{d} συγκλίνει μετά από 5 με 10 επαναλήψεις. Πειραματιζόμαστε τώρα με διαφορετικά κριτήρια σύγκλισης. Συγκεκριμένα, θεωρούμε ότι το διάνυσμα \mathbf{d} συγκλίνει όταν η διαφορά του ανανεωμένου πεδίου οπτικής ροής, \mathbf{d}_{i+1} , από την προηγούμενη εκτίμησή του, \mathbf{d}_i , δεν είναι μεγαλύτερη από 10^{-k} . Μαθηματικά αυτό διατυπώνεται ως εξής:

$$\|\mathbf{d}_{i+1} - \mathbf{d}_i\| < 10^{-k} \Leftrightarrow \|\mathbf{u}\| < 10^{-k}$$

Τροποποιούμε τον κώδικα ώστε να ελέγχει σε κάθε επανάληψη αν ισχύει η παραπάνω συνθήκη. Ως νόρμα του πίνακα \mathbf{u} θεωρούμε την $\|\cdot\|_2$ την οποία υπολογίζουμε με χρήση της συνάρτησης **norm** της numpy.linalg. Για τιμές παραμέτρων ρ και ϵ ίσες με 3 και 0.07 αντίστοιχα τα διανυσματικά πεδία για το πρόσωπο με τις δύο μεθόδους (σύγκλιση μετά από 10 επαναλήψεις και σύγκλιση με βάση το μαθηματικό κριτήριο και $k=1$) φαίνονται παρακάτω:



1.2.2 Υπολογισμός της Μετατόπισης του Προσώπου και των Χεριών από το Πεδίο Οπτικής Ροής

Έχοντας υπολογίσει την οπτική ροή της εικόνας I_n στην περιοχή που έχει ορίσει το bounding box της I_{n-1} , απομένει να βρούμε το συνολικό διάνυσμα μετατόπισης του bounding box ορθογωνίου, με όσο το δυνατόν μεγαλύτερη ακρίβεια. Ισχύει ότι τα διανύσματα του πεδίου της οπτικής ροής έχουν μεγαλύτερο μήκος σε σημεία που ανήκουν σε περιοχές έντονης πληροφορίας όπως οι ακμές και οι κορυφές και σχεδόν μηδενικό μήκος σε σημεία ομοιόμορφης και επίπεδης υφής. Για τον λόγο αυτό ο υπολογισμός της μέσης τιμής όλων των διανυσμάτων μετατόπισης θα οδηγούσε σε ανακριβή αποτελέσματα. Για να πετύχουμε μεγαλύτερη ακρίβεια ορίζουμε αρχικά την ενέργεια της οπτικής ροής σε κάθε σημείο της περιοχής ενδιαφέροντος ως το άθροισμα των τετραγώνων των μετατοπίσεων d_x και d_y στα αντίστοιχα pixels:

$$Energy = \|d\|^2 = d_x^2 + d_y^2$$

Στην συνέχεια υπολογίζουμε την μέση τιμή των διανυσμάτων μετατόπισης εκείνων που έχουν ενέργεια μεγαλύτερη από ένα κατώφλι. Τελικά μετά από δοκιμές επιλέγουμε ως τιμή κατωφλίου τα 3/10 της μέγιστης τιμής της ενέργειας.

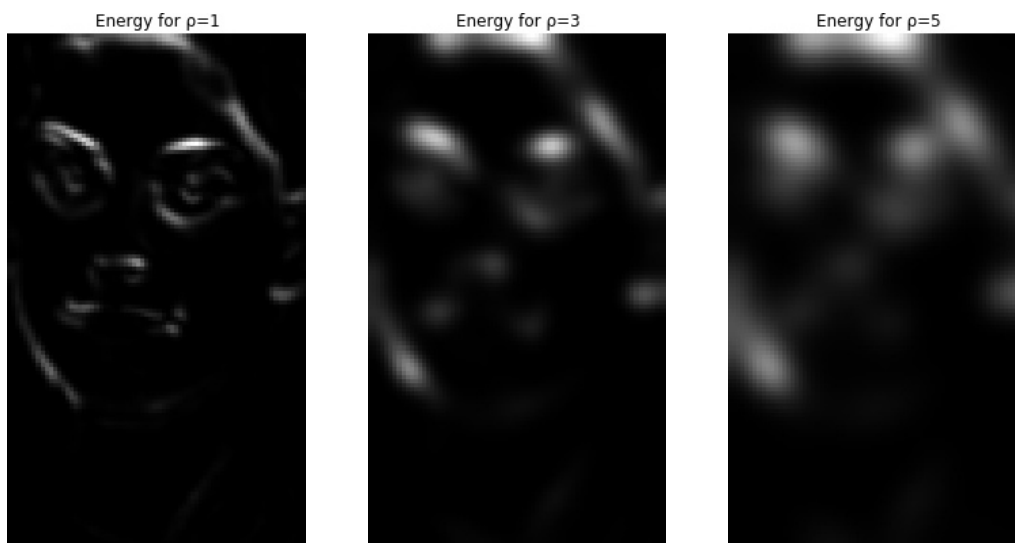
Όλα τα παραπάνω υλοποιούνται εντός της συνάρτησης **displ** η οποία δέχεται για ορίσματα τις συνιστώσες d_x και d_y του διανυσματικού πεδίου οπτικής ροής και επιστρέφει το τελικό διάνυσμα μετατόπισης του ορθογωνίου.

Συνοψίζοντας η όλη διαδικασία παρακολούθησης των περιοχών ενδιαφέροντος μεταξύ διαδοχικών εικόνων (καρέ) του βίντεο αποτελείται από τα ακόλουθα βήματα:

1. Ορίζουμε μέσω της συνάρτησης **boundingBox** τα bounding boxes για τις τρεις περιοχές ενδιαφέροντος (πρόσωπο, αριστερό χέρι, δεξί χέρι).
2. Για κάθε ζεύγος διαδοχικών εικόνων του βίντεο, I_{n-1} και I_n , κόβουμε τα αντίστοιχα πλαίσια στις τρεις περιοχές ενδιαφέροντος σύμφωνα με τα bounding boxes.
3. Βρίσκουμε το πεδίο οπτικής ροής με τον αλγόριθμο των Lucas-Kanade καλώντας την συνάρτηση **lk** τρεις φορές, μία για κάθε περιοχή ενδιαφέροντος.
4. Υπολογίζουμε το συνολικό διάνυσμα μετατόπισης για κάθε bounding box με βάση τη συνάρτηση **displ**.
5. Μετατοπίζουμε τα bounding boxes στη νέα τους θέση και επαναλαμβάνουμε την διαδικασία μέχρι το τέλος του βίντεο.

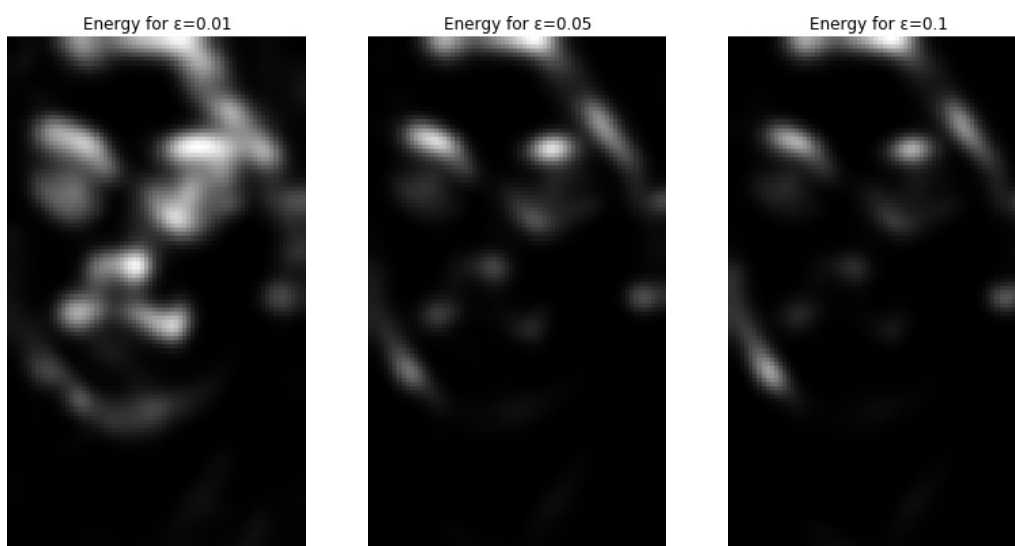
Πειραματιζόμαστε τώρα με διαφορετικές τιμές των παραμέτρων ρ , ϵ και κατωφλίου ενέργειας (*threshold*) προκειμένου να διαπιστώσουμε πώς αυτές επηρεάζουν την ενέργεια των διανυσμάτων μετατόπισης.

Για να διαπιστώσουμε πώς αλλάζει η ενέργεια σε σχέση με το ρ διατηρούμε σταθερές τις τιμές των ϵ , *threshold* και ίσες με 0.07 και 0.3 αντίστοιχα. Τα αποτελέσματα που προκύπτουν για $\rho = 1, 3, 5$ φαίνονται παρακάτω:



Παρατηρούμε πως καθώς η τιμή της παραμέτρου ρ αυξάνεται η γραφική παράσταση της ενέργειας γίνεται πιο θολή. Το γεγονός αυτό είναι αναμενόμενο καθώς η παράμετρος ρ αντιστοιχεί στην τυπική απόκλιση της γκαουσιανής με την οποία συνελίσσουμε και άρα όσο μεγαλύτερη η τιμή της τόσο πιο ομαλοποιημένη η εικόνα που προκύπτει.

Διατηρούμε τώρα σταθερές τις τιμές των παραμέτρων ρ και *threshold* και παρατηρούμε πως μεταβάλλεται η ενέργεια σε σχέση με το ϵ . Τα αποτελέσματα που προκύπτουν φαίνονται παρακάτω:



Παρατηρούμε πως για αυξανόμενες τιμές της παραμέτρου ϵ η ενέργεια γίνεται πιο 'σκοτεινή'. Το αποτέλεσμα αυτό είναι αναμενόμενο γιατί με την αύξηση του ϵ αυτό που επιθυμούμε είναι να έχουμε μειωμένη ενέργεια στα σημεία που αντιστοιχούν σε περιοχές με μειωμένη υφή καθώς τέτοιες περιοχές δεν είναι κατάλληλες για τον υπολογισμό της οπτικής ροής.

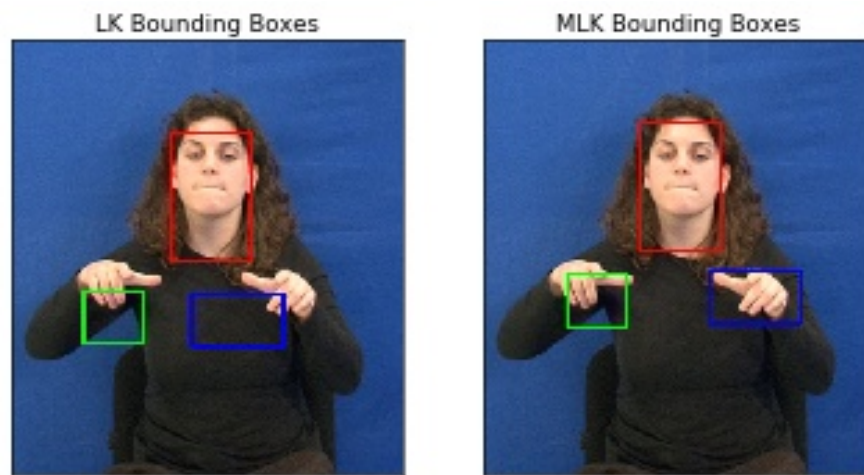
Σημειώνουμε πως το κριτήριο υπολογισμού της μετατόπισης από το πεδίο οπτικής ροής προέκυψε ύστερα από πειραματισμούς. Ως εναλλακτικό τρόπο, υπολογίσαμε την μέση τιμή της ενέργειας και δοκιμάσαμε να λάβουμε υπόψην στον υπολογισμό της μέσης τιμής των διανυσμάτων μετατόπισης μόνο εκείνα τα σημεία των οποίων η ενέργεια υπερβαίνει τη μέση της τιμή κατά ένα παράγοντα 1.5.

1.2.3 Πολυ-Κλιμακωτός Υπολογισμός Οπτικής Ροής

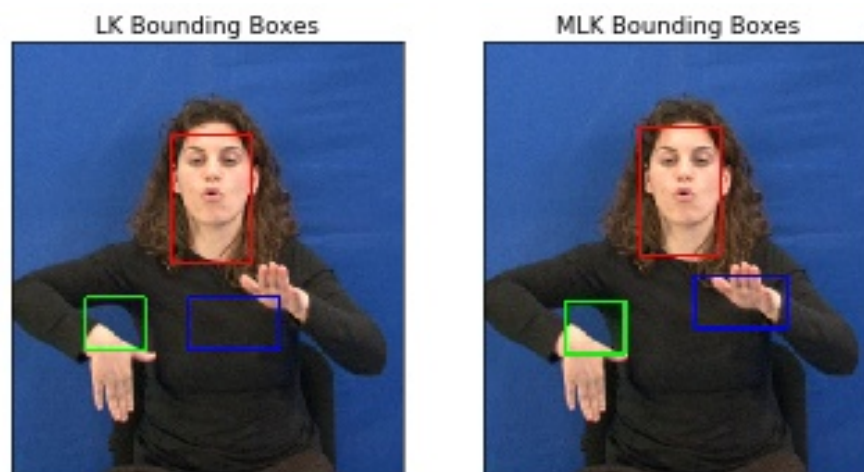
Υλοποιούμε τώρα την πολυ-κλιμακωτή εκδοχή του αλγόριθμου των Lucas-Kanade. Ο αλγόριθμος αναλύει αρχικά τις εικόνες σε γκαουσιανές πυραμίδες και υπολογίζει το πεδίο οπτικής ροής από τις πιο μικρές (τραχείς) στις πιο μεγάλες (λεπτομερείς) κλίμακες, χρησιμοποιώντας τη λύση της μικρής κλίμακας ως αρχική συνθήκη για τη μεγάλη. Για την υλοποίηση του πολυκλιμακωτού αλγορίθμου των Lucas-Kanade ορίζουμε την συνάρτηση **MultiScaleI**k η οποία δέχεται τα ίδια ορίσματα με την αντίστοιχη **I**k, με την διαφορά ότι δέχεται επιπλέον και τον αριθμό των κλιμάκων της πυραμίδας. Για τη μετάβαση από μεγάλες σε μικρές κλίμακες χρειάζεται να φιλτράρουμε την εικόνα με ένα βαθυπερατό φίλτρο πριν την υποδειγματοληψία προκειμένου να μετριάσουμε φαινόμενα φασματικής αναδίπλωσης (aliasing) της εικόνας. Ως βαθυπερατό φίλτρο ορίζουμε μια δισδιάστατη γκαουσιανή τυπικής απόκλισης $\sigma = 3$. Για τον πολυκλιμακωτό υπολογισμό της οπτικής ροής βρίσκουμε αρχικά και αποθηκεύουμε σε δύο λεξικά, `pyramid_prev` και `pyramid_next`, τις εικόνες ανά τα διάφορα επίπεδα της γκαουσιανής πυραμίδας. Με εξαίρεση το πρώτο επίπεδο, στο οποίο η ακολουθία των εικόνων ταυτίζεται με την αρχική, τα επόμενα προκύπτουν από φιλτράρισμα των εικόνων του προηγούμενου επιπέδου με την γκαουσιανή και ύστερα υποδειγματοληψία κάθε φορά στο μισό. Για την εκτέλεση του αλγορίθμου των Lucas-Kanade μονής κλίμακας, ο οποίος χρησιμοποιείται ως υπο-ρουτίνα από την πολυκλιμακωτή εκδοχή του, ξεκινάμε να διατρέχουμε την πυραμίδα ανάποδα. Συγκεκριμένα, ξεκινάμε με μηδενικές εκτιμήσεις κίνησης από το τελευταίο επίπεδο και υπολογίζουμε τις συνιστώσες d_x και d_y του διανυσματικού πεδίου **d**. Στη συνέχεια επαναλαμβάνουμε την διαδικασία και για τα υπόλοιπα επίπεδα της πυραμίδας με την διαφορά ότι οι εκτιμήσεις κίνησης δεν αρχικοποιούνται στο μηδέν αλλά στο αποτέλεσμα της προηγούμενης κλήσης του **I**k αφού πρώτα αυτό διπλασιαστεί.

Έχοντας τώρα υλοποιήσει την πολυκλιμακωτή εκδοχή του αλγορίθμου των Lucas-Kanade συγκρίνουμε τα αποτελέσματά του στην ίδια ακολουθία εικόνων σε σχέση με τον αλγόριθμο μονής κλίμακας. Για λόγους διευκόλυνσης και σύγκρισης δημιουργούμε subplots στα οποία φαίνεται ξεκάθαρα η σύγκριση των δύο μεθόδων ως προς την τελική παρακολούθηση εικόνων και χειρών (**bounding_boxes_subplots**) και την ενέργεια των διανυσμάτων (**energy_subplots**) για όλα τα πλαίσια της ακολουθίας εικόνων του βίντεο. Επίσης για λόγους συμπίεσης τα αντίστοιχα plots για το πεδίο οπτικής ροής επισυνάπτονται στα παραδοτέα ως GIFs.

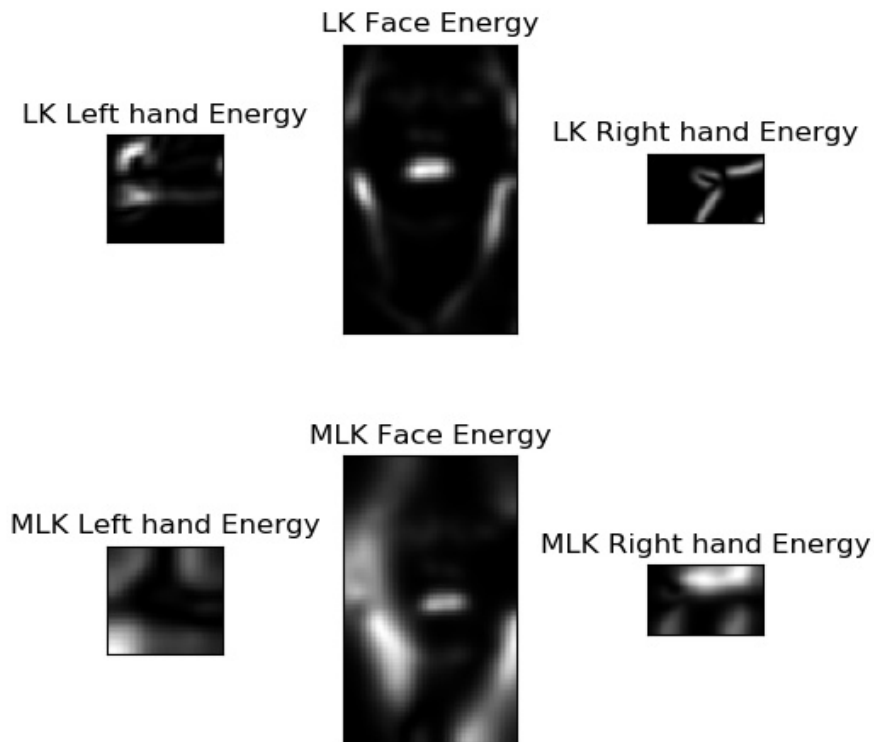
Αναφορικά με την παρακολούθηση των περιοχών ενδιαφέροντος παρατηρούμε πως ο πολυκλιμακωτός αλγόριθμος πετυχαίνει καλύτερα αποτελέσματα σε σχέση με τον μονοκλιμακωτό. Αυτό είναι αναμενόμενο καθώς ο αλγόριθμος μονής κλίμακας μπορεί να δώσει ικανοποιητικά αποτελέσματα μόνο σε περιπτώσεις που έχουμε μικρές και αργές κινήσεις, όπως πχ. οι κινήσεις του κεφαλιού. Για να μπορέσουμε να παρακολουθήσουμε με επιτυχία γρήγορες αλλαγές, όπως είναι οι κινήσεις των χεριών, χρειάζεται η εφαρμογή του πολυκλιμακωτού αλγορίθμου. Παρακάτω φαίνεται η σύγκριση των δύο αλγορίθμων για το πλαίσιο 51 της αρχικής ακολουθίας εικόνων η οποία επιβεβαιώνει τα όσα αναφέραμε. Συγκεκριμένα, τόσο ο LK όσο και ο MultiScale-LK παρακολουθούν επιτυχώς τις κινήσεις του κεφαλιού καθώς αυτές είναι μικρές και αργές. Αντίθετα για τις μεγάλες κινήσεις των χεριών ο LK χάνει το tracking ήδη από τα πρώτα frames.



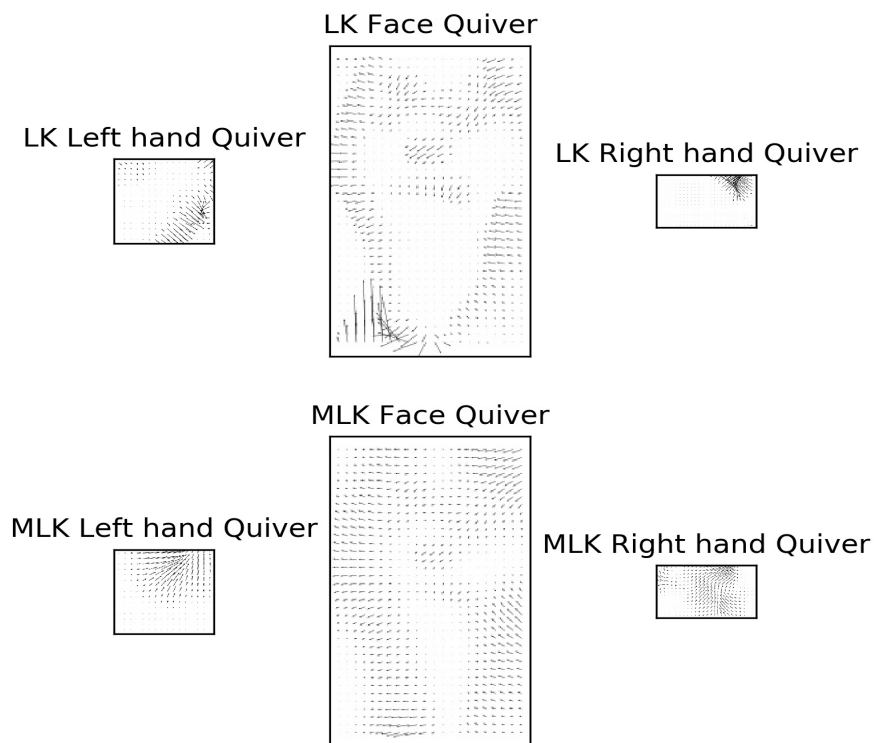
Ωστόσο ενδέχεται και ο πολυκλιμακωτός αλγόριθμος να χάσει την παρακολούθηση των γρήγορων κινήσεων. Μάλιστα για όλο και μεγαλύτερο αριθμό κλιμάκων και επειδή στα υψηλότερα επίπεδα της πυραμίδας εντοπίζεται μεγαλύτερη κίνηση είναι πιθανό το bounding box να μετατοπιστεί σε τέτοιο βαθμό ώστε να βγει εκτός των ορίων της περιοχής που περικλείει. Ένα τέτοιο παράδειγμα φαίνεται ακολουθώντας όπου εντοπίζουμε αστοχία στην παρακολούθηση του αριστερού χεριού και από τον MultiScale-LK πέραν του LK.



Συγκρίνουμε τώρα την ενέργεια των τριών περιοχών ενδιαφέροντος και για τις δύο μεθόδους, μονοκλιμακωτή και πολυκλιμακωτή. Ενδεικτικά παραθέτουμε την ενέργεια για το πλαίσιο 57 της αρχικής ακολουθίας εικόνων.



Παρατηρούμε ότι στην περίπτωση του MultiScale-LK η ενέργεια είναι διασκορπισμένη σε μία μεγαλύτερη περιοχή. Τέλος συγκρίνουμε τις δύο μεθόδους ως προς το πεδίο οπτικής ροής.



Το παραπάνω subplot αφορά το διανυσματικό πεδίο για το πλαίσιο 44 της ακολουθίας εικόνων ενώ παρατηρούμε ότι στην περίπτωση του πολυκλιμακωτού αλγορίθμου είναι πιο ομοιογενές σε σχέση με αυτό του μονοκλιμακωτού. Για το πλαίσιο αυτό όπως φαίνεται και από το αντίστοιχο subplot παρακολούθησης προσώπου και χεριών (βλ. φάκελο **bounding_boxes_subplots**) ο LK έχει χάσει το tracking και των δύο χεριών. Το γεγονός αυτό αιτιολογεί την διαφορά μεταξύ των δύο αλγορίθμων στα διανυσματικά πεδία του αριστερού και δεξιού χεριού.

Συγκρίνοντας τους δύο αλγορίθμους ως προς την ταχύτητα σύγκλισης διαπιστώνουμε πως στην περίπτωση του MultiScale-LK χρειάζονται λιγότερες επαναλήψεις. Το αποτέλεσμα αυτό είναι αναμενόμενο καθώς η λύση της προηγούμενης κλίμακας χρησιμοποιείται ως μία αρχική και πιο αντιπροσωπευτική συνθήκη για την μεγάλη κλίμακα.

Αναφορές

- [1] Μαραγκός Π. *Ανάλυση Εικόνων και Όραση Υπολογιστών*. Ε.Μ.Π., 2014
- [2] S.Baker and I.Matthews. Lucas-Kanade 20 years on: A unifying framework. *Int. J. of Comp. Vis.*, 56(3):221–255, 2004.