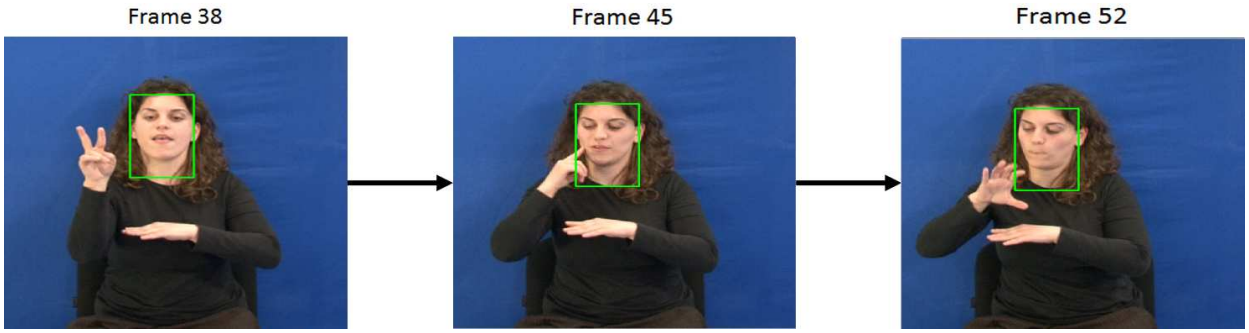
Ε.Μ.Π. – Σ.Η.Μ.Μ.Υ. **ΟΡΑΣΗ ΥΠΟΛΟΓΙΣΤΩΝ** Ακαδ.Ετος 2018–**2019**
Ημερ/νια: 16-05-2019 **2η Εργαστηριακή Ασκηση** Παράδοση: 30-05-2019

[Εξηγείστε περιεκτικά και επαρκώς την εργασία σας. Επιτρέπεται προαιρετικά η συνεργασία εντός ομάδων των 2 ατόμων. Κάθε ομάδα 2 ατόμων υποβάλλει μια κοινή αναφορά που αντιπροσωπεύει μόνο την προσωπική εργασία των μελών της. Αν χρησιμοποιήσετε κάποια άλλη πηγή εκτός του βιβλίου και του εκπαιδευτικού υλικού του μαθήματος, πρέπει να το αναφέρετε. Η παράδοση της αναφοράς και του κώδικα της εργασίας θα γίνει ηλεκτρονικά στο mycourses.ntua.gr και επιπλέον η αναφορά της εργασίας θα παραδίδεται τυπωμένη και προσωπικά στην γραμματεία του εργαστηρίου Ρομποτικής (2.1.12, παλαιό Κτ.Ηλεκ.), ώρες 09.30-14.30].

Θέμα: Εκτίμηση Οπτικής Ροής (Optical Flow) και Εξαγωγή Χαρακτηριστικών σε Βίντεο

Μέρος 1: Παρακολούθηση Προσώπου με Χρήση του Πεδίου Οπτικής Ροής (Optical Flow) με τη Μέθοδο των Lucas-Kanade

Σκοπός της εργαστηριακής άσκησης είναι η υλοποίηση ενός συστήματος Παρακολούθησης Προσώπου (Face Tracking) σε μια ακολουθία βίντεο νοηματικής γλώσσας. Το σύστημα αρχικά θα ανιχνεύει στο πρώτο πλαίσιο την περιοχή του προσώπου με χρήση ενός πιθανοτικού ανιχνευτή ανθρώπινου δέρματος. Στη συνέχεια θα μπορεί να παρακολουθεί αυτή την περιοχή του προσώπου χρησιμοποιώντας το διανυσματικό πεδίο οπτικής ροής, υπολογισμένο με τη μέθοδο Lucas-Kanade.



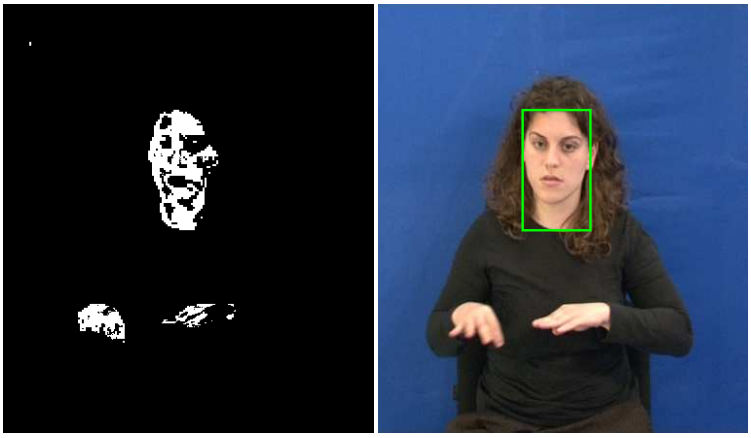
Σχήμα 1: Παράδειγμα παρακολούθησης προσώπου σε ακολουθία βίντεο νοηματικής γλώσσας.

1.1 Ανίχνευση Δέρματος Προσώπου

Στο πρώτο ερώτημα ζητείται η ανίχνευση σημείων δέρματος στο πρώτο πλαίσιο της ακολουθίας και η τελική επιλογή της περιοχής του προσώπου. 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 δισδιάστατη Γκαουσιανή κατανομή, δηλαδή

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

όπου $\mathbf{c}$ είναι το διάνυσμα τιμών Cb και Cr για κάθε σημείο (x, y) της εικόνας. Η Γκαουσιανή κατανομή εκπαιδεύεται υπολογίζοντας το 1×2 διάνυσμα μέσης τιμής $\boldsymbol{\mu} = [\mu_{C_b}\ \mu_{C_r}]$



(α) (β)

Σχήμα 2: 1ο πλαίσιο της ακολουθίας βίντεο νοηματικής γλώσσας: (α) Ανίχνευση σημείων δέρματος. (β) Τελική ανίχνευση περιοχής προσώπου.

και τον 2×2 πίνακα συνδιακύμανσης Σ από τα δείγματα δέρματος που δίνονται στο αρχείο `skinSamplesRGB.mat` σε μορφή RGB. Η δυαδική εικόνα ανίχνευσης δέρματος προκύπτει από την εικόνα πιθανοτήτων $P(\mathbf{c}(x, y) = \text{skin}),\ \forall\,(x, y)$ με κατωφλιοποίηση. Ενδεικτικές τιμές κατωφλίου: στο διάστημα $[0.1, 0.3]$ (για τιμές πιθανοτήτων $[0, 1]$).

Η τελική ανίχνευση της περιοχής δέρματος του προσώπου γίνεται επιλέγοντας την περιοχή με το μεγαλύτερο εμβαδό από όσες βρέθηκαν. Για το σκοπό αυτό απαιτείται μια μορφολογική επεξεργασία της δυαδικής εικόνας δέρματος και συγκεκριμένα κάλυψη των τρυπών, opening με πολύ μικρό δομικό στοιχείο και closing με μεγάλο δομικό στοιχείο, έτσι ώστε να εξαλειφθούν οι μικρές περιοχές και να αποκτήσουν συνοχή οι περιοχές του προσώπου και των χεριών. Το ορθογώνιο που περιβάλλει την τελική περιοχή δέρματος του προσώπου (bounding box) είναι το παράθυρο της εικόνας που θα χρησιμοποιηθεί στο Μέρος 2 για υπολογισμό του πεδίου Οπτικής Ροής και την τελική παρακολούθηση του προσώπου.

Υλοποιείστε την παραπάνω διαδικασία στο περιβάλον Matlab ως αυτόνομη συνάρτηση που να δέχεται ως εισόδους μια εικόνα (την πρώτη της ακολουθίας βίντεο), τη μέση τιμή $\boldsymbol{\mu}$ και την συνδιακύμανση Σ της Γκαουσιανής κατανομής και να επιστρέφει το πλαίσιο οριοθέτησης προσώπου στη μορφή `[x, y, width, height]`, όπου `x, y` οι συντεταγμένες του πάνω αριστερά σημείου, π.χ.

```
boundingBox = fd(I, mu, cov)
```

▶ Βοήθεια για Matlab: συναρτήσεις `rgb2ycbcr, surf, bwlabel, regionprops, rectangle, cov, mvnpdf`.

1.2 Υλοποίηση του Αλγόριθμου των Lucas-Kanade

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

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

Ο αλγόριθμος των Lucas-Kanade υπολογίζει την οπτική ροή σε κάθε σημείο της εικόνας $\mathbf{x}$ με τη μέθοδο των ελάχιστων τετραγώνων, θεωρώντας ότι το $\mathbf{d}$ είναι σταθερό σε ένα μικρό

παράθυρο γύρω από το σημείο και ελαχιστοποιώντας το τετραγωνικό σφάλμα

$$J_{\mathbf{x}}(\mathbf{d}) = \int_{\mathbf{x}' \in \mathbb{R}^2} G_\rho(\mathbf{x} - \mathbf{x}') \left[I_n(\mathbf{x}') - I_{n-1}(\mathbf{x}' + \mathbf{d}) \right]^2 \mathrm{d}\mathbf{x}', \tag{3}$$

όπου $G_\rho(\mathbf{x})$ είναι μια συνάρτηση παραθύρωσης, π.χ. Γκαουσιανή με τυπική απόκλιση ρ.

Θεωρούμε ότι έχουμε μια εκτίμηση $\mathbf{d}_i$ για το $\mathbf{d}$ και προσπαθούμε να τη βελτιώσουμε κατά $\mathbf{u}$, δηλαδή $\mathbf{d}_{i+1} = \mathbf{d}_i + \mathbf{u}$. Αναπτύσσοντας κατά Taylor την έκφραση $I_{n-1}(\mathbf{x}+\mathbf{d}) = I_{n-1}(\mathbf{x}+\mathbf{d}_i+\mathbf{u})$ γύρω από το σημείο $\mathbf{x} + \mathbf{d}_i$, προκύπτει ότι

$$I_{n-1}(\mathbf{x} + \mathbf{d}) \approx I_{n-1}(\mathbf{x} + \mathbf{d}_i) + \nabla I_{n-1}(\mathbf{x} + \mathbf{d}_i)^T \mathbf{u} \tag{4}$$

Βάζοντας αυτήν την έκφραση στην Εξ. (3) μπορεί να δειχθεί ότι η λύση ελάχιστων τετραγώνων για τη βελτίωση του πεδίου οπτικής ροής σε κάθε σημείο είναι

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

όπου

$$A(\mathbf{x}) = \begin{bmatrix} A_1(\mathbf{x}) & A_2(\mathbf{x}) \end{bmatrix} = \begin{bmatrix} \dfrac{\partial I_{n-1}(\mathbf{x} + \mathbf{d}_i)}{\partial x} & \dfrac{\partial I_{n-1}(\mathbf{x} + \mathbf{d}_i)}{\partial y} \end{bmatrix} \tag{6}$$

$$E(\mathbf{x}) = I_n(\mathbf{x}) - I_{n-1}(\mathbf{x} + \mathbf{d}_i) \tag{7}$$

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 $\mathbf{d}_{i+1} = \mathbf{d}_i + \mathbf{u}$, με το $\mathbf{u}$ να υπολογίζεται από την Εξ. (5), επαναλαμβάνεται αρκετές φορές ως τη σύγκλιση.

Υλοποιείστε τον αλγόριθμο των Lucas-Kanade στο περιβάλον Matlab. Ο αλγόριθμος να υλοποιηθεί ως αυτόνομη συνάρτηση, που να δέχεται ως εισόδους δύο εικόνες (κομμένα παράθυρα με βάση το bounding box από δυο διαδοχικά πλαίσια του βίντεο , το εύρος ρ του γκαουσιανού παραθύρου, την θετική σταθερά κανονικοποίησης ϵ, και την αρχική εκτίμηση $\mathbf{d}_0$ για το πεδίο οπτικής ροής, και να επιστρέφει το $\mathbf{d}$, π.χ.

```
[d_x,d_y] = lk(I1, I2, rho, epsilon, d_x0, d_y0)
```

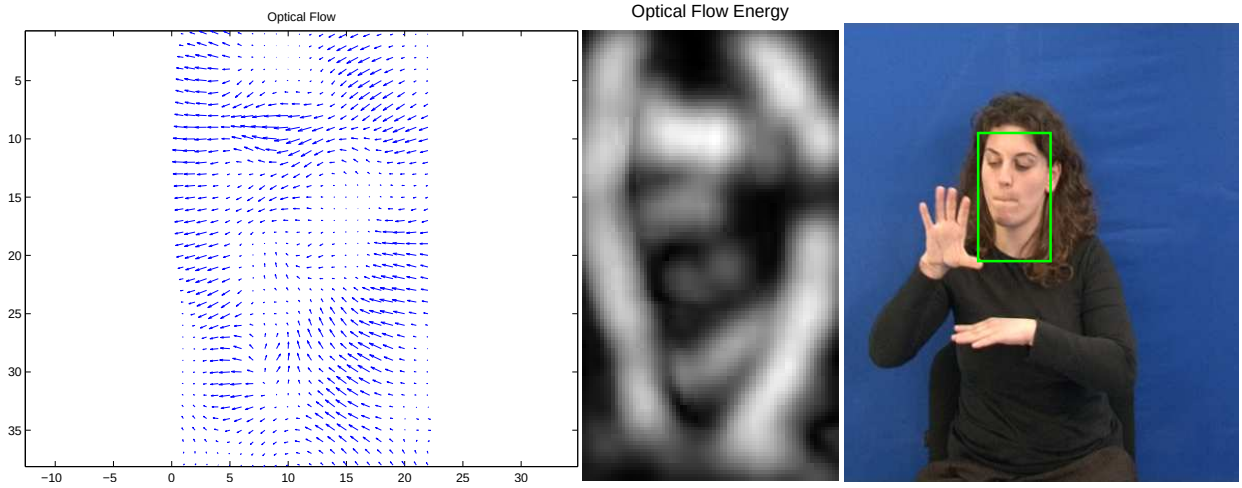
Χρήσιμες οδηγίες:

- Για τον υπολογισμό της $I_{n-1}(\mathbf{x} + \mathbf{d}_i)$ και των $\frac{\partial I_{n-1}}{\partial x}(\mathbf{x} + \mathbf{d}_i)$, $\frac{\partial I_{n-1}}{\partial y}(\mathbf{x} + \mathbf{d}_i)$ θα χρειαστείτε τις τιμές της I_{n-1} και των μερικών παραγώγων της σε ενδιάμεσα σημεία του πλέγματος. Χρησιμοποιείστε την interp2(I1,x_0+dx_i,y_0+dy_i,'linear',0), όπου [x_0,y_0] = meshgrid(1:size(I1,2),1:size(I1,1)), και όμοια για τις μερικές παραγώγους.

- Συνήθως η διαδικασία $\mathbf{d}_{i+1} = \mathbf{d}_i + \mathbf{u}$ συγκλίνει μετά από 5-10 επαναλήψεις. Πειραματιστείτε με εναλλακτικά κριτήρια σύγκλισης.

- Ενδεικτικές τιμές παραμέτρων: $\rho \in [1, 5]$ pixels και $\epsilon \in [0.01, 0.1]$ (για τιμές της εικόνας στο $[0, 1]$). Πειραματιστείτε με πολύ διαφορετικές τιμές των παραμέτρων ρ και ϵ και σχολιάστε πώς το αποτέλεσμα αλλάζει.

- Για να απεικονίσετε το $\mathbf{d}$ ως διανυσματικό πεδίο όπως στο Σχ. 3(α) χρησιμοποιείστε τη συνάρτηση `quiver(-d_x_r,-d_y_r)`, με υποδειγματοληπτημένες εκδοχές των `d_x, d_y`, π.χ. `d_x_r=imresize(d_x,0.3)`.



(α) (β) (γ)

Σχήμα 3: 3ο πλαίσιο της ακολουθίας βίντεο νοηματικής γλώσσας. (α) Πεδίο οπτικής ροής. (β) Ενέργεια διανυσμάτων οπτικής ροής. (γ) Τελική μετατόπιση ορθογωνίου.

▶ Βοήθεια για Matlab: συναρτήσεις `meshgrid`, `interp2`, `fspecial`, `imfilter`, `quiver`.

1.3 Πολυ-Κλιμακωτός Υπολογισμός Οπτικής Ροής

(Προαιρετικό για τους προπτυχιακούς, υποχρεωτικό για τους μεταπτυχιακούς[1])

Υλοποιείστε την πολυ-κλιμακωτή εκδοχή του αλγόριθμου των Lucas-Kanade. Ο αλγόριθμος θα αναλύει τις αρχικές εικόνες σε γκαουσιανές πυραμίδες και θα 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 κλίμακα. Ο αλγόριθμος να υλοποιηθεί ως αυτόνομη συνάρτηση, παρόμοια με πριν, αλλά να δέχεται επίσης ως είσοδο τον αριθμό των κλιμάκων της πυραμίδας, και να χρησιμοποιεί τον αλγόριθμο των Lucas-Kanade μονής κλίμακας ως υπο-ρουτίνα. Χρήσιμες οδηγίες:

- Για τη μετάβαση από μεγάλες σε μικρές κλίμακες κατά την κατασκευή της γκαουσιανής πυραμίδας φιλτράρετε την εικόνα με βαθυπερατό φίλτρο (π.χ. γκαουσιανή τυπικής απόκλισης 3 pixels) πριν την υποδειγματοληψία για να μετριάσετε τη φασματική αναδίπλωση (aliasing) της εικόνας.
- Κατά τη μεταφορά του $\mathbf{d}$ από μικρές σε μεγάλες κλίμακες μην ξεχάσετε να διπλασιάσετε το $\mathbf{d}$: `2*imresize`.

Τρέξτε στη συνέχεια τον πολυ-κλιμακωτό αλγόριθμό σας στην ίδια ακολουθία εικόνων και Σχολιάστε τις διαφορές που παρατηρείτε στην ταχύτητα σύγκλισης και στην ποιότητα του αποτελέσματος σε σχέση με τον αλγόριθμο μονής κλίμακας.

1.4 Υπολογισμός της Μετατόπισης του Προσώπου από το Πεδίο Οπτικής Ροής

Έχοντας υπολογίσει την οπτική ροή της εικόνας I_n στην περιοχή που είχε ορίσει το bounding box της εικόνας I_{n-1}, απομένει να βρούμε το συνολικό διάνυσμα μετατόπισης του bounding box ορθογωνίου, με όσο το δυνατόν μεγαλύτερη ακρίβεια. Είναι εύκολο να παρατηρήσουμε ότι τα διανύσματα του πεδίου της οπτικής ροής έχουν μεγαλύτερο μήκος σε σημεία που ανήκουν σε περιοχές με έντονη πληροφορία υφής (π.χ. ακμές, κορυφές) και σχεδόν μηδενικό μήκος σε

[1] Όσοι προπτυχιακοί επιλέξουν να κάνουν το Μέρος 1.3, θα έχουν bonus 20% επί αυτής της εργαστηριακής άσκησης.

σημεία που ανήκουν σε περιοχές με ομοιόμορφη και επίπεδη υφή. Επομένως, ο υπολογισμός της μέσης τιμής των διανυσμάτων μετατόπισης οδηγεί σε ανακριβή αποτελέσματα. Για να πετύχουμε καλύτερη ακρίβεια, μπορούμε να υπολογίσουμε τη μέση τιμή των διανυσμάτων μετατόπισης που έχουν ενέργεια μεγαλύτερη από μια τιμή κατωφλίου. Ως ενέργεια διανύσματος ταχύτητας ορίζουμε $\|\mathbf{d}\|^2 = {d_x}^2 + {d_y}^2$ και ένα παράδειγμα της εικόνας ενέργειας οπτικής ροής φαίνεται στο Σχ. 3(β).
Υλοποιείστε συνάρτηση που θα δέχεται σαν είσοδο τα διανυσματικά πεδία οπτικής ροής και θα υπολογίζει το τελικό διάνυσμα μετατόπισης του ορθογωνίου του προσώπου, π.χ.

```
[displ_x,displ_y] = displ(d_x, d_y)
```

Εκτελέστε το συνολικό σύστημα παρακολούθησης προσώπου για την ακολουθία εικόνων βίντεο που περιέχονται στο αρχείο **GreekSignLanguage.zip**. Πρόκειται για βίντεο ελληνικής νοηματικής γλώσσας γυρισμένο σε στούντιο με ελεγχόμενο φωτισμό. Πειραματιστείτε με διαφορετικές τιμές των παραμέτρων ρ, ϵ και κατωφλίου ενέργειας οπτικής ροής και παρατηρείστε πως επηρεάζεται το τελικό αποτέλεσμα. Πειραματιστείτε με εναλλακτικά κριτήρια υπολογισμού της μετατόπισης από το πεδίο οπτικής ροής.

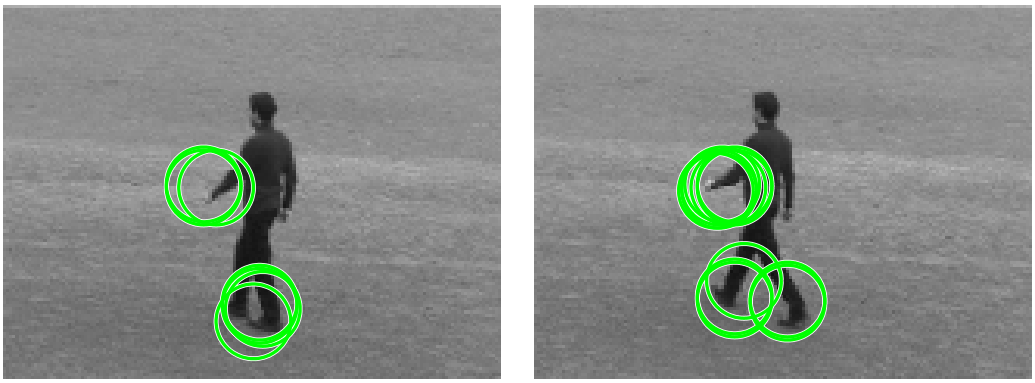
ΠΑΡΑΔΟΤΕΑ: Ζητείται να παραδώσετε τα εξής:

- Image plots (σχήματα εικόνων) για όλα τα βήματα της ανίχνευσης προσώπου με ανίχνευση δέρματος, όπως την επιφάνεια της Γκαουσιανής κατανομής, την εικόνα πιθανότητας δέρματος, τις δυαδικές εικόνες δέρματος πριν και μετά τη μορφολογική επεξεργασία και την τελική ανίχνευση προσώπου.
- Τελικές εικόνες της παρακολούθησης προσώπου με το ορθογώνιο παρακολούθησης σχεδιασμένο σε καθεμιά.
- Εικόνες με το πεδίο οπτικής ροής και την ενέργεια των διανυσμάτων για ενδεικτικά πλαίσια της ακολουθίας εικόνων που σας δίνεται.
- Matlab scripts (λίστες εντολών).
- Τελική αναφορά που να περιλαμβάνει τις πιο σημαντικές εικόνες αποτελεσμάτων και συνοπτική επεξήγηση των αλγορίθμων.

Μέρος 2: Εντοπισμός Χωρο-χρονικών Σημείων Ενδιαφέροντος και Εξαγωγή Χαρακτηριστικών σε Βίντεο Ανθρωπίνων Δράσεων

Στην εργαστηριακή αυτή άσκηση θα ασχοληθούμε με την εξαγωγή χωρο-χρονικών χαρακτηριστικών με στόχο την εφαρμογή τους στο πρόβλημα κατηγοριοποίησης βίντεο που περιέχουν ανθρώπινες δράσεις. Όπως έχουμε ήδη δει τα τοπικά χαρακτηριστικά (local features) έχουν δείξει τεράστια επιτυχία σε διάφορα προβλήματα αναγνώρισης της Όρασης Υπολογιστών, όπως η αναγνώριση αντικειμένων. Οι τοπικές αναπαραστάσεις περιγράφουν το προς παρατήρηση αντικείμενο με μια σειρά από τοπικούς περιγραφητές που υπολογίζονται σε γειτονιές ανιχνευθέντων σημείων ενδιαφέροντος. Τελικά, η συλλογή των τοπικών χαρακτηριστικών ενσωματώνεται σε μια τελική αναπαράσταση global representation (π.χ. bag of visual words) ικανή να αναπαραστήσει τη στατιστική κατανομή τους και να προχωρήσει στα επόμενα στάδια της αναγνώρισης.

Η αναπαράσταση με χρήση τοπικών χαρακτηριστικών έχει επικρατήσει και στην αναγνώριση ανθρώπινων δράσεων, όπου γίνεται μια επιλογή από δεδομένα που αφ΄ ενός μειώνουν κατά πολύ τη διάσταση των βίντεο και αφ΄ ετέρου τα μετασχηματίζουν σε μια αναπαράσταση που τα κάνει



(α) (β)

Σχήμα 4: Παράδειγμα ανίχνευσης χωρο-χρονικών σημείων ενδιαφέροντος (Harris Detector).

κατηγοριοποιήσιμα. Στα πλαίσια της άσκησης θα σας δοθούν βίντεο από 3 κλάσεις δράσεων (walking,running,boxing) από τα οποία θα εξάγεται χωρο-χρονικούς περιγραφητές με σκοπό κατηγοριοποίηση την κατηγοριοποίηση των δράσεων αυτών.

Τα βίντεο θα τα διαβάσετε καλώντας την συνάρτηση read_video(name,nframes,0), όπου name το όνομα του βίντεο και nframes ο αριθμός των frames (200) που θέλετε να διαβάσετε. Το video θα αναπαριστάται με ένα τρισδιάστατο πίνακα, όπου η 3η διάσταση αποτελεί την ακολουθία των frames, τα οποία είναι grayscale εικόνες.

2.1 Χωρο-χρονικά Σημεία Ενδιαφέροντος

Οι ανιχνευτές τοπικών χαρακτηριστικών αναζητούν χωρο-χρονικά σημεία και κλίμακες ενδιαφέροντος που αντιστοιχούν σε περιοχές που χαρακτηρίζονται από σύνθετη κίνηση ή απότομες μεταβολές στην εμφάνιση του video εισόδου μεγιστοποιώντας μια συνάρτηση οπτικής σημαντικότητας. Πολλοί ανιχνευτές έχουν επινοηθεί τα τελευταία χρόνια αντλώντας αραιά αλλά εύρωστα σημεία [7]. Στην εργαστηριακή αυτή άσκηση θα ασχοληθούμε με 2 διαφορετικούς τέτοιους ανιχνευτές: 1) Harris detector [4] και 2) Gabor detector [2].

2.1.1 Υλοποιήστε τον ανιχνευτή Harris ο οποίος αποτελεί μια επέκταση σε 3 διαστάσεις του ανιχνευτή γωνιών Harris-Stephens, που υλοποιήσατε στην 1η εργαστηριακή άσκηση. Για κάθε voxel του βίντεο υπολογίστε τον 3×3 πίνακα $M(x, y, t)$ προσθέτοντας στον 2Δ δομικό τανυστή και τη χρονική παράγωγο:

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

όπου $g(x, y, t; s\sigma, s\tau)$ ένας 3Δ γκαουσιανός πυρήνας ομαλοποίησης και $\nabla L(x, y, t; \sigma, \tau)$ οι χωρο-χρονικές παράγωγοι για την χωρική κλίμακα σ και τη χρονική κλίμακα τ. Τις παραγώγους (χωρικές και χρονικές) μπορείτε να τις υπολογίσετε εφαρμόζοντας συνέλιξη με τον πυρήνα κεντρικών διαφόρων $[-1\,0\,1]^T$ (προσαρμοσμένο στην κατάλληλη διάσταση).

Το 3Δ κριτήριο γωνιότητας ακολουθεί και αυτό την ίδια λογική:

$$H(x, y, t) = det(M(x, y, t)) - k \cdot trace^3(M(x, y, t))$$

2.1.2 Υλοποιήστε τον ανιχνευτή Gabor ο οποίος βασίζεται στο χρονικό φιλτράρισμα του βίντεο με ένα ζεύγος Gabor φίλτρων αφού πρώτα αυτό έχει υποστεί εξομάλυνση στις χωρικές διαστάσεις μέσω ενός 2Δ γκαουσιανού πυρήνα $g(x, y; \sigma)$ με τυπική απόκλιση σ. Τα Gabor ορίζονται ως:

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

Για τον υπολογισμό της κρουστικής απόκρισης των Gabor θεωρήστε μέγεθος παραθύρου $[-2\tau, 2\tau]$ και κανονικοποιήστε με την $L1$ νόρμα.

Η συχνότητα ω του Gabor φίλτρου συνδέεται με την χρονική κλίμακα τ (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:

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

2.1.3 Για κάθε ένα ανιχνευτή υπολογίστε τα σημεία ενδιαφέροντος σαν τα τοπικά μέγιστα του κριτηρίου σημαντικότητας. Για απλότητα, μπορείτε απλά να επιστρέψετε τα σημεία με τις μεγαλύτερες τιμές του κριτηρίου σημαντικότητας (π.χ. τα 500-600 πρώτα). Απεικονίστε για επιλεγμένα frames τα κριτήρια σημαντικότητας καθώς και τα σημεία που προκύπτουν χρησιμοποιώντας τη συνάρτηση `showDetection.m`. Μπορείτε να πειραματιστείτε με διαφορετικές χωρικές και χρονικές κλίμακες ή και με πολλαπλές κλίμακες. Τι παρατηρείτε ως προς τον τύπο των σημείων που ανιχνεύουν οι δύο μέθοδοι;

► *Τα N σημεία που επιστρέφουν οι ανιχνευτές που υλοποιήσατε πρέπει να είναι στην μορφή ενός πινάκα $N \times 4$. Οι δύο πρώτες στήλες αντιστοιχούν στις συντεταγμένες τους (x, y), η τρίτη στην κλίμακα σ στην οποία ανιχνεύθηκαν και η τέταρτη στο frame t στο οποίο ανιχνεύθηκαν. Με αυτή την μορφή τα διαβάζει και η συνάρτηση `showDetection.m`*

► Βοήθεια για Matlab: συναρτήσεις `imfilter,fspecial,convn,norm`.

2.2 Χωρο-χρονικοί Ιστογραφικοί Περιγραφητές

Οι χωρο-χρονικοί περιγραφητές που θα χρησιμοποιηθούν βασίζονται στον υπολογισμό ιστογραμμάτων της κατευθυντικής παραγώγου (HOG) και της οπτικής ροής (HOF - Histograms of Oriented Flow) [4] γύρω από τα σημεία ενδιαφέροντος που υπολογίσατε.

2.2.1 Για κάθε frame του βίντεο υπολογίστε το διάνυσμα κλίσης (gradient και οπτικής ροής χρησιμοποιώντας τις συναρτήσεις που έχετε υλοποιήσει στο πρώτο μέρος της άσκησης.

2.2.2 Στη συνέχεια χρησιμοποιήστε τη συνάρτηση `OrientationHistogram.p` προκειμένου να υπολογίσετε τους 2 ιστογραφικούς περιγραφητές. Η συνάρτηση αυτή δέχεται ως είσοδο το διανυσματικό πεδίο (κατευθυντικές παραγώγους είτε κατεύθυνση ροής), το μέγεθος του grid και το πλήθος των bins και επιστρέφει την ιστογραμματική περιγραφή της αντίστοιχης περιοχής. Εσείς καλείστε να εξάγετε τα διανυσματικά πεδία που απαιτούνται για μια (τετραγωνική) περιοχή $4 \times scale$ γύρω από το εκάστοτε σημείο ενδιαφέροντος (από την εικόνα που αντιστοιχεί στο frame που ανιχνεύσατε σημεία ενδιαφέροντος). Να δώσετε προσοχή στα όρια της εικόνας. Προαιρετικά μπορείτε να κατασκευάσετε μόνοι σας τα ιστόγραμμα με βάση το παράρτημα της 1ης εργαστηριακής άσκησης. Για τη δημιουργία του HOG/HOF περιγραφητή συνενώστε τους δύο επιμέρους περιγραφητές.

2.2.3 Υπολογίστε την τελική αναπαράσταση global representation για κάθε βίντεο υλοποιώντας την bag of visual words BoVW τεχνική που περιγράφεται στην 1η εργαστηριακή άσκηση. Σημειώστε, ότι εδώ δεν έχουμε train και test δεδομένα αλλά μόνο ένα σύνολο δεδομένων από το οποίο θα υπολογιστούν αρχικά οι λέξεις του λεξικού και στη συνέχεια τα ιστογράμματα εμφάνισης.

► *Η συνάρτηση `OrientationHistogram.p` για την εξαγωγή τοπικών περιγραφητών καλείται ως εξής desc = OrientationHistogram(Gx,Gy,nbins,[n m]), όπου Gx, Gy ορίζουν ένα διανυσματικό πεδίο, nbins ειναι το πλήθος των bins και $n \times m$ είναι το μέγεθος του grid. Η έξοδος desc είναι ένα ιστόγραμμα (περιγραφητής) για την περιοχή του διανυσματικού πεδίου εισόδου μεγέθους $n \times m \times nbins$.*

2.3 Κατασκευή Δενδρογράμματος για τον Διαχωρισμό των Δράσεων

Στο ερώτημα αυτό θα γίνει μια προσπάθεια κατανόησης της ικανότητας κατηγοριοποίησης των βίντεο με τις ανθρώπινες δράσεις σε 3 κατηγορίες/κλάσεις (που η κάθε μία θα αντιπροσωπεύει ένα διαφορετικό είδος δράσης) με χρήση των BoVW αναπαραστάσεων (που βασίζονται σε HOG / HOF χαρακτηριστικά) που υπολογίσατε στα προηγούμενα ερωτήματα. Αυτό θα επιτευχθεί ποιοτικά με την οπτικοποίηση της απόστασης των διανυσμάτων χαρακτηριστικών μέσω της κατασκευής ενός δενδρογράμματος αποστάσεων που αντιπροσωπεύει την ικανότητα διαχωρισμού των 3 διαφορετικών κατηγοριών.

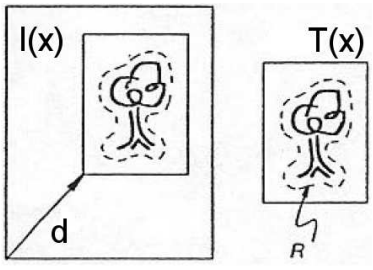
2.3.1 Κατασκευή του δενδρογράμματος αποστάσεων (αποτελεί οπτικοποίηση μιας μορφής ιεραρχικής κατηγοριοποίησης) από το σύνολο των BoVW ιστογραμμάτων, όπως υπολογιστήκαν σε προηγούμενα ερωτήματα. Παρατηρήστε την μορφή του και σχολιάστε τι συμπεράσματα μπορούν να εξαχθούν από αυτή. Προτείνεται η χρήση της χ^2 απόστασης, η οποία είναι κατάλληλη για ιστογράμματα, αλλά μπορείτε να πειραματιστείτε και με άλλες αποστάσεις. Η χ^2 για 2 ιστογράμματα $H_i = \{h_{i1}, h_{i2}, ..., h_{iK}\}$ και $H_j = \{h_{j1}, h_{j2}, ..., h_{jK}\}$ (όπου K το πλήθος των κέντρων) υπολογίζεται ως $D(H_i, H_j) = \frac{1}{2}\sum_{n=1}^{K} \frac{(h_{in}-h_{jn})^2}{h_{in}+h_{jn}}$.

► Βοήθεια για Matlab: συναρτήσεις `linkage,dendrogram,pdist,distChiSq`.

2.3.2 Πειραματιστείτε με τους διαφορετικούς συνδυασμούς ανιχνευτών/περιγραφητών και παρατηρήστε τις μεταβολές τόσο στην μορφή του δενδρογράμματος. Αναφέρετε το καλύτερο συνδυασμό που χρησιμοποιήσατε και σχολιάστε.

Παράρτημα: Αντιστοίχιση Εικόνων με τη Μέθοδο των Lucas και Kanade

Η εργασία των Lucas-Kanade [5] εισήγαγε ένα πλαίσιο στο οποίο εντάσσεται μια ευρεία οικογένεια αλγορίθμων για αντιστοίχιση εικόνων.
Θεωρούμε μια εικόνα I την οποία επιθυμούμε να αντιστοιχίσουμε με μια εικόνα αναφοράς (τεμπλέτα) T. Αυτές μπορεί για παράδειγμα να είναι ολόκληρα διαδοχικά πλαίσια βίντεο ή να έχουν ληφθεί από ζεύγος στέρεο καμερών. Εναλλακτικά, στην περίπτωση που στόχος είναι ο υπολογισμός του πεδίου οπτικής ροής τα I και T είναι μικρές υποπεριοχές (π.χ. μπλοκ διάστασης 8×8 πίξελ) εξαγμένες από τα πλαίσια του βίντεο.



Σχήμα 5: Μεταφορική αντιστοίχιση μεταξύ εικόνας I και τεμπλέτας T. Από το [5].

Υποθέτουμε επίσης ότι η σχετική κίνηση μεταξύ των αντίστοιχων περιοχών στις δυο εικόνες προσεγγίζεται ικανοποιητικά ως μεταφορική, δηλαδή $I(\mathbf{x}+\mathbf{d}) \approx T(\mathbf{x})$, με $\mathbf{d} = (d_x, d_y)^T$, όπως φαίνεται στο Σχ. 5. Αναζητούμε το διάνυσμα μετατόπισης $\mathbf{d}$ που ελαχιστοποιεί κάποιο κριτήριο σφάλματος αντιστοίχισης, π.χ. το μέσο τετραγωνικό σφάλμα

$$J(\mathbf{d}) = \sum_{\mathbf{x}\in R} \left(I(\mathbf{x}+\mathbf{d}) - T(\mathbf{x})\right)^2 . \tag{8}$$

Στον αλγόριθμο Lucas-Kanade θεωρούμε ότι έχουμε μια εκτίμηση $\mathbf{d}_i$ για το $\mathbf{d}$ και προσπαθούμε να τη βελτιώσουμε κατά $\mathbf{u} = (u_x, u_y)^T$, δηλαδή $\mathbf{d}_{i+1} = \mathbf{d}_i + \mathbf{u}$. Εάν αναπτύξουμε κατά Taylor την έκφραση $I(\mathbf{x} + \mathbf{d}) = I(\mathbf{x} + \mathbf{d}_i + \mathbf{u})$ γύρω από το σημείο $\mathbf{x} + \mathbf{d}_i$, προκύπτει η γραμμική προσέγγιση $I(\mathbf{x} + \mathbf{d}) \approx I(\mathbf{x} + \mathbf{d}_i) + A(\mathbf{x})\mathbf{u}$, όπου έχουμε ορίσει τον 1×2 πίνακα μερικών παραγώγων της εικόνας $A(\mathbf{x}) = (\frac{\partial I}{\partial x}(\mathbf{x}+\mathbf{d}_i), \frac{\partial I}{\partial y}(\mathbf{x}+\mathbf{d}_i))$. Αντικαθιστώντας αυτήν την έκφραση στην Εξ. (8) προκύπτει

$$J(\mathbf{u}) = \sum_{\mathbf{x} \in R} \left(E(\mathbf{x}) + A(\mathbf{x})\mathbf{u} \right)^2 , \tag{9}$$

όπου $E(\mathbf{x}) = I(\mathbf{x} + \mathbf{d}_i) - T(\mathbf{x})$ είναι η εικόνα με το σφάλμα αντιστοίχισης κατά την τρέχουσα επανάληψη. Για να βελτιστοποιήσουμε τη J μηδενίζουμε τον πίνακα μερικών παραγώγων ως προς $\mathbf{u}$, οπότε προκύπτει το 2×2 γραμμικό σύστημα

$$0 = \frac{dJ}{d\mathbf{u}} = 2 \sum_{\mathbf{x} \in R} A(\mathbf{x})^T \left(E(\mathbf{x}) + A(\mathbf{x})\mathbf{u} \right) , \tag{10}$$

$$\text{ή} \quad \left(\sum_{\mathbf{x} \in R} A(\mathbf{x})^T A(\mathbf{x}) \right) \mathbf{u} = - \sum_{\mathbf{x} \in R} A(\mathbf{x})^T E(\mathbf{x}) \tag{11}$$

Η Εξ. (5) είναι παραλλαγή της (11) στην οποία έχουμε επίσης εφαρμόσει στο κριτήριο σφάλματος (8) γκαουσιανό παράθυρο στάθμισης, που δίνει μεγαλύτερη έμφαση στα κεντρικά πίξελ του μπλοκ, και επίσης έχουμε προσθέσει την κανονικοποιητική σταθερά μ στη διαγώνιο του πίνακα γραμμικού συστήματος.

Πέρα από το βασικό αλγόριθμο που περιγράφουμε, η τεχνική των Lucas-Kanade μπορεί να επεκταθεί για να καλύψει γενικότερα μοντέλα κίνησης (π.χ. περιστροφή ή αφινική κίνηση), καθώς και μοντέλα φωτομετρικής αλλοίωσης [3]. Μια πρόσφατη επισκόπηση της σχετικής βιβλιογραφίας μπορεί να βρεθεί στα [1, 6].

References

[1] S. Baker and I. Matthews. Lucas-Kanade 20 years on: A unifying framework. *Int. J. of Comp. Vis.*, 56(3):221–255, 2004.

[2] P. Dollár, V. Rabaud, G. Cottrell, and S. Belongie. Behavior recognition via sparse spatio-temporal features. In *Proc. IEEE Int'l Workshop on VS-PETS*, 2005.

[3] C.S. Fuh and P. Maragos. Motion displacement estimation using an affine model for image matching. *Optical Engin.*, (30):7, July 1991.

[4] I. Laptev, M. Marszalek, C. Schmid, and B. Rozenfeld. Learning realistic human actions from movies. In *Proc. IEEE Conf. CVPR*, 2008.

[5] B.D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proc. Int. Joint Conf. on Artificial Intel.*, pages 674–679, 1981.

[6] R. Szeliski. Image alignment and stitching: a tutorial. *Found. and Trends in Comp. Graph. and Vision*, 2(1):1–104, 2006.

[7] H. Wang, M. M. Ullah, A. Kläser, I. Laptev, and C. Schmid. Evaluation of local spatio-temporal features for action recognition. In *Proc. BMVC*, 2009.