

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΑΤΡΩΝ  
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ Η/Υ ΚΑΙ ΠΛΗΡΟΦΟΡΙΚΗΣ  
ΕΡΓΑΣΤΗΡΙΟ ΕΠΕΞΕΡΓΑΣΙΑΣ ΣΗΜΑΤΩΝ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

## ΥΠΟΛΟΓΙΣΤΙΚΗ ΟΡΑΣΗ

### SCALE INVARIANT FETUARE TRANSFORM (SIFT)

Διδάσκων Επίκουρος Καθηγητής Εμμανουήλ Ζ. Ψαράκης  
Επικουρικό έργο μεταπτυχιακή φοιτήτρια Καλλιόπη Μπασιώτη  
Πάτρα Δεκέμβριος 2016



## ΣΤΟΙΧΕΙΩΔΗΣ ΘΕΩΡΙΑ

Ο αλγόριθμος SIFT μπορεί να χωριστεί στα ακόλουθα τέσσερα (4) στάδια :

1. **Ανίχνευση Ακροτάτων στον Χώρο Κλίμακας:** Στο πρώτο στάδιο αναζητούμε ακρότατα στον χώρο κλίμακας αλλά και σε γειτονίες των εικόνων. Αυτό το στάδιο μπορεί να υλοποιηθεί αποδοτικά χρησιμοποιώντας την διαφορά Γκαουσιανών για την ανίχνευση ενδιαφέροντων σημείων (interesting points).
2. **Εύρεση Τοπικών Σημείων-Κλειδιών:** Κοιτάζοντας σε περιοχές των εικόνων επιλέγουμε από τα ενδιαφέροντα σημεία που βρήκαμε στο προηγούμενο στάδιο τα σημεία κλειδιά (Keypoints).
3. **Διόρθωση Κατεύθυνσης:** Για κάθε σημείο κλειδί εκτελούμε μία ή και παραπάνω περιστροφές ανάλογα με την κατεύθυνση της τοπικής κλίσης σε μια περιοχή της εικόνας.
4. **Περιγραφέας (descriptor) Σημείων-Κλειδιών** Υπολογίζονται οι τοπικές κλίσεις των εικόνων για κάποια συγκεκριμένη κλίμακα σε μια περιοχή γύρω από σημεία κλειδιά.

Τα παραπάνω στάδια αναλύονται στην συνέχεια.

### 1. Ανίχνευση Ακροτάτων στον Χώρο Κλίμακας

Αρχικά δημιουργούμε τον χώρο κλίμακας για την εικόνα ενδιαφέροντος. Μαθηματικά ο χώρος κλίμακας μιας εικόνας μπορεί να περιγραφεί από μια συνάρτηση  $L_\sigma(x, y)$ , που παράγεται από την συνέλιξη ενός γκαουσιανού πυρήνα  $G_\sigma(x, y)$  με την αρχική εικόνα  $I(x, y)$ :

$$L_\sigma(x, y) = G_\sigma(x, y) * I(x, y)$$

όπου  $(x, y)$  είναι οι συντεταγμένες των εικονοστοιχείων της εικόνας και του γκαουσιανού πυρήνα,  $\sigma$  η διασπορά του γκαουσιανού πυρήνα που ορίζεται ως:

$$G_\sigma(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$$

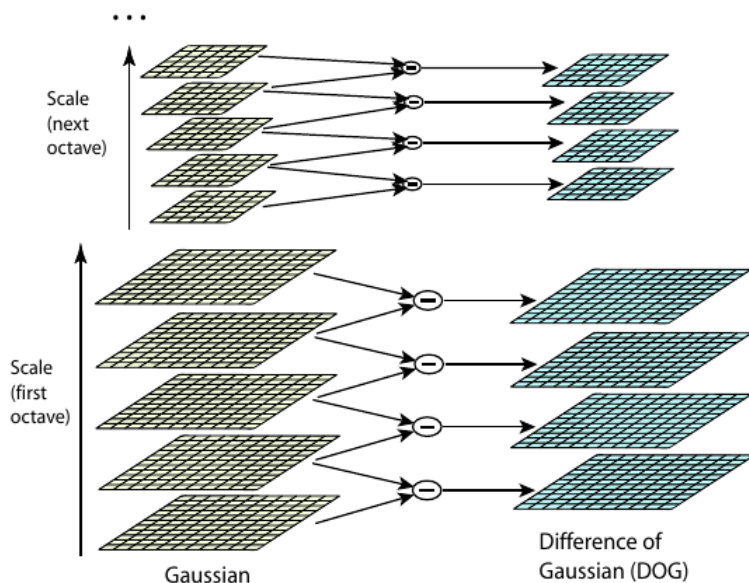
Στην συνέχεια υπολογίζουμε τις διαφορές των Γκαουσιανών  $D_\sigma(x, y)$  που μπορεί να υπολογιστούν από την διαφορά δύο γειτονικών εικόνων του χώρου κλίμακας που διαφέρουν κατά έναν σταθερό παράγοντα  $k$ , δηλαδή:

$$D_\sigma(x, y) = (G_{k\sigma}(x, y) - G_\sigma(x, y)) * I(x, y) = L_{k\sigma}(x, y) - L_\sigma(x, y)$$

όπως φαίνεται στο Σχήμα 1.

### 2. Ανίχνευση Τοπικών Ακροτάτων

Για την εύρεση των τοπικών ακροτάτων (ελαχίστων και μεγίστων) της  $D_\sigma(x, y)$  κάθε εικονοστοιχείο συγκρίνεται με τα οκτώ γειτονικά του εικονοστοιχεία στην τρέχουσα κλίμακα και τους εννιά γείτονες του στην προηγούμενη και την επόμενη κλίμακα, όπως φαίνεται στο Σχήμα 2. Το συγκεκριμένο εικονοστοιχείο επιλέγεται ως ακρότατο αν και μόνο αν είναι μεγαλύτερο ή μικρότερο από όλους τους γείτονές του.



Σχήμα 1: Για κάθε οκτάβα του χώρου κλίμακας, η αρχική εικόνα συνελίσσεται διαδοχικά με γκαουσιανούς πυρήνες για την δημιουργία του συνόλου των εικόνων που φαίνονται αριστερά. Γειτονικές εικόνες στον χώρο κλίμακας αφαιρούνται για την δημιουργία των *DOG* εικόνων που φαίνονται δεξιά.

### 3. Ακριβής Εύρεση των Σημείων Κλειδιών

Από τα παραπάνω εντοπισθέντα σημεία τα οποία είναι τοπικά ακρότατα, σε αυτό το βήμα θα επιλέξουμε από αυτά τα σημεία κλειδιά. Συγκεκριμένα, θα απορρίψουμε σημεία που μπορεί να οφείλονται σε θόρυβο ή η γειτονιά τους δεν εξασφαλίζει μεγάλη αντίθεση. Με αυτό τον τρόπο θα μπορέσουμε να έχουμε πιο σταθερά (stable) σημεία κλειδιά, και επομένως σημεία των οποίων ο περιγραφέας θα εξασφαλίζει μοναδικότητα και κατά συνέπεια εύκολη διακρισιμότητα του σημείου.

### 4. Διόρθωση Κατεύθυνσης

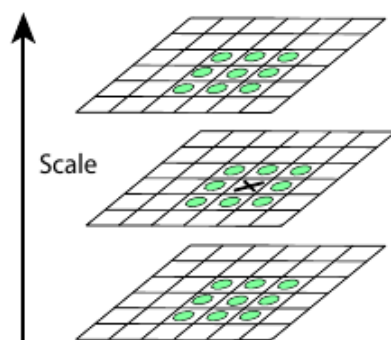
Εφαρμόζοντας κατάλληλες περιστροφές σε κάθε σημείο-κλειδί που βασίζονται σε κριτήρια που αφορούν μια τοπική περιοχή γύρω από το σημείο κλειδί, μπορούμε να έχουμε έναν περιγραφέα σημείων-κλειδιών ανεξάρτητο των μετασχηματισμών περιστροφής.

Επομένως για κάθε εικόνα  $L_{\sigma_0}(x, y)$  της κλίμακας που αντιστοιχεί το κάθε σημείο κλειδί υπολογίζουμε το μέτρο  $m(x, y)$  της κλίσης και την κατεύθυνσή της  $\theta(x, y)$ :

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2}$$

$$\theta(x, y) = \tan^{-1} \left( \frac{L(x, y+1) - L(x, y-1)}{L(x+1, y) - L(x-1, y)} \right).$$

Στην συνέχεια κατασκευάζουμε ιστογράμματα "κατεύθυνσης" ή "προσανατολισμού" υπολογίζοντας τα παραπάνω σε μια περιοχή γύρω από το σημείο κλειδί. Το ιστόγραμμα κατεύθυνσης έχει 36 bins που καλύπτουν τις 360 μοίρες δυνατής



Σχήμα 2: Τα τοπικά ακρότατα των *DoG* εικόνων ανιχνεύονται συγκρίνοντας το εικονοστοιχείο που έχει μαρκαριστεί με  $x$  με τους 26 γείτονές του σε μια περιοχή  $3 \times 3$  στην τρέχουσα και τις προσκείμενες κλίμακες.

περιστροφής. Κάθε δείγμα προστίθεται στο ιστόγραμμα σύμφωνα με την κατεύθυνση της κλίσης και με βάρος που δίνεται από το γινόμενο του μέτρο της και ενός γκαουσιανού κυκλικού παραθύρου με  $\sigma$  που είναι μιάμιση φορά το μέγεθος της κλίμακας που εντοπίστηκε το σημείο κλειδί. Οι μεγαλύτερες τιμές στο ιστόγραμμα μας δείχνουν την κυρίαρχη κατεύθυνση των τοπικών κλίσεων. Εντοπίζουμε λοιπόν την μεγαλύτερη τιμή του ιστογράμματος, και στην συνέχεια και κάθε άλλο τοπικό μέγιστό του που έχει τιμή πάνω από το 80% του ολικού μέγιστου μπορεί να χρησιμοποιηθεί για την δημιουργία σημείου κλειδού με αυτή την κατεύθυνση.

### 5. Τοπικός Περιγραφέας Σημείου Κλειδιού

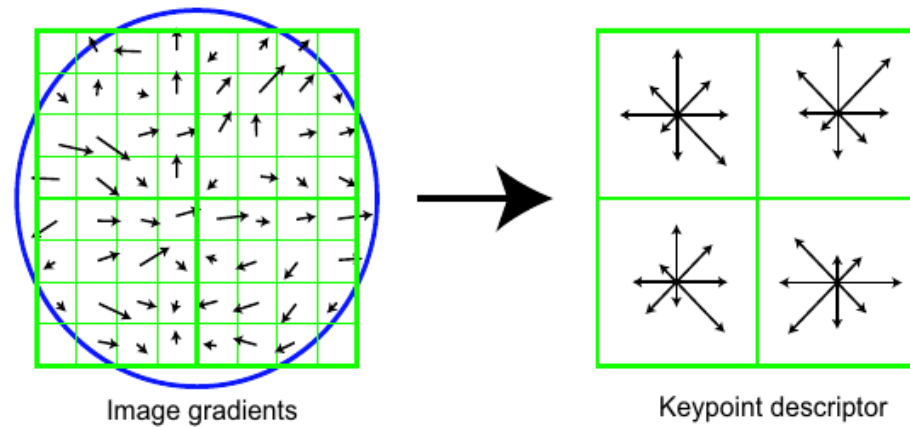
Με τα προηγούμενα βήματα μπορέσαμε να βρούμε για κάθε σημείο κλειδί την θέση του στην εικόνα, την κλίμακα και την κατεύθυνση του.

Στο Σχήμα 3 βλέπουμε ένα  $2 \times 2$  array ιστογράμματος κατεύθυνσης.

## ΔΙΑΔΙΚΑΣΙΑ

1. Μελετήστε το m-file `SIFT_feature.m`<sup>1</sup> που σας δίνεται και εντοπίστε μέσα σε αυτό τα τέσσερα (4) στάδια που περιγράψαν παραπάνω.
2. Για καθένα από αυτά καταγράψτε με αυστηρό τρόπο τα βασικά βήματα που ακολουθούνται.
3. Πόσα επίπεδα χρησιμοποιούνται ανά οκτάδα στον αλγόριθμο;
4. Περιγράψτε διαφορετικές τεχνικές με τις οποίες θα μπορούσαμε να εντοπίσουμε την βέλτιστη, με ακρίβεια μεγαλύτερη από αυτήν που μας προσφέρει η ομοιόμορφη διακριτοποίηση του άξονα της κλίμακας, ενός σημείου κλειδιού. Καταγράψτε τα πλεονεκτήματα και τα μειονέκτηματά κάθε μίας από τις προτεινόμενες τεχνικές.
5. Για πέντε (5) εικόνες της επιλογής σας τρέξτε το m-file `SIFT_feature.m` και καταγράψτε και σχολιάστε τα αποτελέσματά σας.

<sup>1</sup>Στο *m - file* αυτό έχει υλοποιηθεί ο SIFT που περιγράφεται στην αναφορά 1, από τον συγγραφέα της.



Σχήμα 3: Ο περιγραφέας ενός σημείου κλειδιού δημιουργείται υπολογίζοντας αρχικά το μέτρο και την κατεύθυνση της κλίσης για κάθε εικόνα σε μια περιοχή γύρω από το σημείο κλειδί, όπως περιγράψαμε στην παραπάνω παράγραφο, και όπως φαίνεται αριστερά της εικόνας.

6. Υλοποιείτε αλγόριθμο αντιστοίχισης που θα βασίζεται στην αντιστοίχιση των περιγραφών των σημείων κλειδιών που προέκυψαν από την εφαρμογή του SIFT σε οποιοδήποτε ζευγάρι εικόνων της επιλογής σας.
7. Περιγράψτε τον αλγόριθμο RANSAC και καταγράψτε τον τρόπο με τον οποίο τον χρησιμοποιείται στον αλγόριθμο σας.
8. Σχολιάστε την ποιότητα της λύσης του προβλήματος αντιστοίχισης. Καταγράψτε τρόπους με τους οποίους θα μπορούσαμε να βελτιώσουμε το βασικό πρόβλημα της λύσης.
9. Καταγράψτε τις βασικές διαφορές μεταξύ των αλγορίθμων SIFT, GLOH (Gradient Location and Orientation Histogram) και SURF (Speeded Up Robust Features).
10. Ποιό είναι το υπολογιστικό κόστος των παραπάνω αλγορίθμων;
11. Καταγράψτε τις βιβλιογραφικές πηγές που χρησιμοποιήσατε.

## 0.1 Αναφορές

1. David G. Lowe, *Distinctive Image Features from Scale-Invariant Keypoints*, International Journal of Computer Vision, pp. 91-110, 2004.