

Πανεπιστήμιο Πατρών
Τμήμα Μηχανικών Η/Υ και Πληροφορικής

Θέματα
Όρασης Υπολογιστών & Γραφικής

Scale Invariant Feature Transform
(SIFT)

Αναφορά Άσκησης 3

Όνομα: Βασιλική
Επώνυμο: Στάμου
Α.Μ.: 1059543

Πάτρα Αύγουστος 2022

1. Μελετήστε το SIFT_feature.m που σας δίνεται και εντοπίστε μέσα σε αυτό τα τέσσερα (4) στάδια που περιγράφουν παραπάνω.

Ο κώδικας που δίνεται, είναι χωρισμένος στα 4 στάδια της διαδικασίας που περιγράφηκε στην στοιχειώδη θεωρία.

Αναλυτικότερα έχουμε:

1. Ανίχνευση Ακροτάτων στο Χώρο κλίμακας: γραμμές 20 - 68
2. Εύρεση Τοπικών Σημείων – Κλειδιών: γραμμές 69 - 189
Το συγκεκριμένο τμήμα κώδικα χωρίζεται σε δύο υπομέρη. Αρχικά, στις γραμμές 69 – 116 βρίσκονται τα τοπικά σημεία – κλειδιά και έπειτα, στις γραμμές 117 – 189 απορρίπτονται τα σημεία που έχουν χαμηλή αντίθεση ή είναι κακώς εντοπισμένα σε κάποια άκρη.
3. Διόρθωση Κατεύθυνσης: γραμμές 190 – 260
4. Περιγραφέας (descriptor) Σημείων – Κλειδιών: γραμμές 261 – 344

2. Για κάθε ένα από αυτά καταγράψτε με αυστηρό τρόπο τα βασικά βήματα που ακολουθούνται.

1. Ανίχνευση Ακροτάτων στο Χώρο κλίμακας:

Στο στάδιο αυτό, στόχος είναι ο προσδιορισμός θέσεων στην εικόνα, οι οποίες να είναι αναλλοίωτες σε αλλαγές κλίμακας.

Αυτό είναι κάτι που επιτυγχάνεται αναζητώντας ευσταθή χαρακτηριστικά σε όλες τις δυνατές κλίμακες, χρησιμοποιώντας μια συνάρτηση κλίμακας γνωστή ως χώρος κλίμακας. Ο χώρος κλίμακας αναπαριστά μια εικόνα ως μια μονοπαρμετρική οικογένεια ομαλών καμπυλών με στόχο την προσομοίωση της απώλειας πληροφορίας που θα υπάρξει αν η κλίμακα μιας εικόνας ελαττωθεί. Η παράμετρος που ελέγχει αυτήν την εξομάλυνση ονομάζεται παράμετρος κλίμακας.

Στον αλγόριθμο SIFT, για την υλοποίηση αυτής της εξομάλυνσης χρησιμοποιούνται οι πυρήνες Gauss και άρα η παράμετρος κλίμακας είναι η τυπική απόκλιση σ . Η χρήση τέτοιων πυρήνων επιλέχθηκε, καθώς επιτυγχάνουν την εξομάλυνση ικανοποιώντας ταυτόχρονα σημαντικούς περιορισμούς, όπως είναι η γραμμικότητα και το αναλλοίωτο ως προς την μετατόπιση.

Επομένως, ο χώρος κλίμακας μιας εικόνας (κατά προτίμηση αποχρώσεων του γκρι) κατασκευάζεται από την (χωρική) συνέλιξη της εικόνας, με έναν πυρήνα Gauss μεταβλητής κλίμακας.

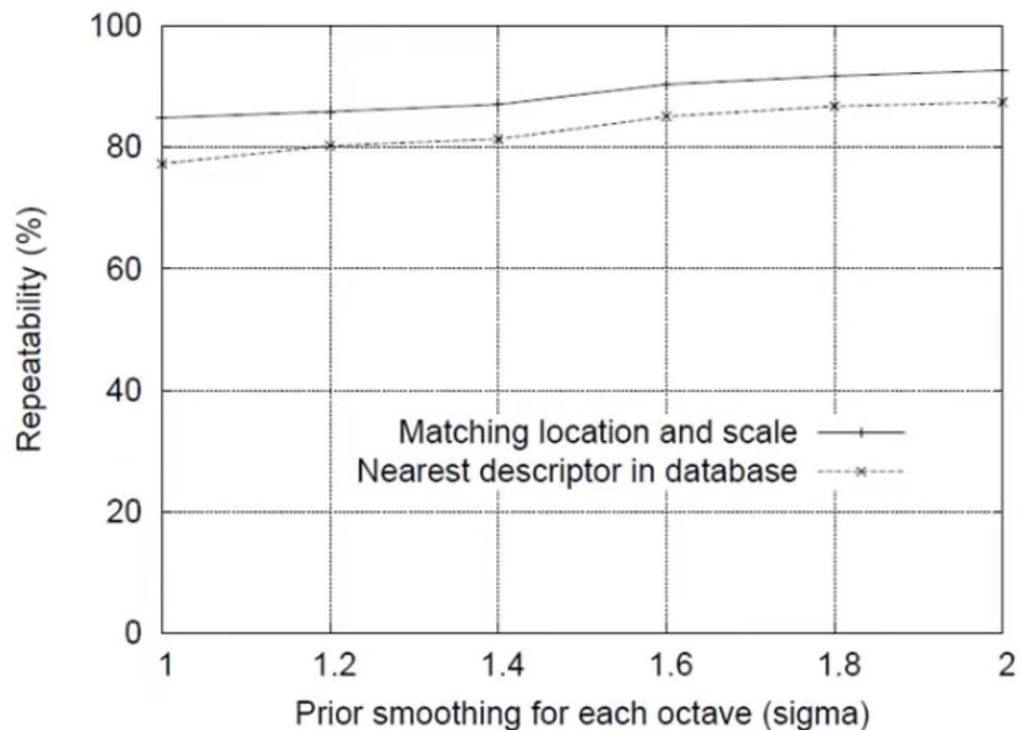
$$L(x, y, \sigma) = G(x, y, \sigma) \star f(x, y)$$

Ο υπολογισμός των διάφορων gaussian, γίνεται από την διαφορά των συναρτήσεων Gauss δύο γειτονικών εικόνων του χώρου κλίμακας σε μια οκτάβα που συνελίσσονται με την εικόνα εισόδου που αντιστοιχεί σε εκείνη την οκτάβα. Δηλαδή, τη διαφορά δύο γειτονικών εικόνων στο χώρο κλίμακας που διαφέρουν κατά σταθερό παράγοντα k . Μαθηματικά αποτυπώνεται ως εξής:

$$D(x, y, s) = [G(x, y, k\sigma) - G(x, y, \sigma)] \star f(x, y)$$

$$D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma)$$

Σχετικά με την επιλογή της αρχικής τιμής του σ :



2. Εύρεση Τοπικών Σημείων – Κλειδιών:

Αφού υπολογιστεί η θέση των σημείων κλειδιών με την διαδικασία που περιγράφηκε παραπάνω, ο αλγόριθμος SIFT προχωρά στην βελτίωση των θέσεων και της εγκυρότητας αυτών των σημείων κλειδιών με δύο βήματα επεξεργασίας. Για αυτό το λόγο, όπως είδαμε και ερώτημα 1, το συγκεκριμένο τμήμα κώδικα είναι χωρισμένο σε δύο υπομέρη.

Για τον προσδιορισμό των ακροτάτων, σε κάθε θέση μιας εικόνας $D(x,y,\sigma)$, η τιμή του εικονοστοιχείου σε αυτή τη θέση συγκρίνεται με τις τιμές των 8 γειτόνων στη συγκεκριμένη εικόνα και με τους 9 γείτονες στην πάνω και στην κάτω εικόνα. Το σημείο επιλέγεται ως ακρότατο αν η τιμή του είναι μεγαλύτερη (μέγιστο) ή μικρότερη (ελάχιστο) από όλες τις τιμές των γειτόνων του. Αξίζει να σημειωθεί, πως στην πρώτη και στην τελευταία κλίμακα μιας οκτάβας δεν ανιχνεύονται ακρότατα σημεία, επειδή δεν υπάρχει εικόνα χαμηλότερης ή υψηλότερης, αντίστοιχα, κλίμακας ίδιου μεγέθους.

Για την βελτίωση στην ακρίβεια των θέσεων των σημείων κλειδιών, ο Lowe χρησιμοποιεί το γραμμικό και τον δευτεροβάθμιο όρο του αναπτύγματος Taylor (για το προηγούμενο βήμα είχαμε εκφράσει το DoG με την βοήθεια του Heat Equation και προσεγγιστικά με την πρώτη παράγωγο) της συνάρτησης $D(x,y,\sigma)$ με μια μετατόπιση, ώστε η αρχή της να βρίσκεται στη θέση του σημείου – δείγματος που εξετάζεται.

$$D(x) = D + \left(\frac{\partial D}{\partial x}\right)^T x + \frac{1}{2} x^T \frac{\partial}{\partial x} \left(\frac{\partial D}{\partial x}\right) x$$

Θέτοντας την παράγωγο της συνάρτησης ίση με το 0, υπολογίζουμε την θέση του ακροτάτου σημείου.

$$\hat{x} = -H^{-1}(\nabla D)$$

Όπου H ο εσσιανός πίνακας και ∇D η κλίση του D .

Η τιμή της συνάρτησης $D(\hat{x})$ στη θέση του ακροτάτου σημείου χρησιμοποιείται για την απόρριψη ασταθών ακροτάτων με χαμηλή αντίθεση. Συνδυάζοντας τα παραπάνω, έχουμε:

$$D(\hat{x}) = D + \frac{1}{2} (\nabla D)^T \hat{x}$$

Τα άλλα σημεία που θέλουμε να απορριφθούν, προκύπτουν από μεταπτώσεις έντασης που προκαλούνται από ακμές, καθώς τα σημεία που μας ενδιαφέρουν στον αλγόριθμο SIFT είναι χαρακτηριστικά που συνάδουν με κορυφές. Για τον σκοπό αυτό, εστιάζουμε στην τοπική καμπυλότητα. Μία ακμή χαρακτηρίζεται από μεγάλη καμπυλότητα προς την μια διεύθυνση και από χαμηλή καμπυλότητα ως προς την ορθογώνια προς αυτήν. Για τον υπολογισμό αυτό, χρησιμοποιείται ο 2x2 διαστάσεων εσσιανός πίνακας.

$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{yx} & D_{yy} \end{bmatrix}$$

Οι ιδιοτιμές του πίνακα H είναι ανάλογες προς τις καμπυλότητες του D . Αντί, όμως, να υπολογίσουμε απευθείας τις ιδιοτιμές, βασιζόμαστε στο ίχνος ($=\lambda_1 + \lambda_2$, ομοιότητα με Harris Corner Detector) και την ορίζουσα του H ($=\lambda_1 \lambda_2$), που ως γνωστόν για διαστάσεις 2x2 είναι εύκολο να υπολογιστούν. Επομένως, για να ελέγξουμε εάν ο λόγος των κύριων καμπυλοτήτων είναι μικρότερος από μια τιμή $r (= \lambda_1 / \lambda_2)$ που έχουμε επιλέξει ως κατώφλι, αρκεί:

$$\frac{[Tr(H)]}{Det(H)} < \frac{(r + 1)^2}{r}$$

3. Διόρθωση Κατεύθυνσης

Μέχρι αυτό το σημείο έχουμε υπολογίσει τα σημεία κλειδιά που θεωρούμε ως ευσταθή και καθώς γνωρίζουμε τη θέση κάθε σημείου – κλειδιού στο χώρο κλίμακας, έχουμε πετύχει την ανεξαρτησία από την κλιμάκωση. Σε αυτό το στάδιο, θα δούμε πως ο SIFT επιτυγχάνει το αναλλοίωτο ως προς την περιστροφή της εικόνας, εκχωρώντας σε κάθε σημείο – κλειδί έναν συνεπή προσανατολισμό με βάση τις τοπικές ιδιότητες της εικόνας.

Αρχικά, για να είναι ανεξάρτητοι από τις κλιμακώσεις οι υπολογισμοί που θα χρειαστούν, η κλίμακα του keypoint χρησιμοποιείται για την επιλογή της εξομαλυμένης gaussian εικόνας L , που είναι πιο κοντά σε αυτή την κλίμακα. Για κάθε δείγμα της εικόνας $L(x, y)$ σε αυτή την κλίμακα, υπολογίζεται το μέτρο της κλίσης $M(x, y)$ και η γωνία του προσανατολισμού $\theta(x, y)$, χρησιμοποιώντας διαφορές εικονοστοιχείων.

$$M(x, y) = [(L(x + 1, y) - L(x - 1, y))^2 + (L(x, y + 1) - L(x, y - 1))^2]^{1/2}$$

$$\theta(x, y) = \tan^{-1} [(L(x, y + 1) - L(x, y - 1)) / (L(x + 1, y) - L(x - 1, y))]$$

Κατασκευάζεται, στη συνέχεια, ένα ιστόγραμμα προσανατολισμών από τις κλίσεις (gradient) των προσανατολισμών των σημείων – δειγμάτων σε μία γειτονιά του κάθε keypoint.

Το ιστόγραμμα έχει 36 bins που καλύπτουν την περιοχή προσανατολισμών των 360° στο επίπεδο της εικόνας. Κάθε δείγμα που προστίθεται στο ιστόγραμμα σταθμίζεται από το μέτρο της κλίσης του και μια κυκλική συνάρτηση τύπου Gauss, με τυπική απόκλιση ίση με 1.5 επί την κλίμακα του keypoint. Οι κορυφές αυτού του ιστογράμματος αντιστοιχούν στις κυρίαρχες τοπικές διευθύνσεις των τοπικών κλίσεων.

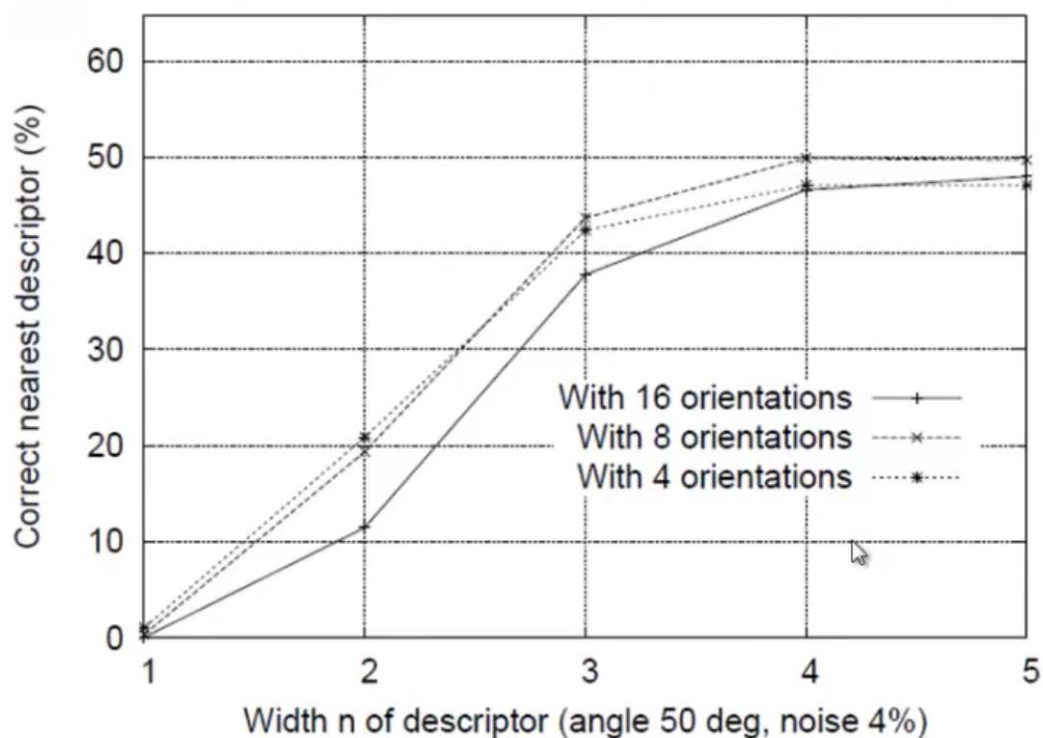
Επιπλέον, ανιχνεύεται η υψηλότερη κορυφή του ιστογράμματος και για την δημιουργία άλλου σημείου – κλειδιού με αυτό τον προσανατολισμό χρησιμοποιούνται και άλλες τοπικές κορυφές που βρίσκονται μέσα στο 80% της τιμής της μέγιστης κορυφής.

4. Περιγραφέας Σημείων- Κλειδιών

Στο τελευταίο αυτό στάδιο, θέλουμε χρησιμοποιώντας τα χαρακτηριστικά που βρήκαμε παραπάνω, να υπολογίσουμε ένα περιγραφέα για μια τοπική περιοχή γύρω από ένα keypoint, ώστε να την χαρακτηρίζει σε πολύ υψηλό βαθμό αλλά ταυτόχρονα, να είναι όσο το δυνατόν πιο αναλλοίωτη σε μεταβολές κλίμακας, προσανατολισμού, φωτισμού και θέσης του σημείου. Με τους περιγραφείς, μπορούμε να βρούμε ομοιότητες ανάμεσα στις τοπικές περιοχές δύο ή περισσότερων εικόνων.

Για κάθε σημείο – κλειδί, ορίζεται μια περιοχή κεντραρισμένη σε αυτό το σημείο και στη συνέχεια, υπολογίζεται το μέτρο και η διεύθυνση κλίσης κάθε σημείου αυτής της περιοχής, χρησιμοποιώντας διαφορές εικονοστοιχείων. Δημιουργείται, ακόμα, μια συνάρτηση βάρους τύπου Gauss με τυπική απόκλιση ίση με το ήμισυ της περιοχής για την εκχώρηση μιας τιμής βάρους που πολλαπλασιάζεται με το μέτρο της κλίσης κάθε στοιχείου. Ο σκοπός αυτής της συνάρτησης, είναι η ελάττωση των απότομων μεταβολών στον περιγραφέα, οι οποίες αντιστοιχούν σε μικρές μεταβολές στη θέση της συνάρτησης. Για την μείωση της επίδρασης μεγάλων μέτρων κλίσης, γίνεται κατωφλίωση των τιμών των χαρακτηριστικών, ώστε να όλες οι συνιστώσεις να έχουν τιμή μικρότερη από την πειραματικώς ορισμένη, 0.2. Μετά την κατωφλίωση, το διάνυσμα κανονικοποιείται στο μοναδιαίο μήκος.

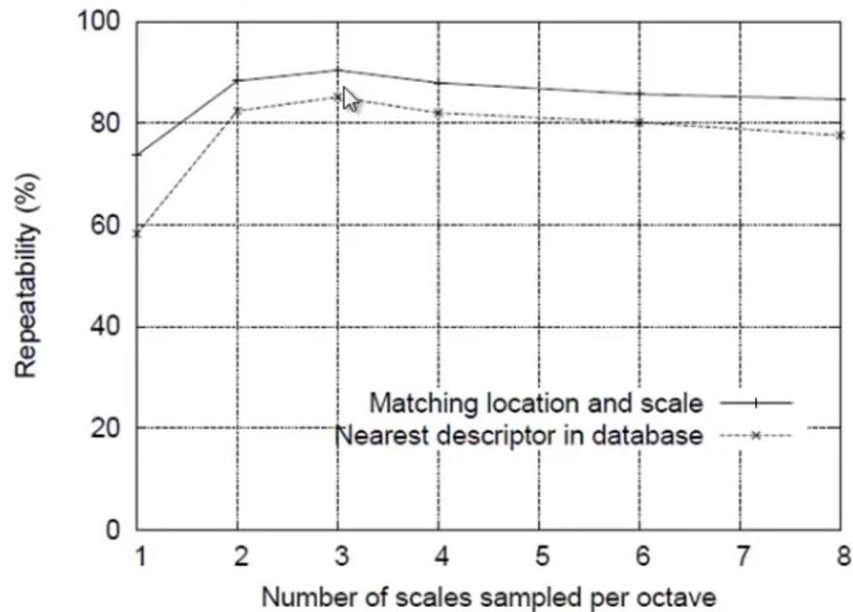
Έτσι, δημιουργείται ένα διάνυσμα χαρακτηριστικών 128 τιμών για το κάθε σημείο κλειδί. Στο κώδικα μας, βλέπουμε αυτό το διάνυσμα ως το μητρώο feature που έχει μέγεθος $4 \times 4 \times 8 = 128$, όπου 4×4 ο πίνακας διαστάσεων (διαστάσεις υπο-περιοχής) με 8 κατευθυντικές τιμές (8 bins κατευθύνσεων).



3. Πόσα επίπεδα χρησιμοποιούνται ανά οκτάβα στον αλγόριθμο

Όπως αναφέρεται στην αρχή του κώδικα(γραμμές 25-26), δημιουργούνται 3 οκτάβες και 3 επίπεδα ανά οκτάβα.

Πειραματικό αποτέλεσμα:



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 σημείο κλειδιού. Καταγράψτε τα πλεονεκτήματα και τα μειονεκτήματα κάθε μιας από τις προτεινόμενες τεχνικές.

Όταν πραγματοποιείται δειγματοληψία μιας συνεχούς συνάρτησης, το πραγματικό μέγιστο ή ελάχιστο σημείο της ενδέχεται να βρίσκεται μεταξύ των δειγμάτων που παίρνουμε. Για να πλησιάσουμε πιο κοντά στο πραγματικό ακρότατο επιθυμούμε να πετύχουμε ακρίβεια υποεικονοστοιχείου. Όπως είδαμε παραπάνω, ο αλγόριθμος SIFT χρησιμοποιεί για αυτό το σκοπό το γραμμικό και το δευτεροβάθμιο όρο του αναπτύγματος Taylor.

Επεκτείνοντας αυτήν την σκέψη, θα μπορούσαμε να αναπτύξουμε και πάλι σειρά Taylor, κρατώντας όμως και τον 3ο όρο, δηλαδή και τη δεύτερη παράγωγο. Με

μια επαναληπτική διαδικασία, υπολογίζουμε μέχρι την επιθυμητή μας ακρίβεια, με το αντίστοιχο κόστος λόγω χρόνου.

Ωστόσο, για την δημιουργία αναπτυσγμάτων Taylor, πρέπει να έχουμε γνώση και να μπορούμε να υπολογίσουμε την παράγωγο. Για να ξεφύγουμε από αυτό, μπορούμε να προσαρμόσουμε μία συνάρτηση παρεμβολής σε κάθε ακρότατο σημείο που βρίσκεται στην ψηφιακή συνάρτηση και στην συνέχεια, να αναζητήσουμε μια βελτιωμένη θέση ακροτάτου στην συνάρτηση παρεμβολής. Ο συγκεκριμένος τρόπος είναι σημαντικά γρηγορότερος από τους προηγούμενους, παρόλα αυτά χάνει σε ακρίβεια.

5. Για πέντε (5) εικόνες της επιλογής σας τρέξτε το SIFT_feature.m και καταγράψτε και σχολιάστε τα αποτελέσματά σας.

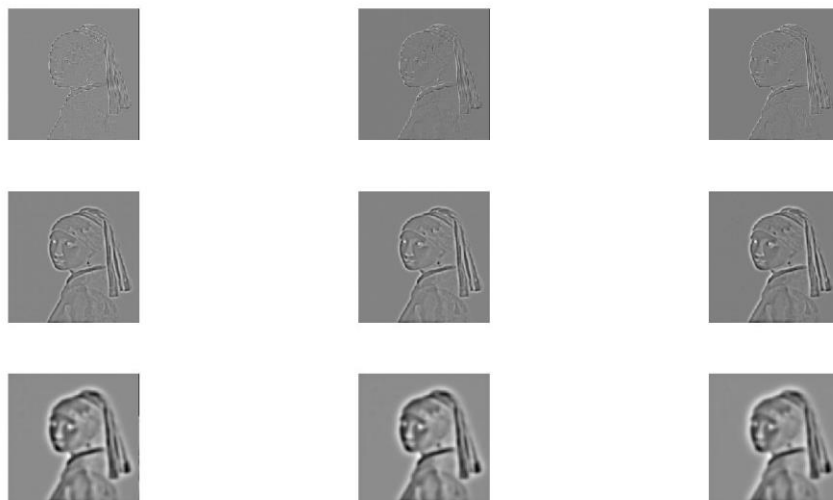
Οι 5 εικόνες που επιλέχθηκαν, παρουσιάζουν ποικιλία στην ποσότητα των ακμών, την αντίθεση και το μέγεθος των ομοιόμορφων περιοχών, για την καλύτερη σύγκριση των αποτελεσμάτων. Οι αρχικές εικόνες φαίνονται ακολούθως.



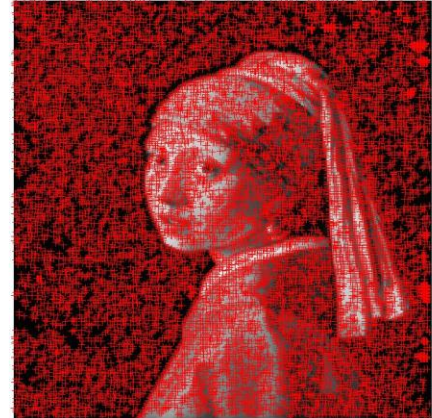


Όλες οι εικόνες μετατρέπονται από RGB σε grayscale, όπως απαιτείται από τον κώδικα που δίνεται.

Εκτελώντας τον αλγόριθμο, αρχικά παίρνουμε την DoG πυραμίδα, όπου βλέπουμε την εικόνα σε διαφορετικά επίπεδα και οκτάβες.



Στη συνέχεια, ο αλγόριθμος εντοπίζει όλα τα ακρότατα που θα μπορούσαν να χαρακτηριστούν ως σημεία – κλειδιά και τα σημειώνει πάνω στην εικόνα με κόκκινους σταυρούς.

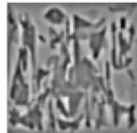
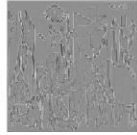


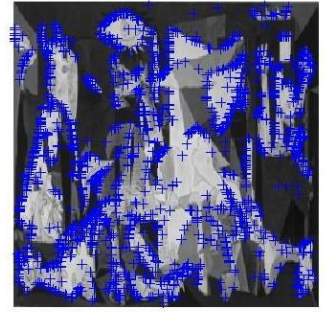
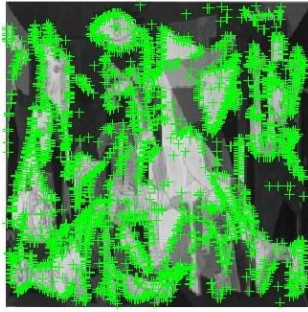
Αφού υπολογίστηκαν όλα τα πιθανά ακρότατα, ο αλγόριθμος προχωράει στην απόρριψη των σημείων που έχουν χαμηλή αντίθεση ή είναι κακώς εντοπισμένα σε κάποια άκρη, όπως μελετήσαμε και στην θεωρία μας. Έτσι, με πράσινο χρώμα βλέπουμε τα σημεία – κλειδιά που έμειναν μετά το φιλτράρισμα αυτών που η γειτονιά τους δεν εξασφάλιζε μεγάλη αντίθεση και έπειτα με μπλε, βλέπουμε τα τελικά keypoints που παρέμειναν μετά την αφαίρεση αυτών που ήταν κακώς εντοπισμένα. Οι αντίστοιχες εικόνες φαίνονται παρακάτω.



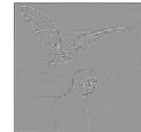
Στη συνέχεια, βλέπουμε τα ίδια αποτελέσματα για τις υπόλοιπες εικόνες.

Guernica



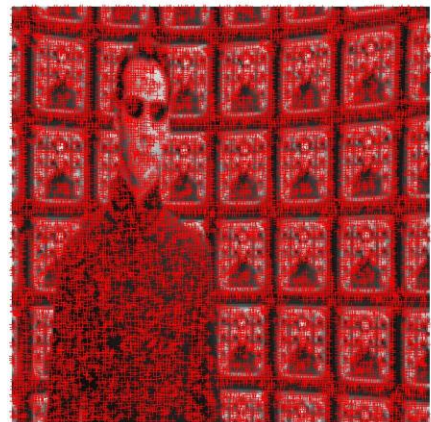
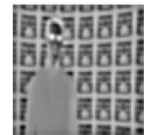
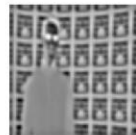
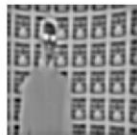
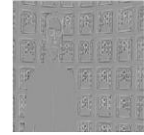
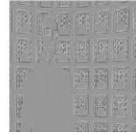
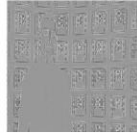


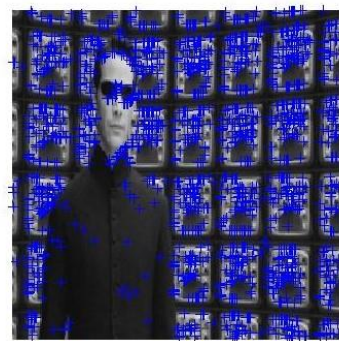
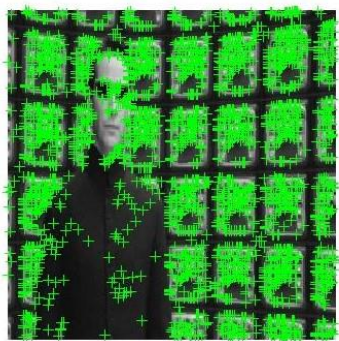
Psycho



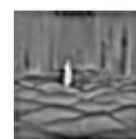
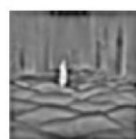
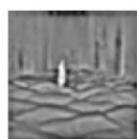
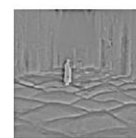
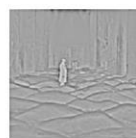
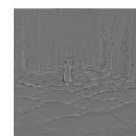
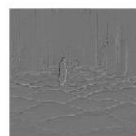
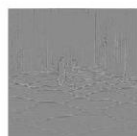


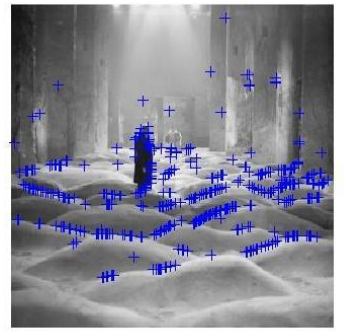
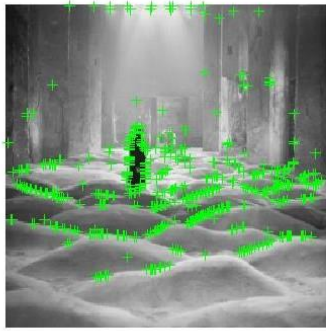
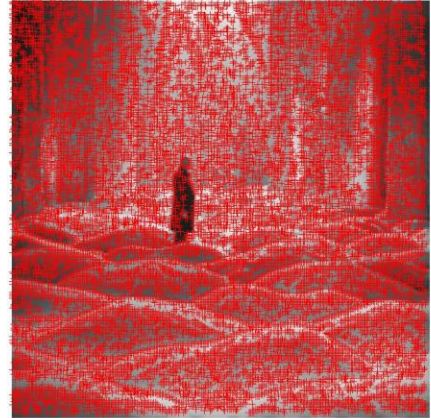
The Matrix





Stalker

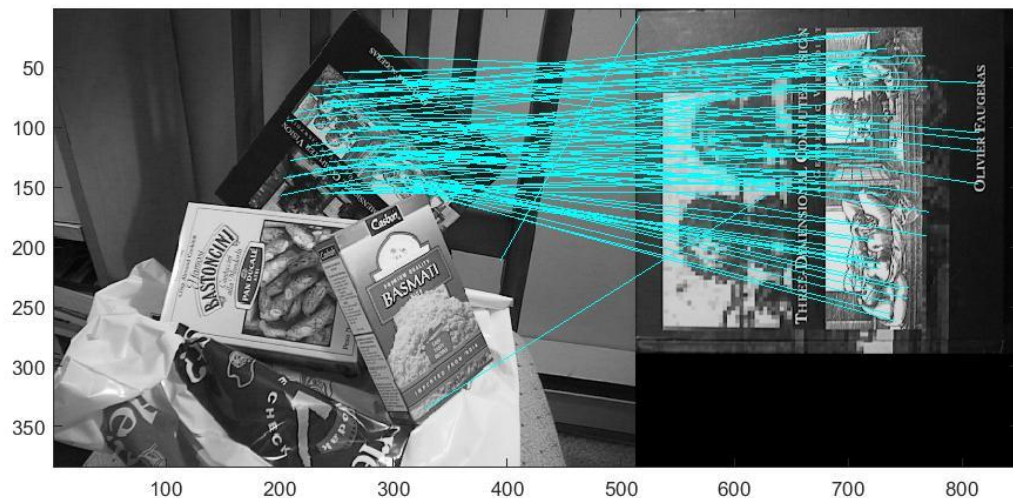




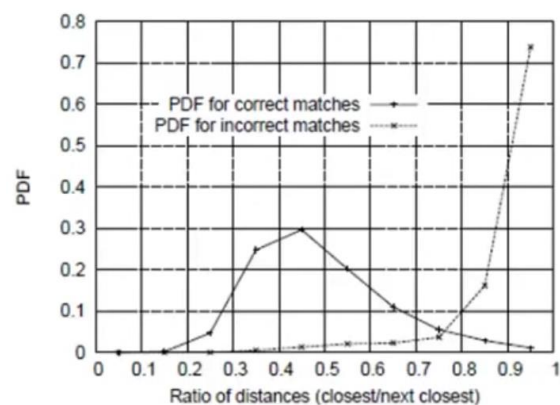
6. Υλοποιείτε αλγόριθμο αντιστοίχισης που θα βασίζεται στην αντιστοίχιση των περιγραφών των σημείων κλειδιών που προέκυψαν από την εφαρμογή του SIFT σε οποιοδήποτε ζευγάρι εικόνων της επιλογής σας.

Αφού βρεθούν τα SIFT features, υπολογίζονται τα matching points. Ένα ζευγάρι είναι αποδεκτό αν η απόσταση είναι μικρότερη από την απόσταση μεταξύ ενός κατωφλίου επί την απόσταση στο δεύτερο πιθανό ταίρι.

Το επιθυμητό αποτέλεσμα φαίνεται στην παρακάτω εικόνα.



Χαρακτηριστικά εξάγονται από καθεμία από τις δύο εικόνες και χαράσσονται γραμμές μεταξύ των χαρακτηριστικών που έχουν στενή αντιστοίχιση. Σε αυτό το παράδειγμα, βρέθηκαν πολλές αντιστοιχίσεις και μόνο ένα μικρό μέρος εξ αυτών είναι λανθασμένο.



7. Περιγράψτε τον αλγόριθμο RANSAC και καταγράψτε τον τρόπο με τον οποίο τον χρησιμοποιείται στον αλγόριθμο σας.

Ο Random Sample Consensus (RANSAC) αλγόριθμος προτάθηκε το 1981 από τους Fischler και Bolles. Αποτελεί μια επαναληπτική διαδικασία για την εκτίμηση παραμέτρων ενός μαθηματικού μοντέλου από ένα σύνολο δεδομένων, το οποίο περιέχει σημαντικό μέγεθος από outliers.

Σε ένα σύνολο δεδομένων θα υπάρχουν inliers, δεδομένα τα οποία ακολουθούν κάποια κατανομή που γίνεται να περιγραφεί από κάποιο μαθηματικό μοντέλο και outliers, δεδομένα τα οποία οφείλονται σε θόρυβο ή/και δεν ταιριάζουν στο μοντέλο, οπότε δεν πρέπει να επιδρούν στις εκτιμήσεις των παραμέτρων.

Σε αντίθεση με παρόμοιες τεχνικές, ο RANSAC χρησιμοποιεί για την παραγωγή του ζητούμενου μοντέλου τον ελάχιστο αριθμό παρατηρήσεων (σημεία δεδομένων). Το μειονέκτημά του, ωστόσο, είναι πως δεν αποτελεί ντετερμινιστικό αλγόριθμο. Παράγει ένα επιθυμητό αποτέλεσμα με μία ορισμένη πιθανότητα(e) που αυξάνεται με τον αριθμό των επαναλήψεων, χωρίς όμως να υπάρχει άνω όριο για αυτές.

Ο βασικός αλγόριθμος συνοψίζεται παρακάτω:

- a. Επιλέγεται τυχαία ο μικρότερος αριθμός(s) σημείων που απαιτείται για να καθοριστούν οι παράμετροι του μοντέλου.
- b. Λύνουμε ως προς τις παραμέτρους του μοντέλου.
- c. Καθορίζεται πόσα δεδομένα, από το σύνολο των δεδομένων ταιριάζουν στο υπολογισμένο μοντέλο, με μια προκαθορισμένη ανοχή ε.
- d. Εάν το κλάσμα του πλήθους των inliers προς το συνολικό μέγεθος των σημείων υπερβαίνει κάποιο ορισμένο κατώφλι τ, γίνεται επανεκτίμηση των παραμέτρων του μοντέλου χρησιμοποιώντας όλα τα σημεία που έχουν θεωρηθεί inliers και ο αλγόριθμος τερματίζει.
- e. Αλλιώς, επαναλαμβάνουμε τα βήματα 1 έως 4 (για μέγιστο αριθμό Τ επαναλήψεων).

$$T = \frac{\log(1 - p)}{\log(1 - (1 - e)^s)}$$

Η χρήση του RANSAC στο πρόβλημα της αντιστοίχισης, είναι ότι μέσω της επαναληπτικής του διαδικασίας θα απομακρύνει τα σημεία που δεν μπορούν να χρησιμοποιηθούν για ταίριασμα ή συμμετέχουν σε λάθος αντιστοιχήσεις (outliers) και επομένως, θα συμβάλει στην ευρωστία του αλγορίθμου.

8. Καταγράψτε τις βασικές διαφορές μεταξύ των αλγορίθμων SIFT, GLOH (Gradient Location and Orientation Histogram) και SURF (Speeded Up Robust Features).

SIFT:

Για την καλύτερη σύγκριση μεταξύ των αλγορίθμων, θα αναφέρουμε εν συντομία τα βασικά βήματα του SIFT, όπως τα έχουμε δει μέχρι τώρα.

Αρχικά, κατασκευάζεται ο χώρος κλίμακας. Οι παράμετροι που πρέπει να καθοριστούν είναι οι σ (τυπική απόκλιση), s (από το s υπολογίζεται ο συντελεστής k) και ο αριθμός των οκτάβων. Από τις εξομαλυμένες εικόνες στον χώρο κλίμακας, υπολογίζονται οι διαφορές των συναρτήσεων Gauss, $D(x, y, \sigma)$.

Έτσι, βρίσκουμε τα ακρότατα σε κάθε εικόνα D , τα οποία αποτελούν και τα αρχικά σημεία – κλειδιά. Χρησιμοποιώντας το ανάπτυγμα Taylor στις τιμές της D , βελτιώνουμε την ακρίβεια της θέσης των σημείων - κλειδιών.

Έπειτα, διαγράφουμε τα σημεία κλειδιά που θεωρούνται ακατάλληλα, καθώς χαρακτηρίζονται από χαμηλή αντίθεση ή δεν έχουν εντοπιστεί με μεγάλη ακρίβεια. Εξαλείφονται, επίσης, τα σημεία – κλειδιά που σχετίζονται με ακμές. Με χρήση του ιστογράμματος, υπολογίζουμε το μέτρο και τον προσανατολισμό του κάθε keypoint και τέλος, παράγεται ο ζητούμενος περιγραφέας, δηλαδή ένα διάνυσμα χαρακτηριστικών για το κάθε σημείο – κλειδί.

GLOH:

Ο Gradient Location and Orientation Histogram αλγόριθμος προτάθηκε από τους Mikolajczyk and Schmid (2005). Θεωρείται ότι έχει στενή σχέση με τον SIFT, καθώς και αυτός κατασκευάζει ιστόγραμμα κλίσης των προσανατολισμών σε μία γειτονιά γύρω από το σημείο ενδιαφέροντος. Σε αντίθεση με τον SIFT, βέβαια, ο περιγραφέας GLOH υπολογίζεται σε ένα λογαριθμικό – πολικό πλέγμα (log polar grid) αντί του ορθογώνιου, χρησιμοποιεί 16 bins, αντί για 8, ώστε να κβαντίσει τις κλίσεις των κατευθύνσεων και κάνει χρήση της τεχνικής PCA (ανάλυση κύριας συνιστώσας) για την μείωση της διαστατικότητας του περιγραφέα.

SURF:

Ο Speeded Up Robust Features αλγόριθμος προτάθηκε από τους Bay, Tuytelaars και Van Gool(2006, 2008). Βασίζεται και αυτός μερικώς στον SIFT, καθώς αποτελεί επίσης διάνυσμα χαρακτηριστικών που προέρχεται από τη γειτονιά του σημείου ενδιαφέροντος και θεωρείται τρεις φορές πιο γρήγορος από τον SIFT. Η υλοποίηση του SURF διαφέρει στα ακόλουθα σημεία:

- i. αντί του gaussian πυρήνα, για να πετύχει την εξομάλυνση των εικόνων (υπολογισμός DoG), χρησιμοποιεί box filters τα οποία είναι σημαντικά γρηγορότερα με τη βοήθεια integral εικόνων.
- ii. για να ανιχνεύσει σημεία ενδιαφέροντος, χρησιμοποιεί blob detector που βασίζεται στον Hessian πίνακα.
- iii. βασίζεται σε Haar wavelets αντί για προσεγγίσεις παραγώγων σε μια πυραμίδα εικόνας, τόσο για τους προσανατολισμούς όσο και για τον περιγραφέα.

9. Ποιο είναι το υπολογιστικό κόστος των παραπάνω αλγορίθμων·

Σύμφωνα με το paper “Analyzing and Exploring Feature Detectors in Images”, που φαίνεται και στην βιβλιογραφία παρακάτω, η θεωρητική και πειραματικά επαληθεύσιμη πολυπλοκότητα των αλγορίθμων SIFT και SURF είναι $O(mn + k)$, όπου m , n οι διαστάσεις της εικόνας και k το πλήθος των ακροτάτων που βρέθηκαν. Αν και η πολυπλοκότητα είναι ίδια, υπάρχουν σταθερές που λόγω της ασυμπτωτικής γραφής παραλείπονται, ωστόσο καθιστούν τον SIFT πιο αργό από τον SURF.

Αναμένουμε ότι η πολυπλοκότητα του GLOH, σε ασυμπτωτική αναπαράσταση, θα είναι ίδια με παραπάνω, όμως λόγω της PCA, θεωρούμε ότι ο τελευταίος θα είναι πιο αργός από τον SIFT.

Συνοπτικά, μπορούμε να κατατάξουμε τους 3 αλγορίθμους κατά φθίνον υπολογιστικό κόστος ως εξής: GLOH, SIFT, SURF.

10. Καταγράψτε τις βιβλιογραφικές πηγές που χρησιμοποιήσατε.

- ✚ “Distinctive Image Features from Scale Invariant Keypoints”, International Journal of Computer Vision, pp. 91110, 2004, David G. Lowe.
- ✚ “Digital Image Processing”, 4th Edition, Rafael C. Gonzalez, Richard E. Woods.
- ✚ “Scale Invariant Feature Transform”, Scholarpedia, Prof. Tony Lindeberg.
- ✚ “Random Sample Consensus: A paradigm for model Fitting with Applications to Image Analysis and Automated Cartography”, Martin A. Fischler, Robert C. Bolles.
- ✚ “Overview of the RANSAC algorithm”, Konstantinos G. Derpanis.
- ✚ “Analyzing and Exploring Feature Detectors in Images”, Paulo Drews Jr, Rodrigo de Bem, Alexandre de Melo.
- ✚ “A performance evaluation of local descriptors”, Krystian Mikolajczyk, Cordelia Schmid.