Scale & Affine Invariant Interest Point Detectors

KRYSTIAN MIKOLAJCZYK AND CORDELIA SCHMID [2004]

Shreyas Saxena
Gurkirit Singh

23/11/2012

Introduction

- We are interested in finding interest points.
- What is an interest point?
- Why is invariance required?

  -Scale

  -Rotation

  -Affine

Why a new approach is required for Detecting Interest points?

- Classical Approach
- Flaws-
- Detection and matching are resolution dependent.

Figure 1. An example of matching a low-resolution image with a high-resolution one.

Intuitive Idea to solve the problem of Scale Variation

- Extract the information at different scales!
- Issues-
- Space for representation
- Mismatches due to a large feature space.

<u>Way about this problem-</u>

Extraction on feature points at a characteristic scale.

Example

Scale Invariant Detectors (1/2)

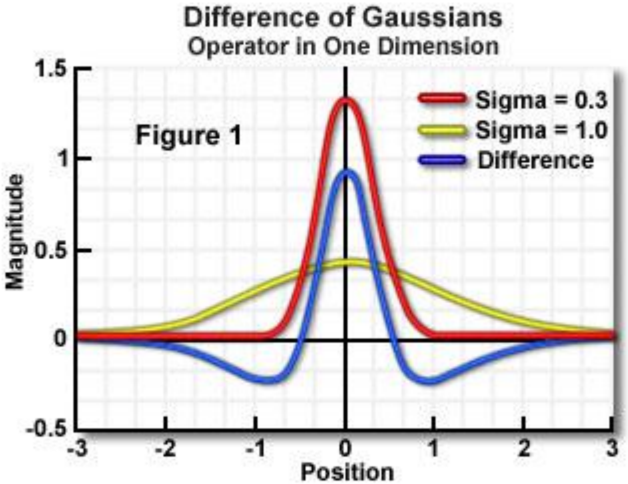
- Assumption- Scale Change is Isotropic.
- Robust to minor affine transformations
- Introduced in 1981 by Crowley-
  - Pyramid Construction
  - Difference of Gaussians

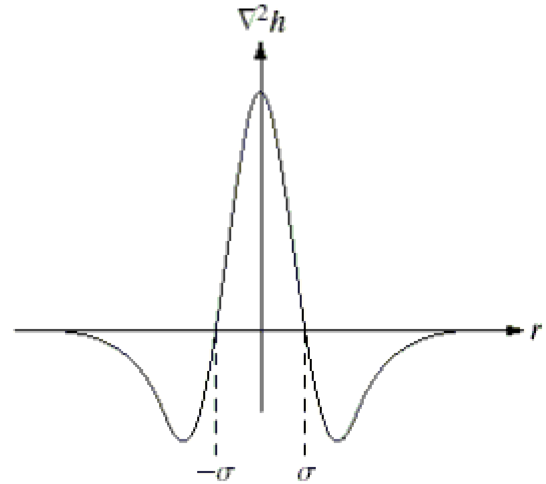
  3D Extremum as a feature point, if it more that a specific threshold.

Scale Invariant Detectors (2/2)

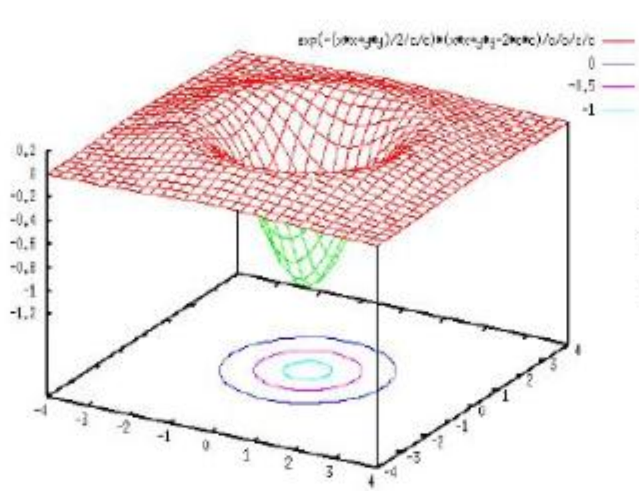
- Other works-
- Lindberg 1998 uses LoG to form the pyramids. Later, automatic scale selection is also proposed.
- Lowe 1999 Scale Space pyramid based with Difference of Gaussian. (Why?)

DoG vs LoG

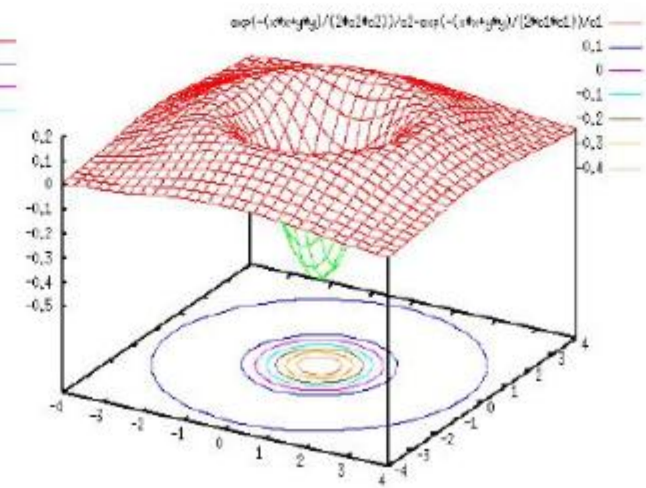




Laplacian of Gaussian



Difference of Gaussians



Drawbacks

- Detection of maxima even at places where, the signal change is present in one direction.
- Also, they are not stable to noise.

Way about-

- We penalize the feature points, having variation only in one direction.
- Also, use of second order derivative insures a maxima for a localized neighborhood.

Scale Adapted Harris Detector

- Second moment matrix- scale adapted

$$\mu(\mathbf{x}, \sigma_I, \sigma_D) = \begin{bmatrix} \mu_{11} & \mu_{12} \\ \mu_{21} & \mu_{22} \end{bmatrix} = \sigma_D^2\, g(\sigma_I) * \begin{bmatrix} L_x^2(\mathbf{x}, \sigma_D) & L_x L_y(\mathbf{x}, \sigma_D) \\ L_x L_y(\mathbf{x}, \sigma_D) & L_y^2(\mathbf{x}, \sigma_D) \end{bmatrix}$$

- Here,
  - Sigma D is the differentiation scale
  - Sigma I is the integration scale
  - Lx and Ly are the first order derivatives in X and Y
  - Differentiation Scale= 0.7 * Integration Scale

Finding Corners

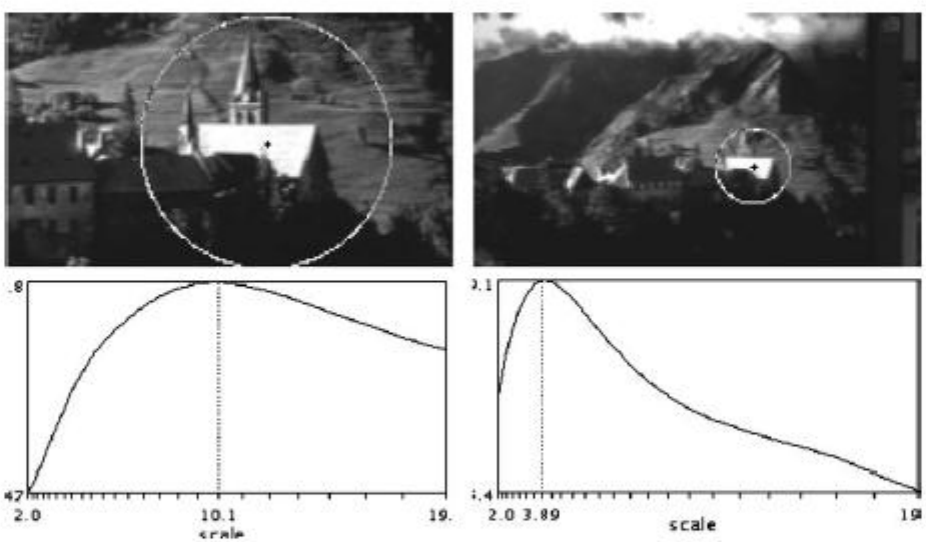
- Our interest points are the one where both Eigen values are significant.

$$cornerness = \det(\mu(\mathbf{x}, \sigma_{\mathbf{I}}, \sigma_{\mathbf{D}})) - \alpha \text{trace}^2(\mu(\mathbf{x}, \sigma_{\mathbf{I}}, \sigma_{\mathbf{D}}))$$

- If λ_1 and λ_2 are the two Eigen values of a matrix then the above expression becomes-
- $\lambda_1 \lambda_2 - \alpha(\lambda_1 + \lambda_2)^2$
  - α is generally 0.07
  - We select points, for which response is greater than threshold.

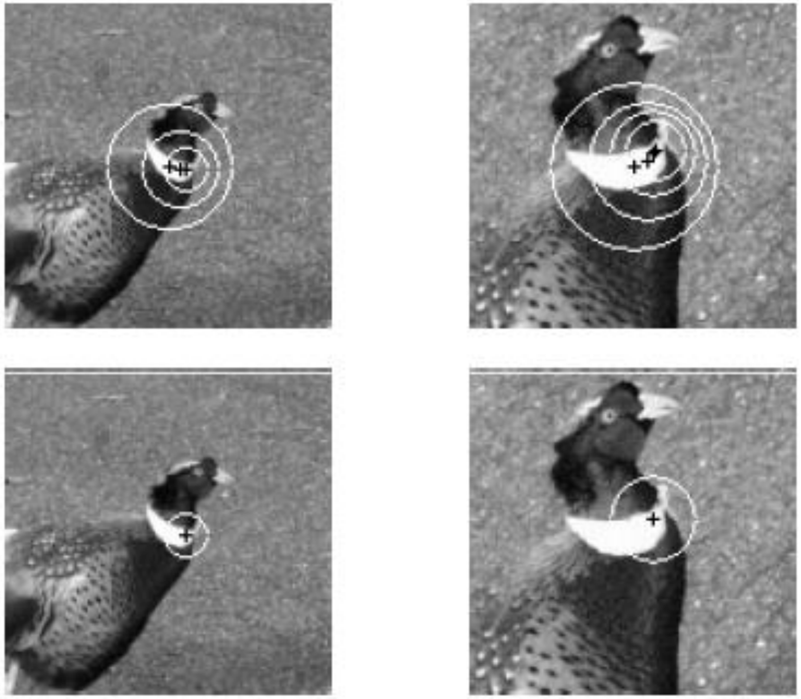
Characteristic Scale

- As explained, we are interested to extract the feature points only for a range of scales.
- We evaluate the number of features, found at each scale.
- Harris measure does not validate as a good benchmark, and LoG performance is much better.



$$|\mathrm{LoG}(\mathbf{x}, \sigma_n)| = \sigma_n^2 |L_{xx}(\mathbf{x}, \sigma_n) + L_{yy}(\mathbf{x}, \sigma_n)|$$

Example



Harris Laplace Detector

- Construct the scale space at different scales. (Scale Factor is 1.4)
- Detect Harris points, with a threshold for the minimum value.
- Once, points are found for each of them we scan the neighboring scales for a extrema of LoG. (Scale Selection)

$$\sigma_I^{(k+1)} = t\sigma_I^{(k)} \quad \text{with } t \in [0.7, \ldots, 1.4]$$

- After, this we take maxima of Harris measure at that scale, and update our point.
- Why do we scan again for different scales?

Simplified Harris Laplace Detector

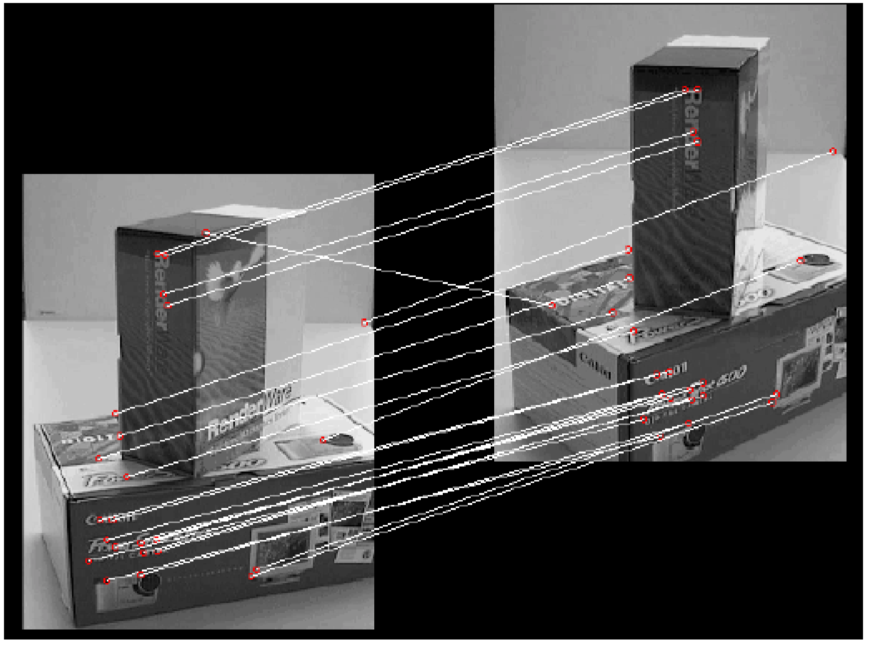
(Mikolajczyk and Schmid,2001)

- At each scale, find Harris points having a maxima.
- On each point, we use LoG measure to see if it is a local maxima greater than a threshold.
- The ratio between scales is 1.2

This method is a tradeoff for speed versus accuracy, whereas the previous approach takes time but gives a more accurate location and scale.

Problem?

- Change in Perspective causes more problems than scale and rotation.
- Scale Change is not isotropic.



Affine Variation

- Perspective transformation can be modeled as an affine variation up to a certain extent, for a planar region.
- The detection scale should vary independently in orthogonal directions in order to deal with affine scaling.

Basic Theory

- The second order moment is given by-

$$\mu(\mathbf{x}, \Sigma_I, \Sigma_D) = \det(\Sigma_D)\, g(\Sigma_I) * ((\nabla L)(\mathbf{x}, \Sigma_D)(\nabla L)(\mathbf{x}, \Sigma_D)^T)$$

- The affine relation- $\mathbf{x}_L = A\mathbf{x}_R$,
- This should change the other kernels of Integration and differentiation by same

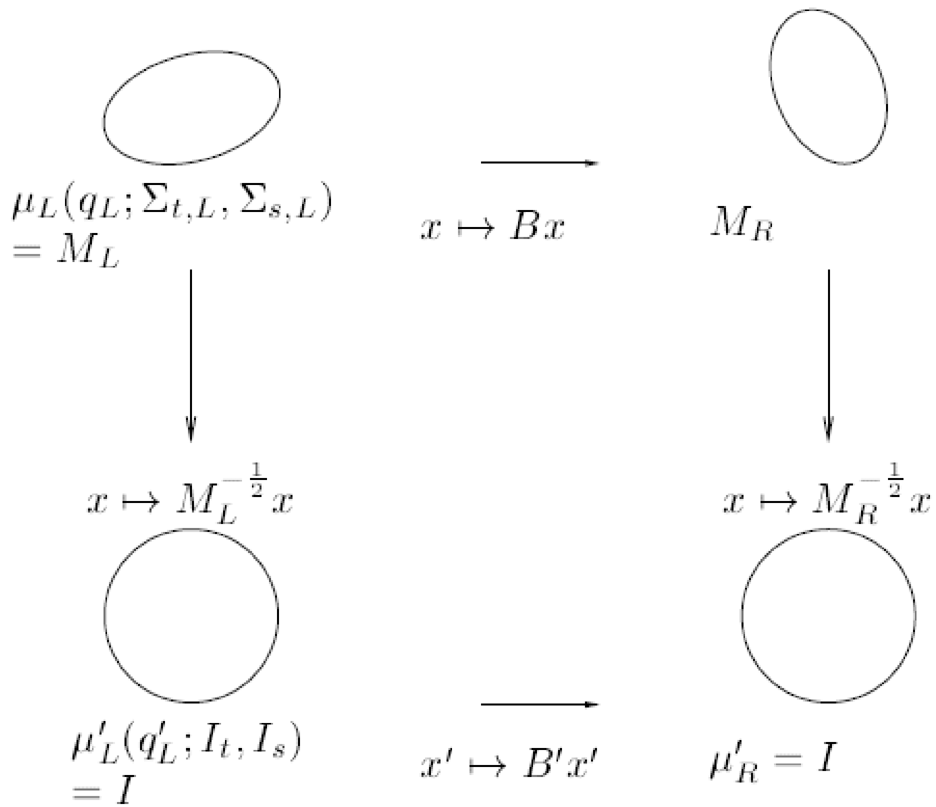
$$\mu(\mathbf{x}_L, \Sigma_{I,L}, \Sigma_{D,L}) = A^T \mu(\mathbf{x}_R, \Sigma_{I,R}, \Sigma_{D,R}) A = A^T \mu(A\mathbf{x}_L, A\Sigma_{I,L}A^T, A\Sigma_{D,L}A^T) A$$

$$\mu(\mathbf{x}_L, \Sigma_{I,L}, \Sigma_{D,L}) = M_L \quad \mu(\mathbf{x}_R, \Sigma_{I,R}, \Sigma_{D,R}) = M_R$$

$$M_L = A^T M_R A$$

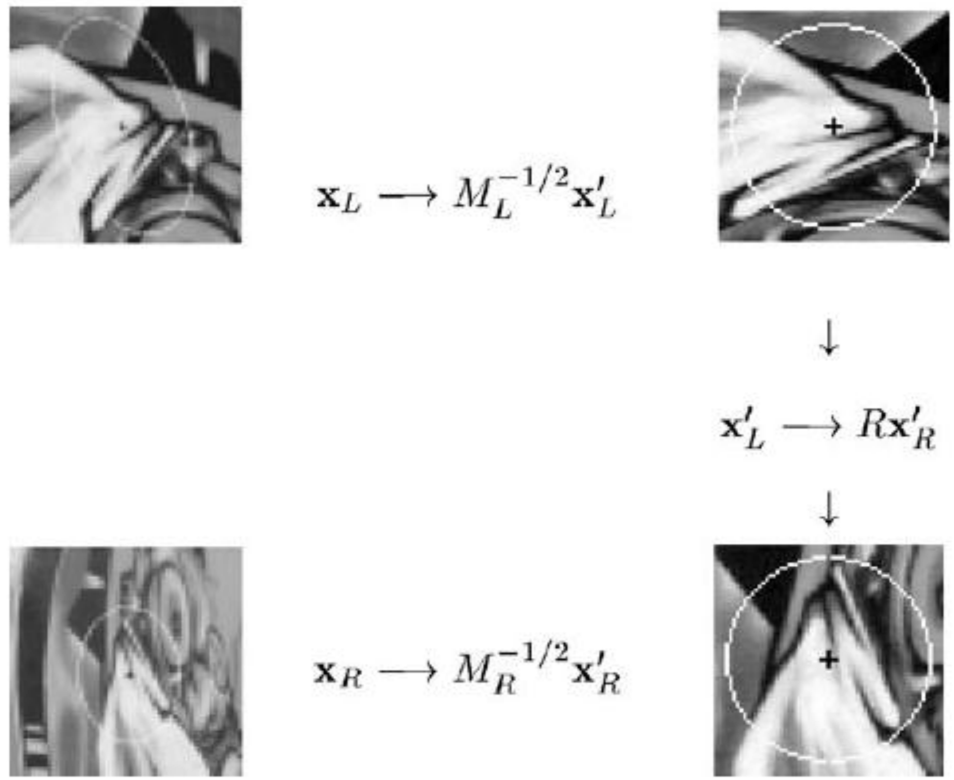
What is really happening-

- We want to normalize the neighborhood of a point.



[Baumberg 2000]

Another example



Eigen Vectors

- What does this mean in terms of Eigen Vectors?
- The Eigen Vector, having the smallest value in A, gets the highest Eigen value in A inverse.
- In a way we stretch the image patch in the direction with less variance.
- Final measure, is ratio of Eigen values which in perfect case should approach 1.

How do we go about it?
Harris Affine Interest Point Detector

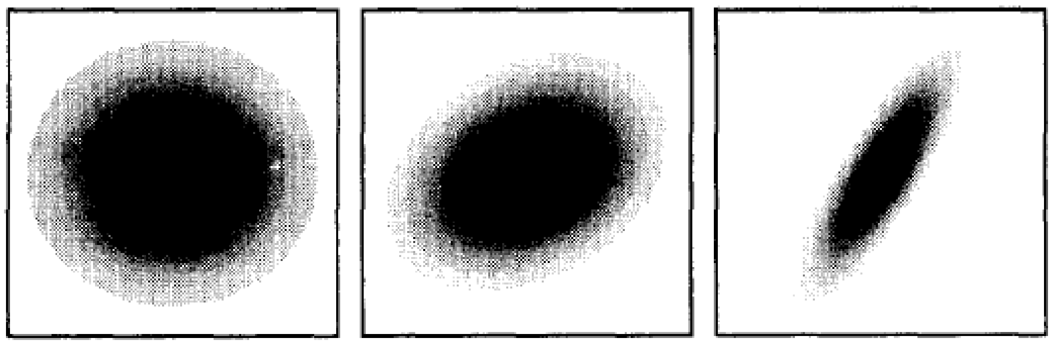
- Spatial Location- Determined by the Harris Detector
- Integration Scale- Maxima of LoG, taken same from above
- Shape Adapted matrix- Computed from the second moments, to normalize the neighborhood.
- Differentiation Scale- is initially taken from the integration scale, but is then varied to get a maxima for Isotropy.

- What is Isotropy here?

$$Q = \frac{\lambda_{\min}(\mu)}{\lambda_{\max}(\mu)}$$

Shape Adapted Matrix

- Initially Lindberg had proposed used of affine Gaussian kernels. [Lindberg 1997]



- But, it is better to compute the affine on image patch so, that we can recursively apply the same Gaussian.

$$U = \prod_k \left(\mu^{-\frac{1}{2}}\right)^{(k)} U^{(0)}$$

- One thing is ensured,

$$\lambda_{\max}(U) = 1.$$

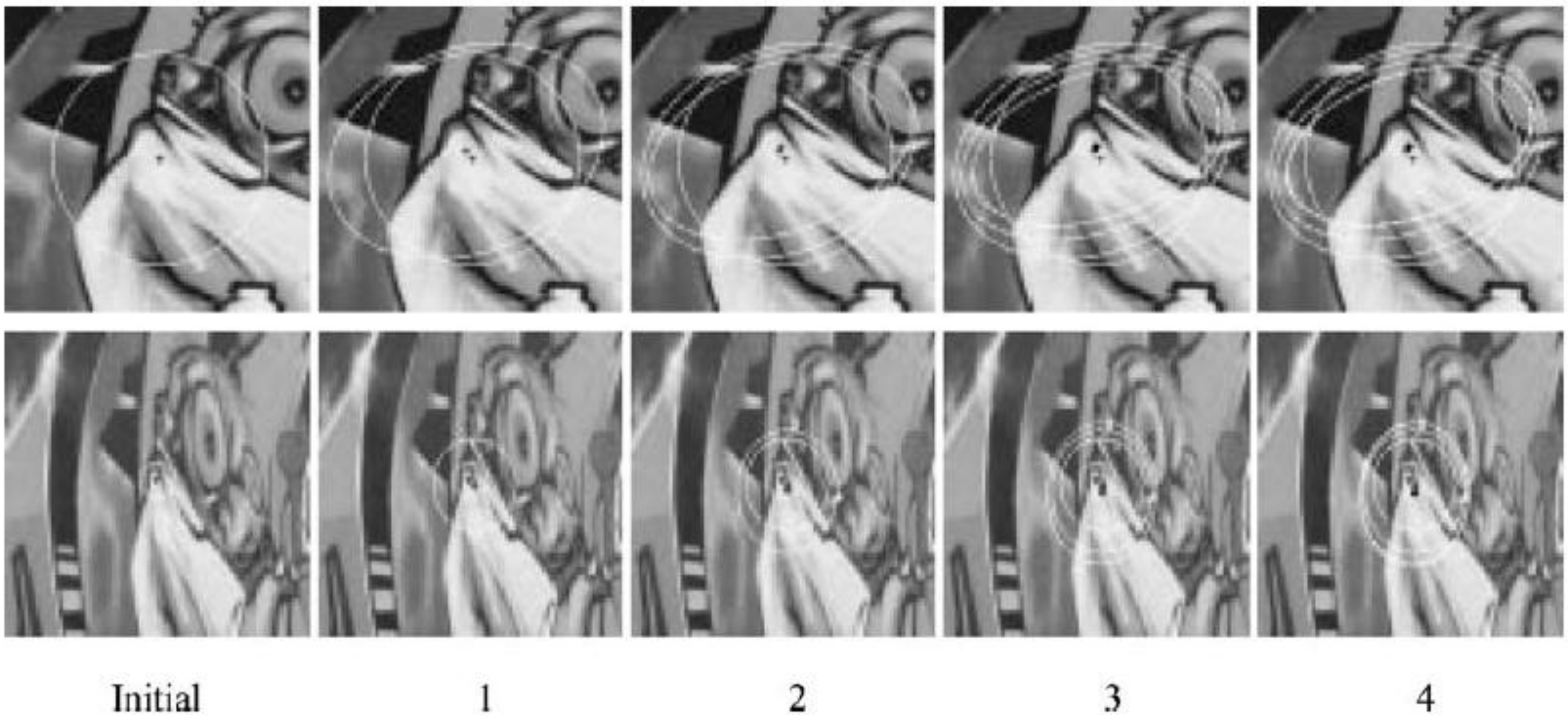
Integration Scale

- The starting value is chosen from the Harris Laplace detector.
- Strong affine transformations, it is essential to select the integration scale after each estimation of the U transformation.
- This allows to converge towards a solution where the scale and the second moment matrix do not change any more.

Differentiation Scale

- Diff. Scale < Inte. Scale
- Should be in an optimum Range,
  - If too less, then smoothing dominates
  - Should be less enough such that, integrating kernel smoothens out the noise without suppressing information.
- Its value is varied, in order to get a higher isotropy measure Q.
- Scales help to converge faster in case Eigen values of selected points are not similar.
- We can have Diff. Scale = Constant * Inte. Scale; not always efficient.

Example



Convergence Criterion

- Either we can see if the matrix U^k is almost a rotation matrix; or we can say both the Eigen Values are same.
- Generally we allow $\frac{\lambda_{\max}(D)}{\lambda_{\min}(D)} > \epsilon_l$ n of error,

$$1 - \frac{\lambda_{\min}(\mu)}{\lambda_{\max}(\mu)} < \epsilon_C$$

$$\epsilon_C = 0.05$$

- Termination Criterion, in case of a step edge-
  - If $\frac{\lambda_{\max}(D)}{\lambda_{\min}(D)} > 6$

Iterative Detection Algorithm

Step 1. Initialization

- Initialization is done with multi scale Harris detector of point.
- Scale space, Integration scale σ_I ,

  Differentiation scale σ_D,
- initial points X(0)
- Shape Adaption Matrix U(0) as Identity matrix

[1]

Algorithm continue

Step 2. Normalize the window

$$U^{(k-1)}\mathbf{x}_w^{(k-1)} = \mathbf{x}^{(k-1)}$$

Step 3. Integration scale selection

scale that maximizes LOG

$$\sigma_I^{(k)} = \underset{\substack{\sigma_I = t\sigma_I^{(k-1)} \\ t\in[0.7,\ldots,1.4]}}{\operatorname{argmax}} \sigma_I^2 \det(L_{xx}(\mathbf{x},\sigma_I) + L_{yy}(\mathbf{x},\sigma_I))$$

[1] and [2]

Algorithm continue

Step 4. Integration scale selection

Scale that maximize the isotropic measure.

$$\sigma_D^{(k)} = \underset{\sigma_D = s\sigma_I^{(k)},\ s \in [0.5,\ldots,0.75]}{\operatorname{argmax}} \frac{\lambda_{\min}(\mu(\mathbf{x}_w^{(k)}, \sigma_I^k, \sigma_D))}{\lambda_{\max}(\mu(\mathbf{x}_w^{(k)}, \sigma_I^k, \sigma_D))}$$

This maximization process will try to converge the eigenvalues of second-momtent matrix to same value

[1] and [2]

Algorithm continue

Step 5. Spatial Localization

- Maximizes the Harris corner measure (Cornerness) within the 8 neighborhood of previous point.

$$\mathbf{x}_w^{(k)} = \underset{\mathbf{x}_w \in W(\mathbf{x}_w^{(k-1)})}{\operatorname{argmax}} \det(\mu(\mathbf{x}_w, \sigma_I^k, \sigma_D^{(k)})) - \alpha \operatorname{trace}^2(\mu(\mathbf{x}_w, \sigma_I^k, \sigma_D^{(k)}))$$

- Then New point should transformed to U normalized frame. Localization is done in that.

$$\mathbf{x}^{(k)} = \mathbf{x}^{(k-1)} + U^{(k-1)} \cdot (\mathbf{x}_w^{(k)} - \mathbf{x}_w^{(k-1)})$$

[1] and [2]

Algorithm continue

Step 6. Updating

- Square root of second moment matrix define the reference frame. So,

$$\mu_i^{(k)} = \mu^{-\frac{1}{2}}(\mathbf{x}_w^{(k)}, \sigma_I^{(k)}, \sigma_D^{(k)})$$

- Henceforth , Transformation or shape adaption matrix

$$U = \prod_k \mu_i^{(k)} \cdot U^{(0)} = \prod_k (\mu^{-\frac{1}{2}})^{(k)} \cdot U^{(0)}$$

- Fix the maximum eigenvalue to 1 (to ensure the expansion in least change direction)

[1] and [2]

Algorithm continue

Step 7 Stopping criterion

- Algorithm solve for anisotropic region and try to converge to isotropic region by U matrix.
- When close enough to isotopic shape then stop the iterative algorithm, So,
- Stop when $$1 - \frac{\lambda_{\min(\mu_i^{(k)})}}{\lambda_{\max(\mu_i^{(k)})}} < \varepsilon_C$$
- Where ε_c is 0.05

[1] and [2]

Results

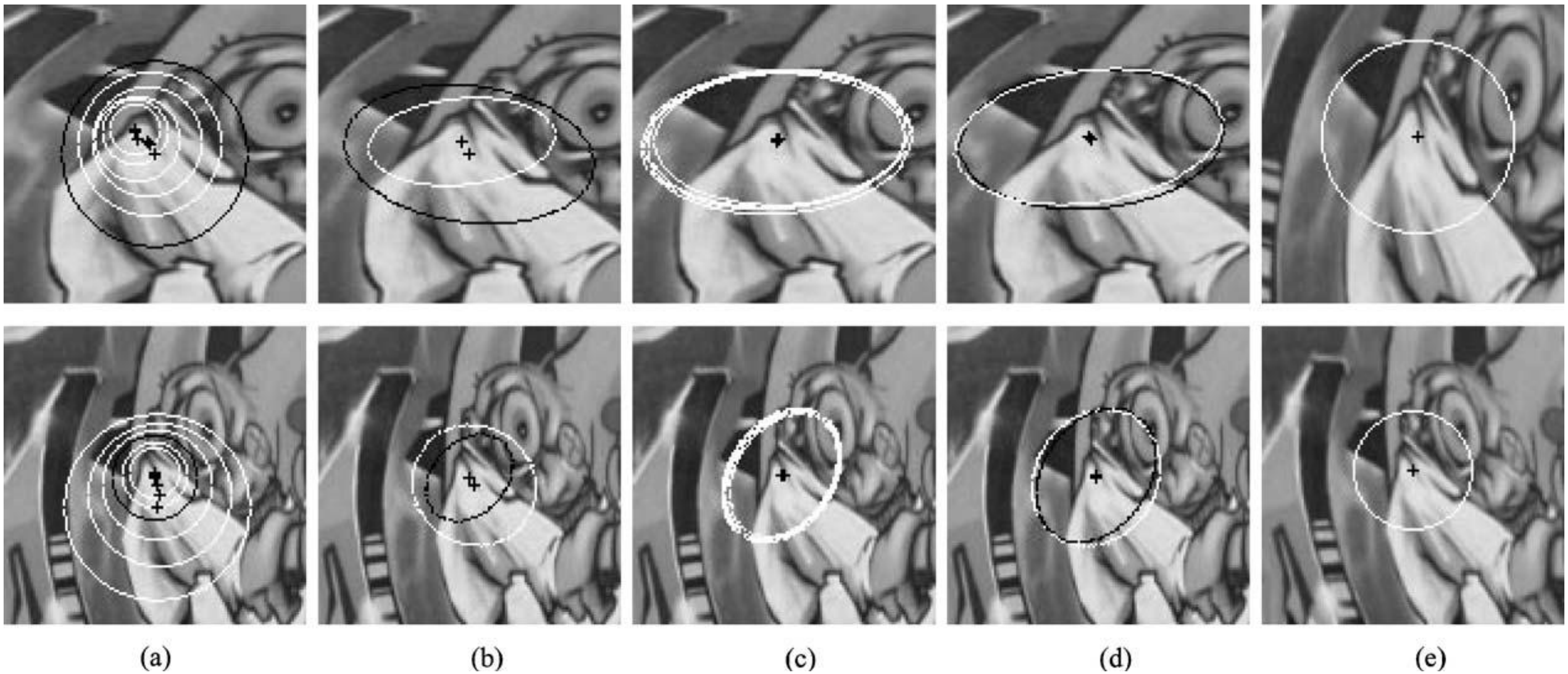


Image taken from [1]

Evaluation of Interest point detector

1. Number of corresponding point detected in images under different geometric transformations.
2. Localization and region overlapping accuracy

[1]

Data Set

- Scale change 1.4 to 4.5
- View point change up to 70 degree.
- 160 Images, 10,000 interest point

[1]

Repeatability Criterion

- [Repeatability %] Ratio between number of point to point correspondences and minimum number of point detect in images.

$$R_{\text{score}} = \frac{C(A,B)}{\min(n_A, n_B)}$$

- Point detected in both images.
- [Localization error] Xa and Xb point correspondences and related by image homography H if error: |Xa-H.Xb| is less than 1.5.

[1] and [2]

Scale overlap Error

- Scale invariant   points the surface error ε_s is:

$$\epsilon_S = \left| 1 - s^2 \frac{\min(\sigma_a^2, \sigma_b^2)}{\max(\sigma_a^2, \sigma_b^2)} \right|$$

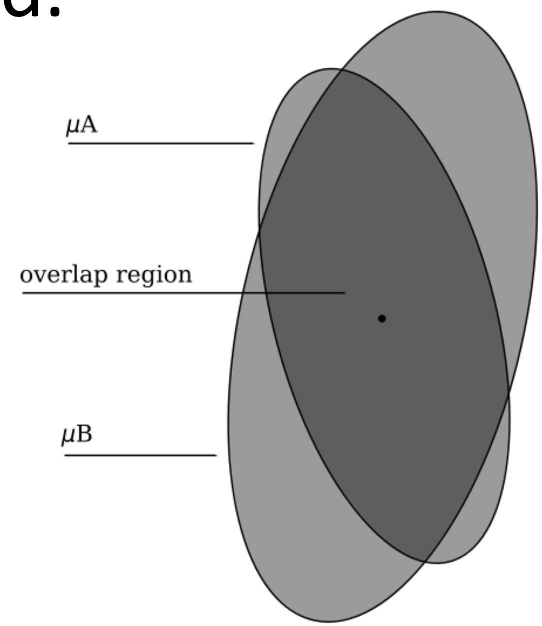
- where σa and σb are the selected point scales and s is the actual scale factor recovered from the homography between the images ($s > 1$).
- $\varepsilon s < 0.4$.

[1] and [2]

Affine Overlap Error

- Surface error ε_s of two affine points must be less than a specified threshold.

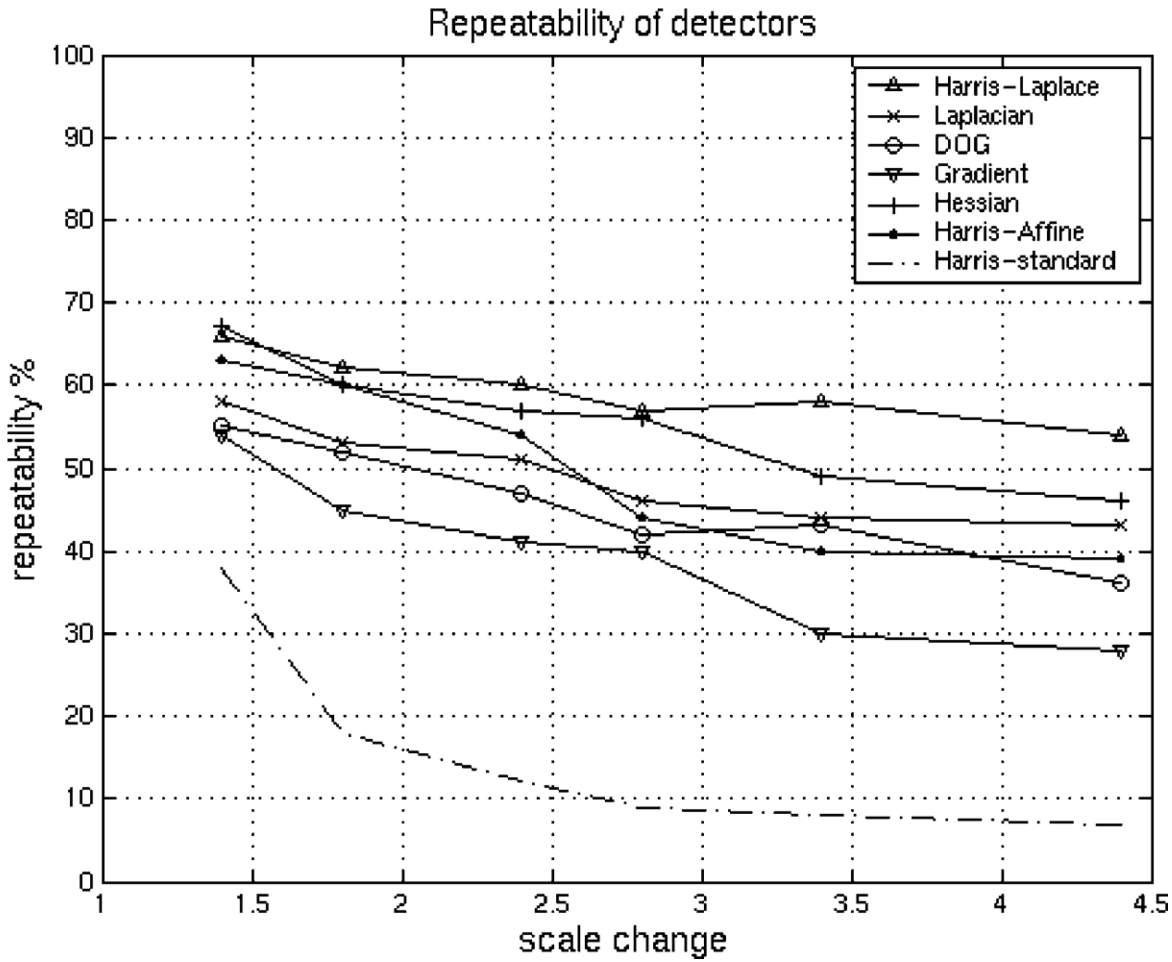
$$\epsilon_S = 1 - \frac{\mu_a \cap (A^T \mu_b A)}{(\mu_a \cup A^T \mu_b A)} \varepsilon_s$$



Where μa and μb are the elliptic regions defined by $X^T \mu X = 1$.

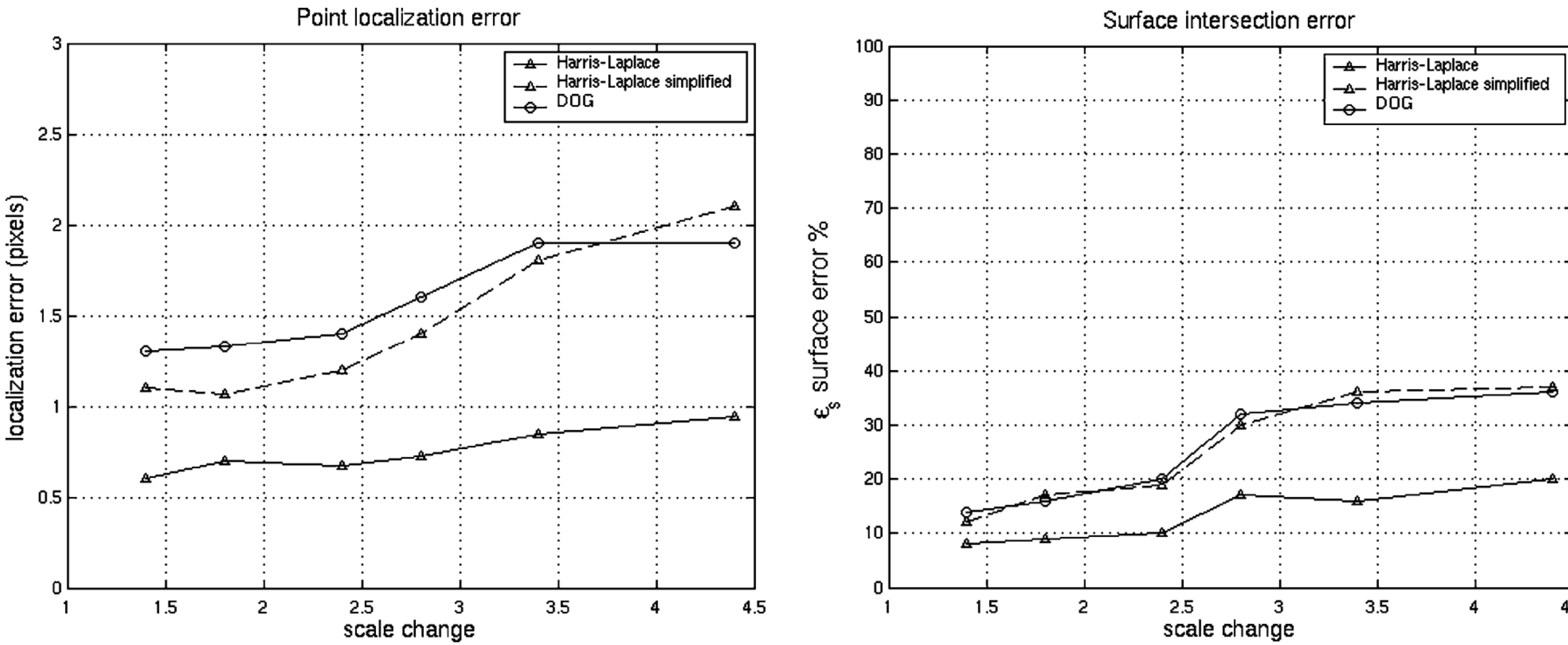
[1] and image taken from [2]

Repeatability % with scale change



Graph taken from [1]

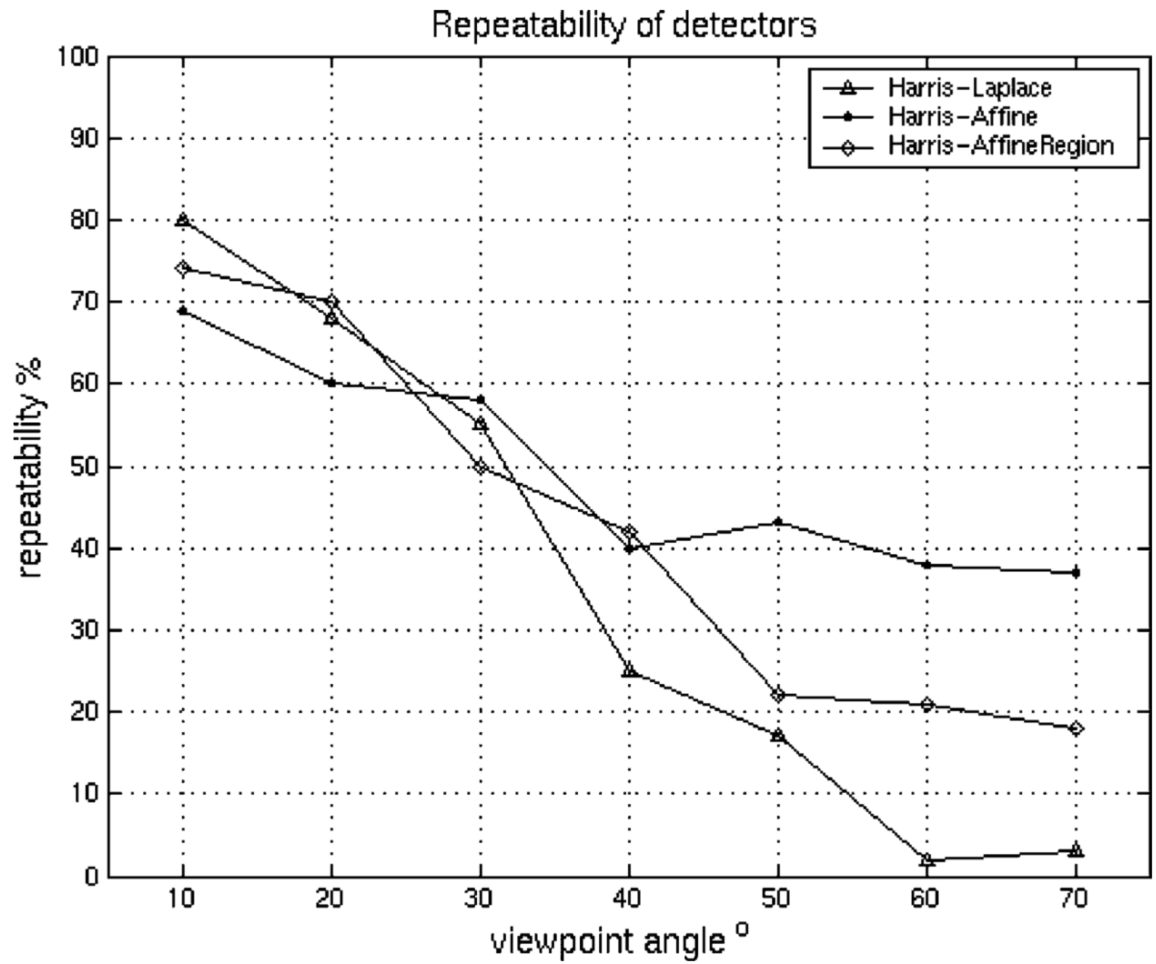
Localization and surface overlap Error with scale change



(a) (b)

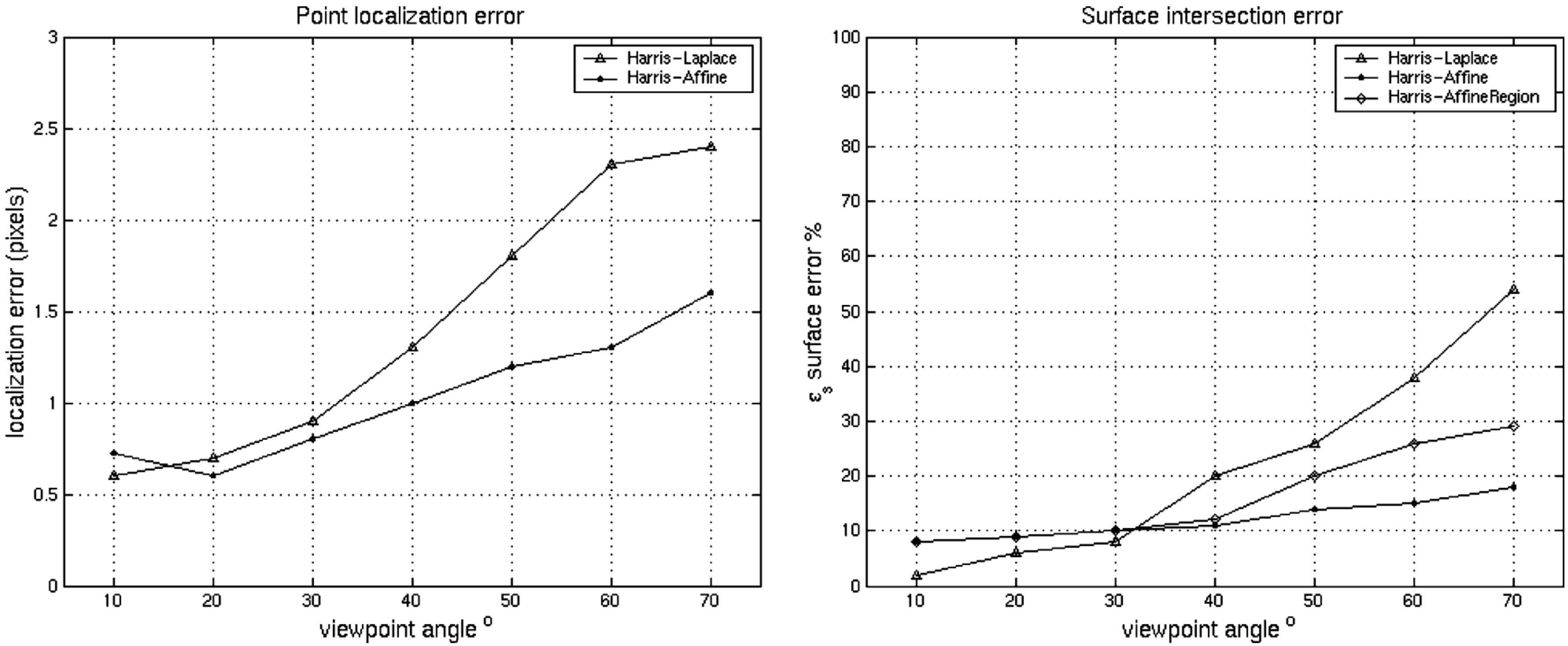
Graph taken from [1]

Repeatability with view point angle change in degrees



Graph taken from [1]

Localization and surface overlap Error with view point change



(a) (b)

Computational Complexity

- Image size 800 X 640.
- Pentium II 500 MHz
- Harris Laplace is O(n), n is number of pixel.
- Harris affine is O((m+k)p)
- where p is the number of initial points,
- m is the size of the search space for the automatic scale selection and
- k is the number of iterations required to compute the affine adaptation matrix

Computational Complexity

Table Complexity of the detectors. $g(I)$ denotes Gaussian smoothing. $H(I)$ denotes the Hessian matrix and $\mu(I)$ the second moment matrix computed for every image point. $(d_{xx} + d_{yy})$ is a convolution of a point neighborhood with a 2D Laplacian kernel. $\#n$ denotes the number of iterations per point patch, and can vary for different initial points.

Detector	Operation on image (initial points)	Operation on patch (scale)	Operation on patch (shape)	Run time seconds	Number of points
DoG	$\#12\ g(I)$			0.7	1527
Hessian	$\#12\ H(I)$			0.9	1832
H-L simplified	$\#12\ \mu(I)$	$\#3\ (d_{xx} + d_{yy})$		1.4	1625
H-L	$\#12\ \mu(I)$	$\#n\ (d_{xx} + d_{yy})$		7	1438
H-AR	$\#12\ \mu(I)$	$\#3\ (d_{xx} + d_{yy})$	$\#n\ \mu(\mathbf{x})$	12	1463
H-A	$\#12\ \mu(I)$	$\#7n\ (d_{xx} + d_{yy})$	$\#5n\ \mu(\mathbf{x})$	36	1123

Image taken from [1]

Application : Matching

- Descriptor: A set of Gaussian derivative up to 4^{th} order derivative, So 12 dimensional vector.
- Derivatives are computed on image patches normalized with the matrix U , which is estimated independently for each point
- Invariance to affine intensity changes is obtained by dividing the higher order derivatives by the first derivative.

[1]

Similarity Measures

- Mahalanobis distance is used to compute the similairty between two interest point.
- D(x,y)= sqrt((X-Y)T *inv(C)*(X-Y))
- X and Y are interest points.
- C is covariance matrix.
- Covariance matrix is estimated over a large set of images.

Wiki

Why Mahalanobis distance

- Why Mahalanobis distance
- Because takes into account the correlations of the data set and is scale-invariant.

- Outlier are removed by using RANSAC (RANdom SAmple Consensus).

Conclusion

- Scale invariant detector deals with large scale changes.
- Harris affine can deal with significant view changes transformation but it fails with large scale changes.
- Affine invariant detector gives more degree of freedom but it is not very discriminative.

References

- [1] Mikolajcyk, K. and Schmid, C. 2004. “An affine invariant interest point detector”. In Proceedings of the International Journal of Computer Vision 60(1), pp 63–86.
- [2] http://en.wikipedia.org/wiki/Harris-Affine last accessed 22/11/2012
- [3] Mikolajczyk, K. and Schmid, C. 2002. An affine invariant interest point detector. In *Proceedings of the 7th European Conference on Computer Vision*, Copenhagen, Denmark, vol. I, pp. 128–142.
- [4] T. Lindeberg (1998). "Feature detection with automatic scale selection". International Journal of Computer Vision 30 (2): pp 77—116.
- [5] Baumberg, A. 2000. Reliable feature matching across widely separated views. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, Hilton Head Island, South Carolina, USA, pp. 774–781.