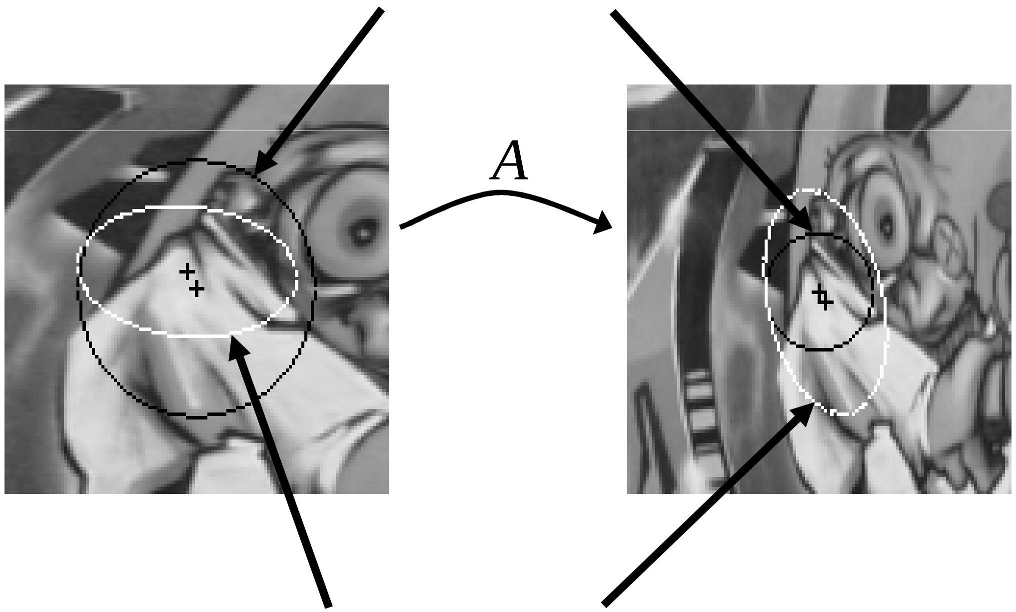
Overview

- Introduction to local features
- Harris interest points + SSD, ZNCC, SIFT
- Scale & **affine invariant interest point detectors**
- Evaluation and comparison of different detectors
- Region descriptors and their performance

Affine invariant regions - Motivation

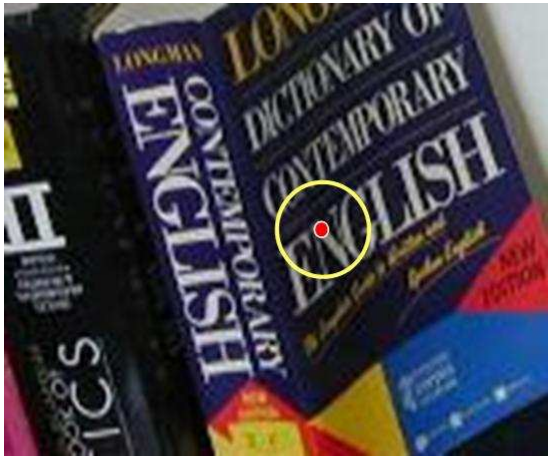
- Scale invariance is not sufficient for large baseline changes

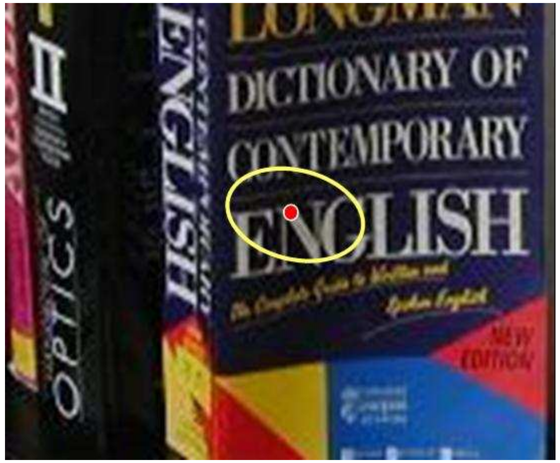
detected scale invariant region



projected regions, viewpoint changes can locally be approximated by an affine transformation A

Affine invariant regions - Motivation





Affine invariant regions - Example

Harris/Hessian/Laplacian-Affine

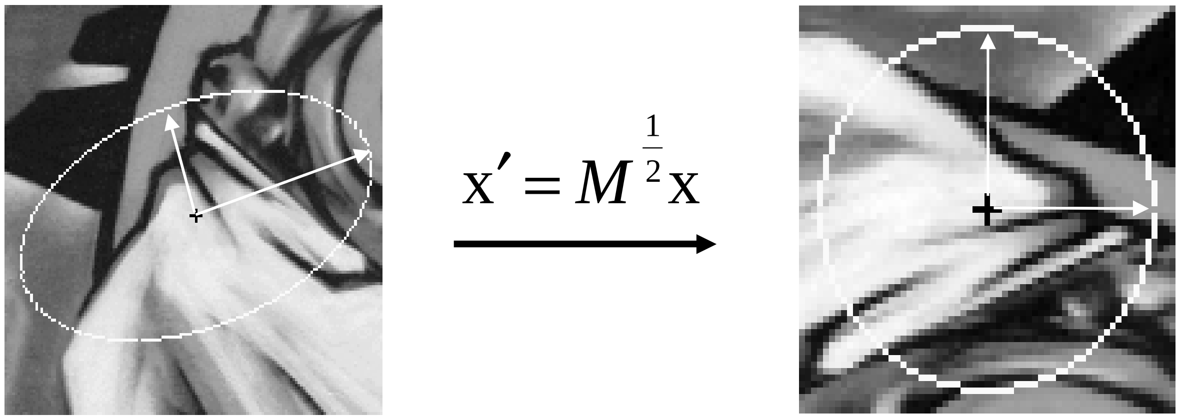
- Initialize with scale-invariant Harris/Hessian/Laplacian points
- Estimation of the affine neighbourhood with the second moment matrix [Lindeberg'94]
- Apply affine neighbourhood estimation to the scale-invariant interest points [Mikolajczyk & Schmid'02, Schaffalitzky & Zisserman'02]
- Excellent results in a comparison [Mikolajczyk et al.'05]

Affine invariant regions

- Based on the second moment matrix (Lindeberg'94)

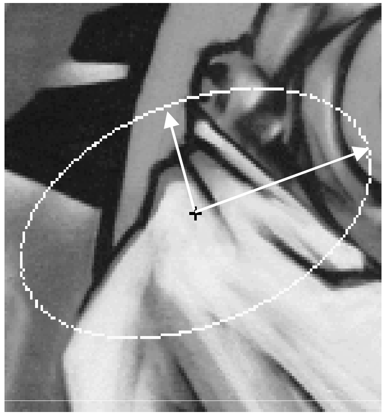
$$M = \mu(\mathbf{x}, \sigma_I, \sigma_D) = \sigma_D^2 G(\sigma_I) \otimes \begin{bmatrix} L_x^2(\mathbf{x}, \sigma_D) & L_x L_y(\mathbf{x}, \sigma_D) \\ L_x L_y(\mathbf{x}, \sigma_D) & L_y^2(\mathbf{x}, \sigma_D) \end{bmatrix}$$

- Normalization with eigenvalues/eigenvectors

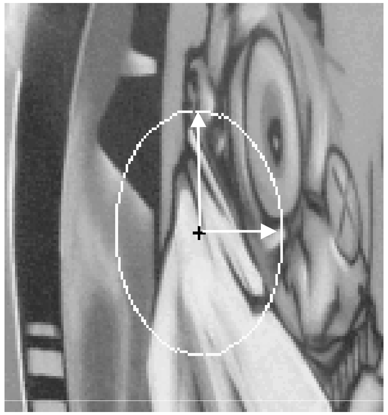


$$\mathrm{x}' = M^{\frac{1}{2}}\mathrm{x}$$

Affine invariant regions

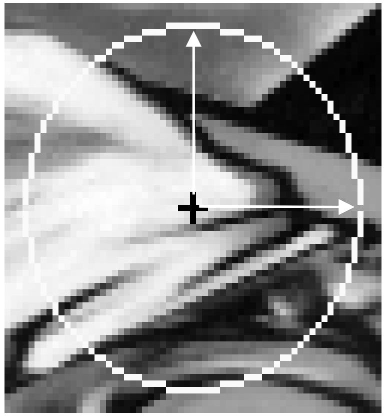


$$x_R = A x_L$$

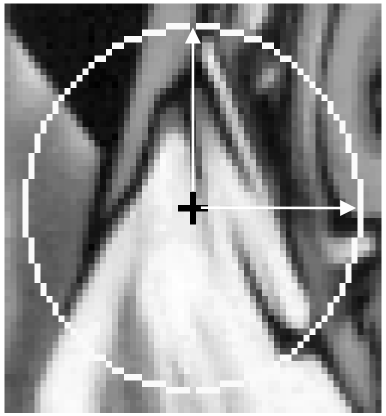


$$x_L' = M_L^{\frac{1}{2}} x_L$$

$$x_R' = M_R^{\frac{1}{2}} x_R$$



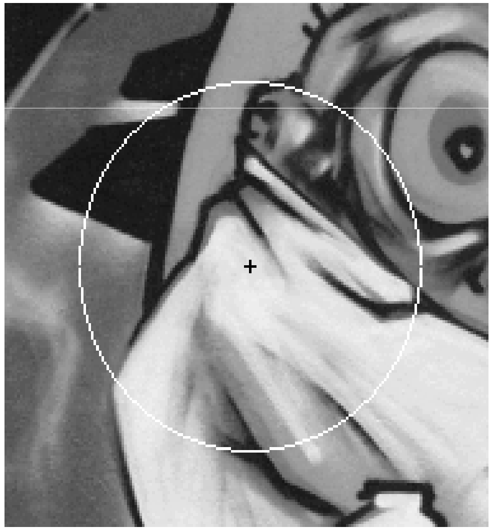
$$x_R' = R x_L'$$

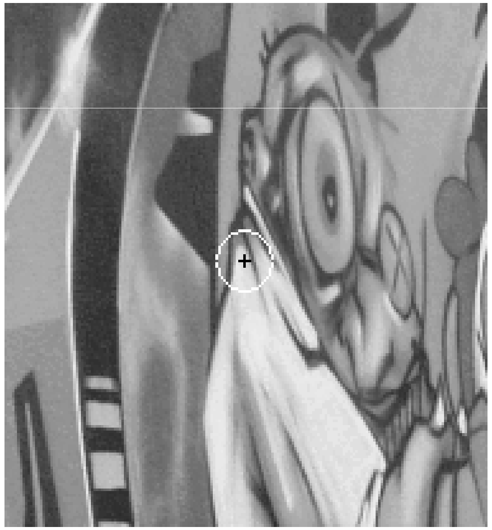


Isotropic neighborhoods related by image rotation

Affine invariant regions - Estimation

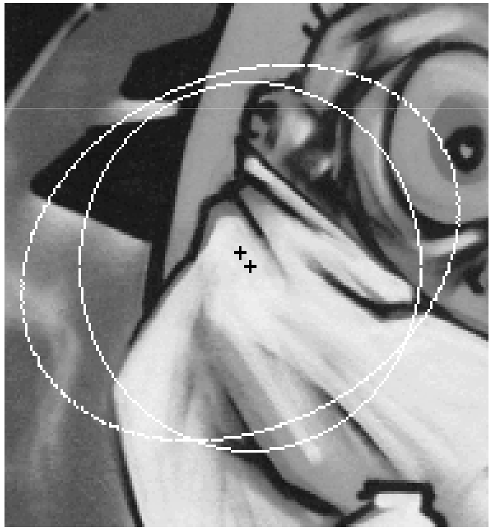
- Iterative estimation – initial points





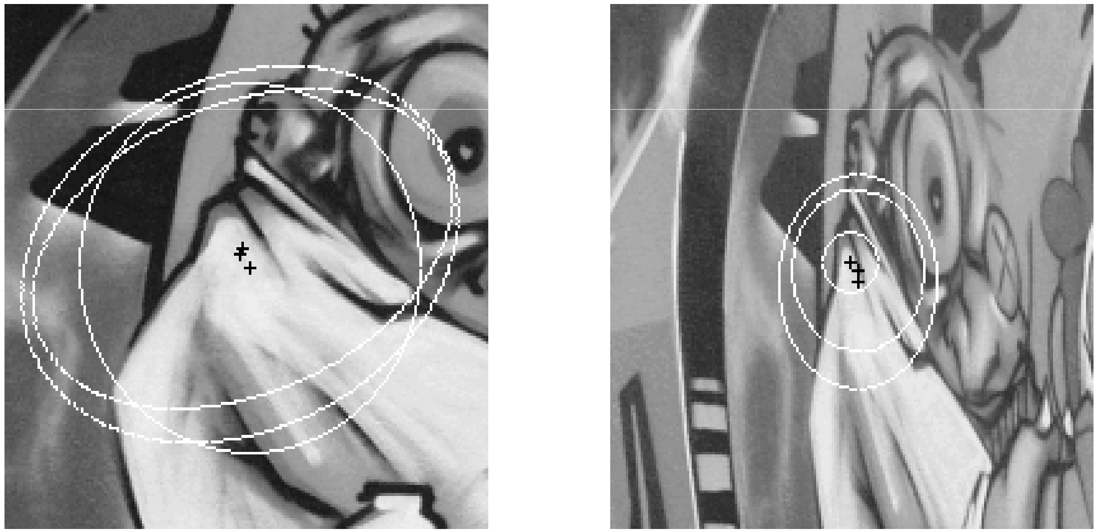
Affine invariant regions - Estimation

- Iterative estimation – iteration #1



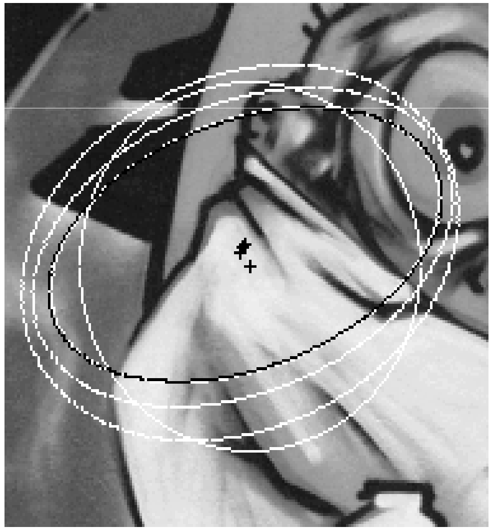
Affine invariant regions - Estimation

- Iterative estimation – iteration #2

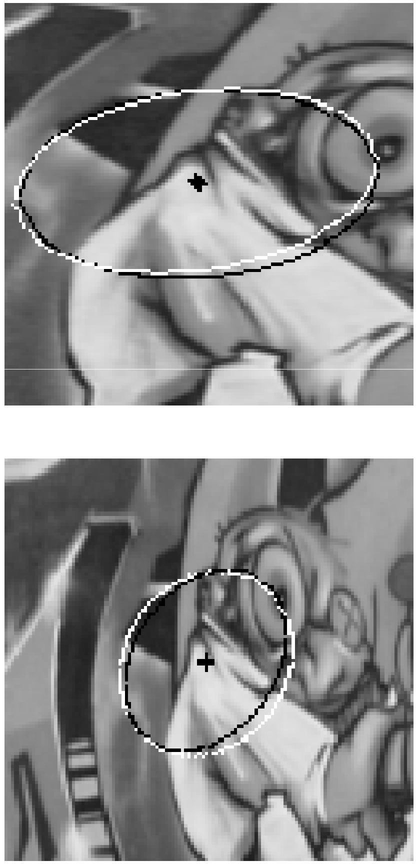


Affine invariant regions - Estimation

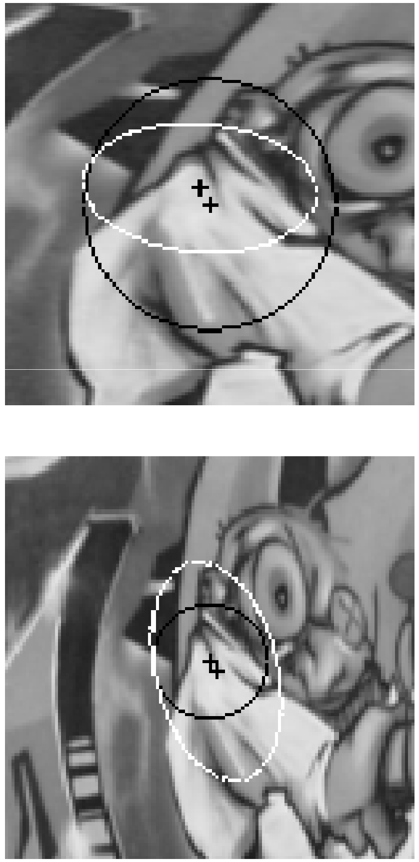
- Iterative estimation – iteration #3, #4



Harris-Affine versus Harris-Laplace

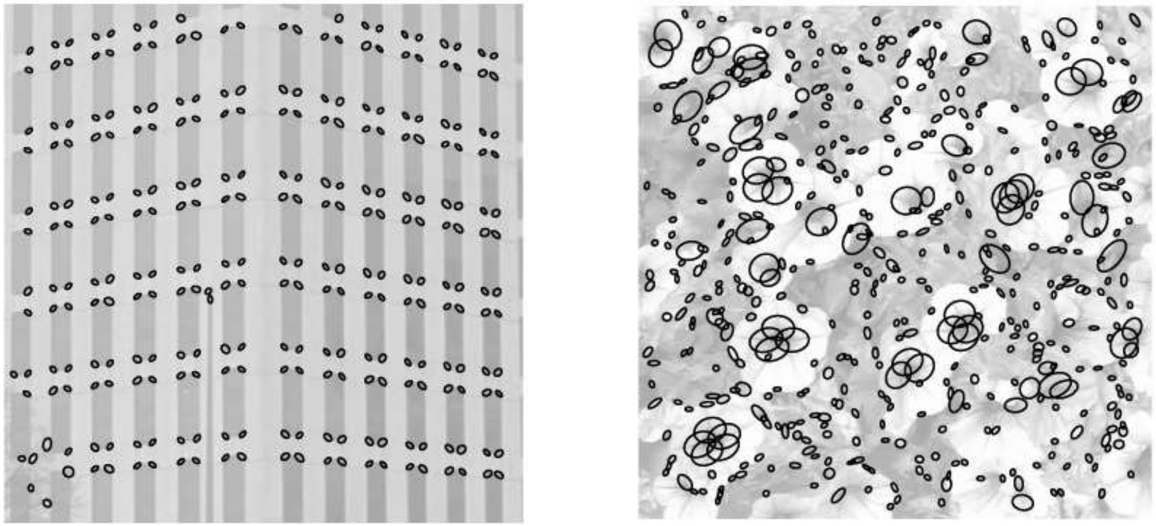


Harris-Affine

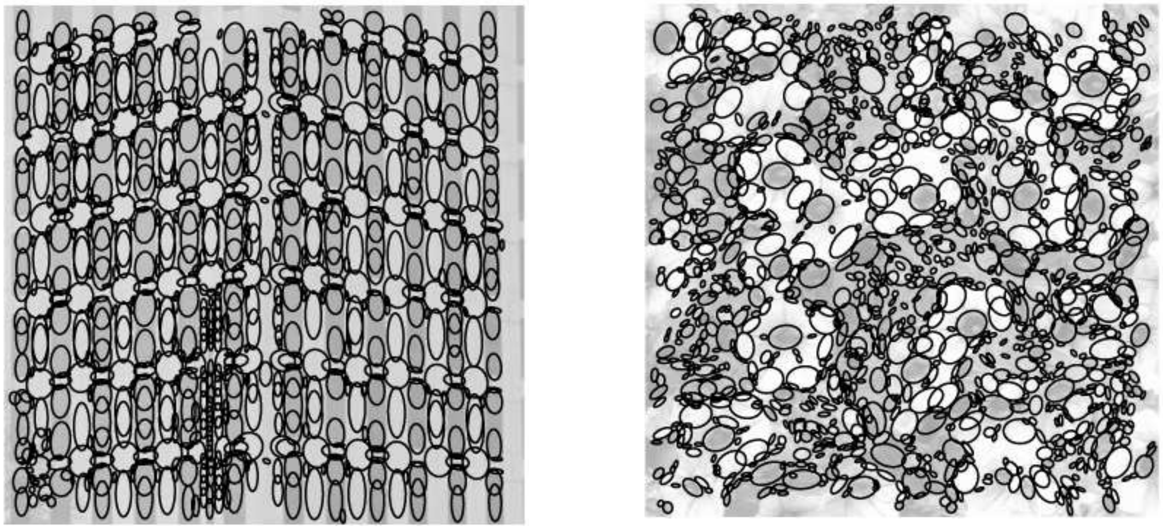


Harris-Laplace

Harris/Hessian-Affine

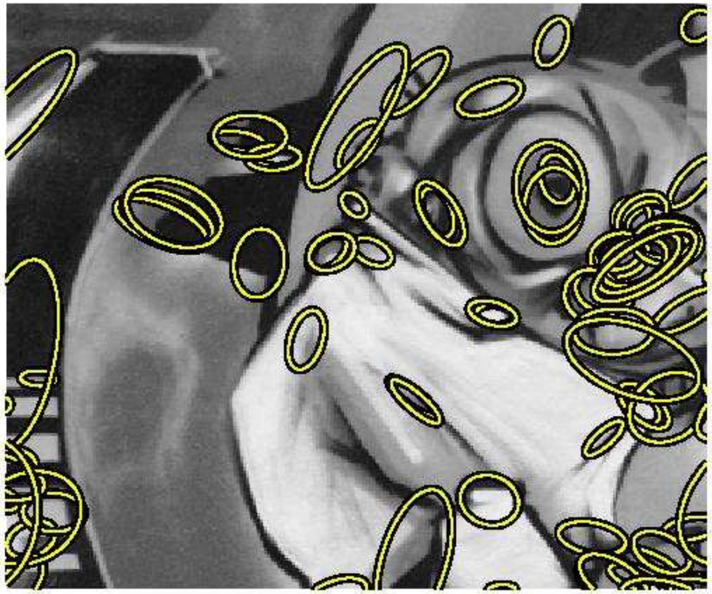


Harris-Affine



Hessian-Affine

Harris-Affine

Hessian-Affine

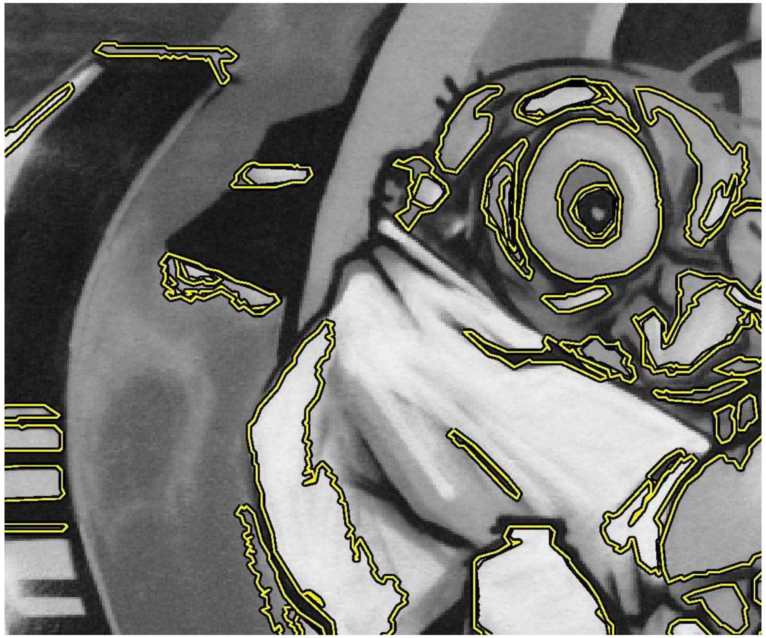
Matches

22 correct matches

Matches

33 correct matches

Maximally stable extremal regions (MSER) [Matas'02]



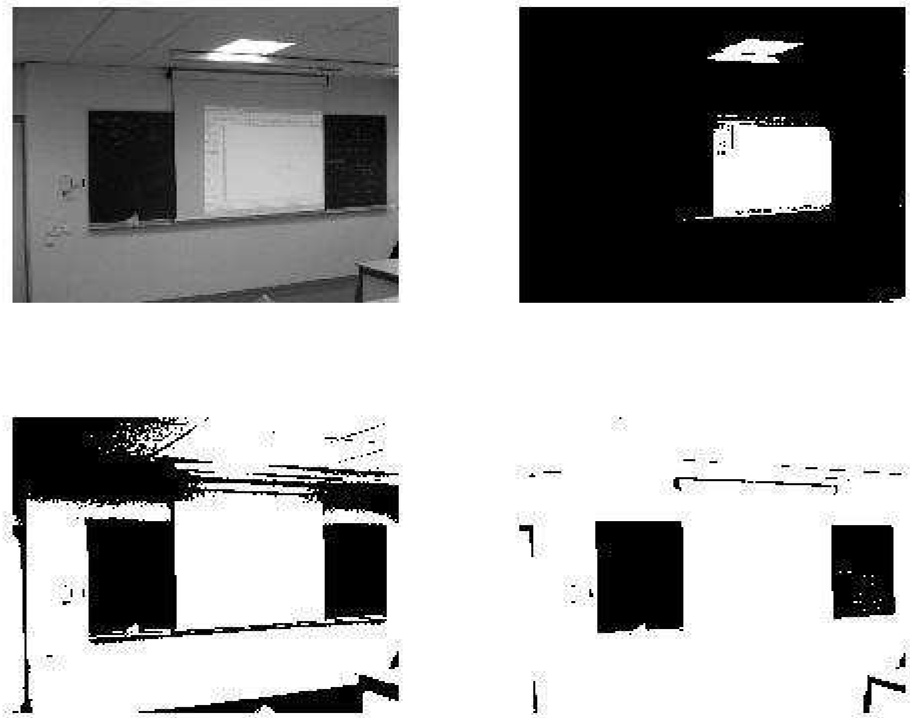
- Based on the idea of region segmentation
- State of the art results

Maximally stable extremal regions (MSER) [Matas'02]

- Extremal regions: connected components in a thresholded image (all pixels above/below a threshold)
- Maximally stable: minimal change of the component (area) for a change of the threshold, i.e. region remains stable for a change of threshold

Maximally stable extremal regions (MSER)

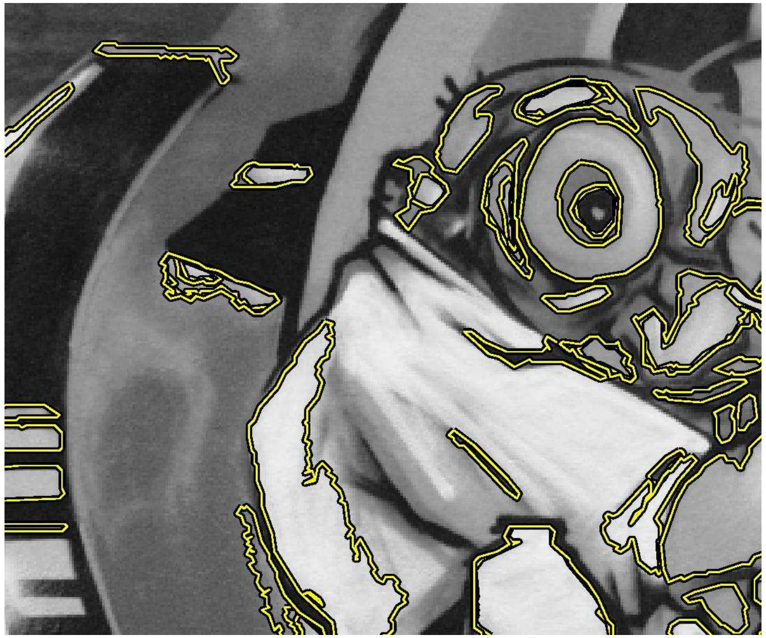
Examples of thresholded images

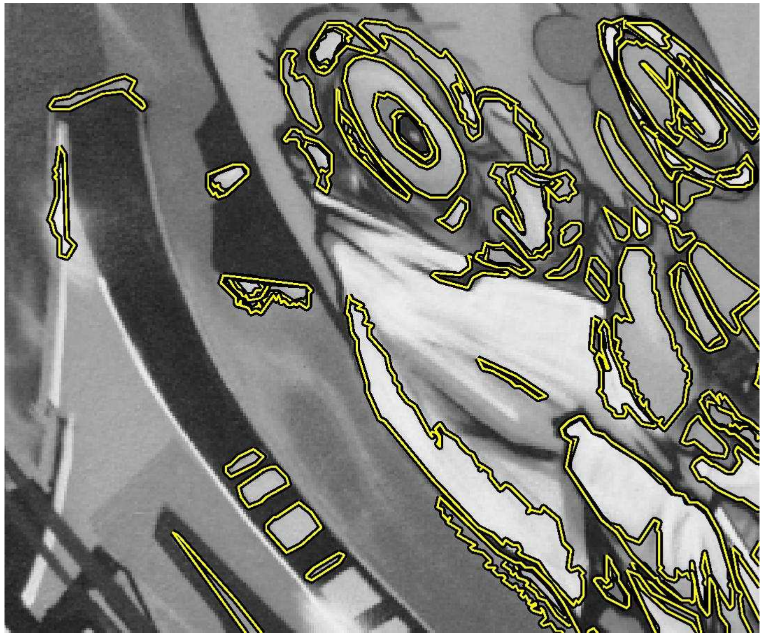


high threshold

low threshold

MSER





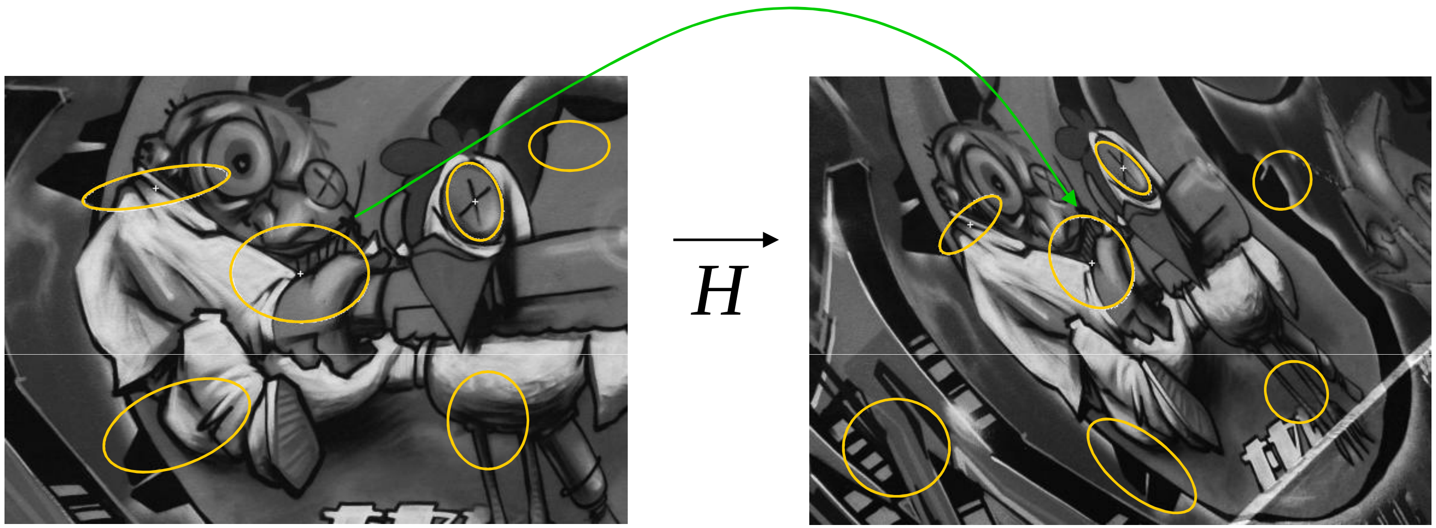
Overview

- Introduction to local features
- Harris interest points + SSD, ZNCC, SIFT
- Scale & affine invariant interest point detectors
- **Evaluation and comparison of different detectors**
- Region descriptors and their performance

Evaluation of interest points

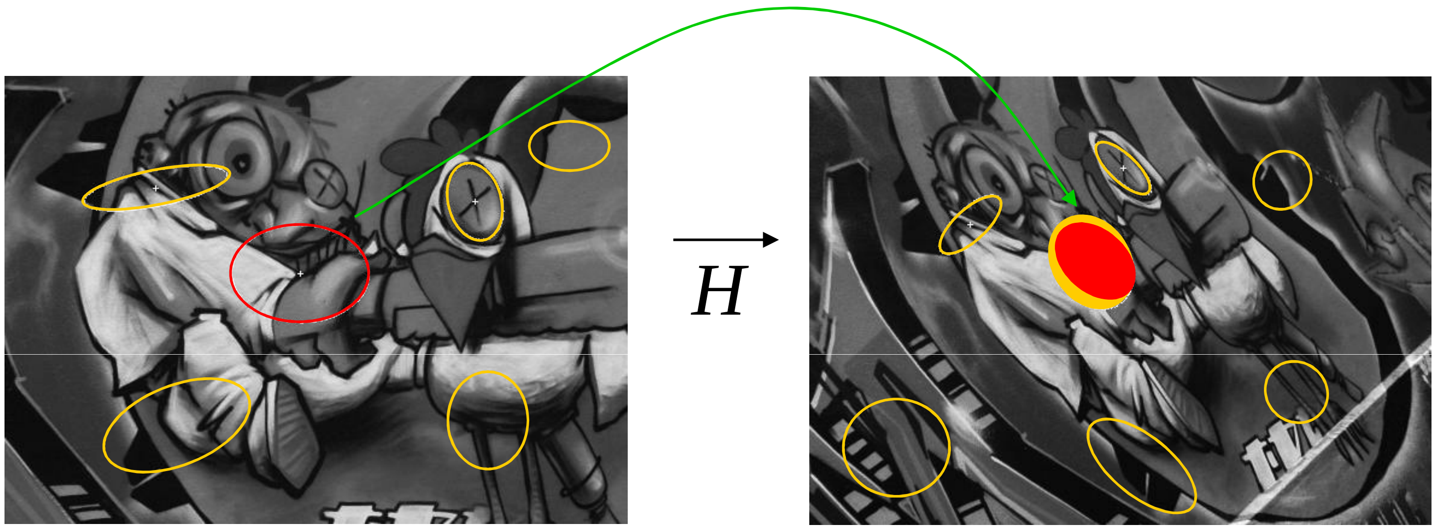
- Quantitative evaluation of interest point/region detectors
  - points / regions at the same relative location and area
- Repeatability rate : percentage of corresponding points
- Two points/regions are corresponding if
  - location error small
  - area intersection large
- [K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir & L. Van Gool ’05]

Evaluation criterion



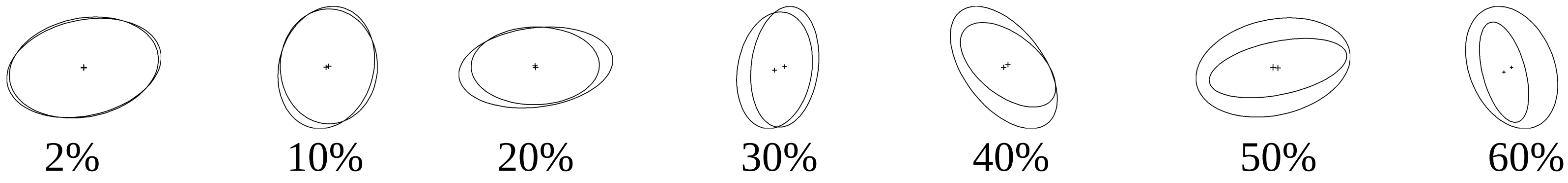
$$repeatability = \frac{\#\,corresponding\ regions}{\#\,detected\ regions} \cdot 100\%$$

Evaluation criterion



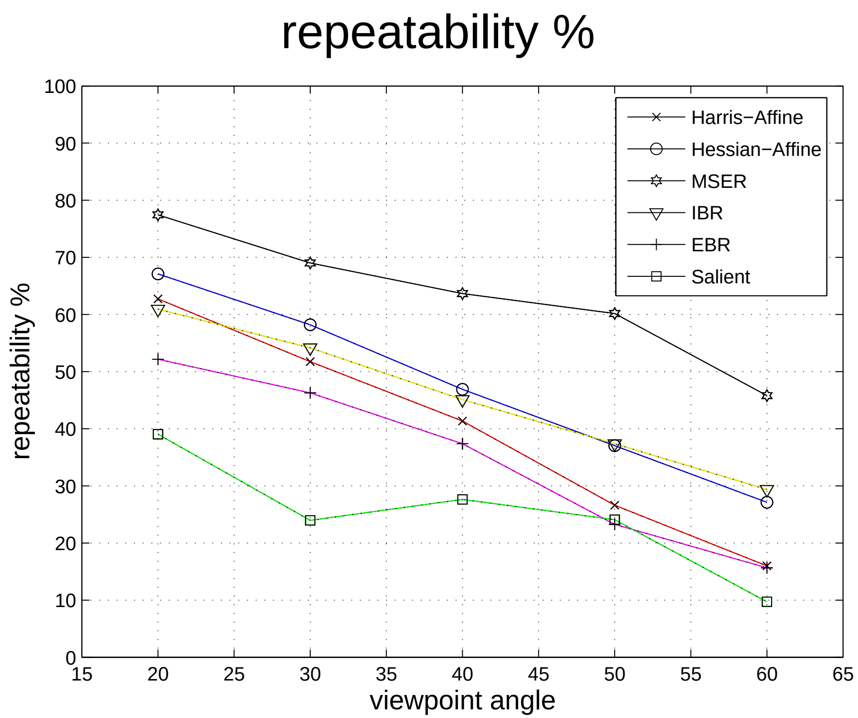
$$repeatability = \frac{\#corresponding\ regions}{\#detected\ regions} \cdot 100\%$$

$$overlap\ error = (1 - \frac{intersection}{union}) \cdot 100\%$$



Comparison of affine invariant detectors

Viewpoint change - structured scene

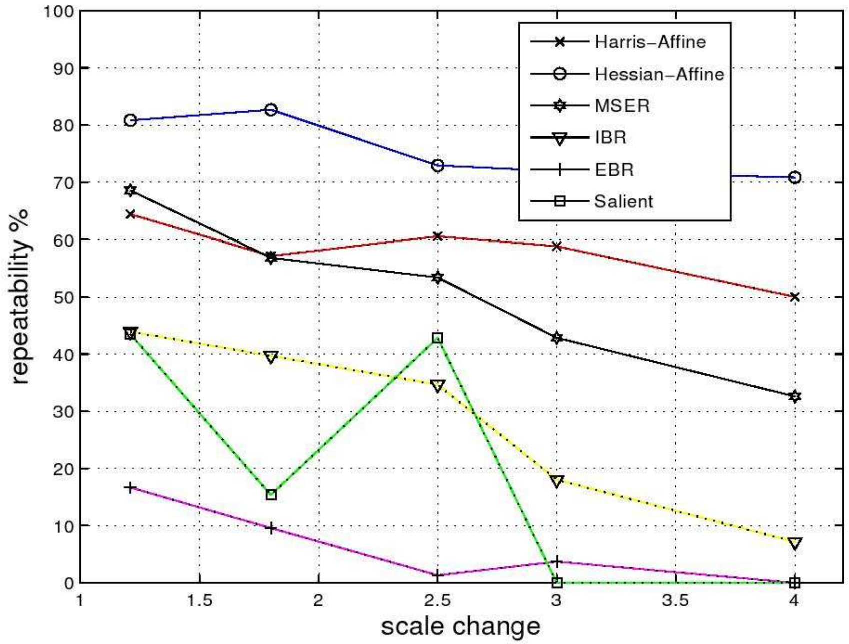


reference image 20 40 60

Comparison of affine invariant detectors

Scale change – textured scene

repeatability %



reference image

4

Conclusion - detectors

- Good performance for large viewpoint and scale changes

- Results depend on transformation and scene type, no one best detector

- Detectors are complementary
  - MSER adapted to structured scenes
  - Harris and Hessian adapted to textured scenes

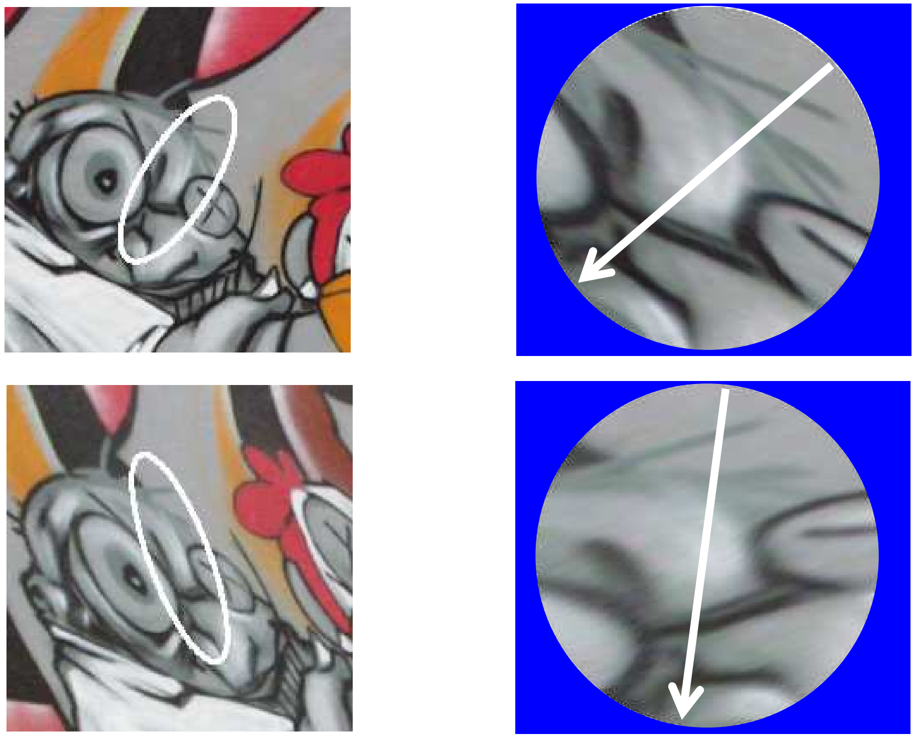
- Performance of the different scale invariant detectors is very similar (Harris-Laplace, Hessian, LoG and DOG)

- Scale-invariant detector sufficient up to 40 degrees of viewpoint change

Overview

- Introduction to local features
- Harris interest points + SSD, ZNCC, SIFT
- Scale & affine invariant interest point detectors
- Evaluation and comparison of different detectors
- **Region descriptors and their performance**

Region descriptors



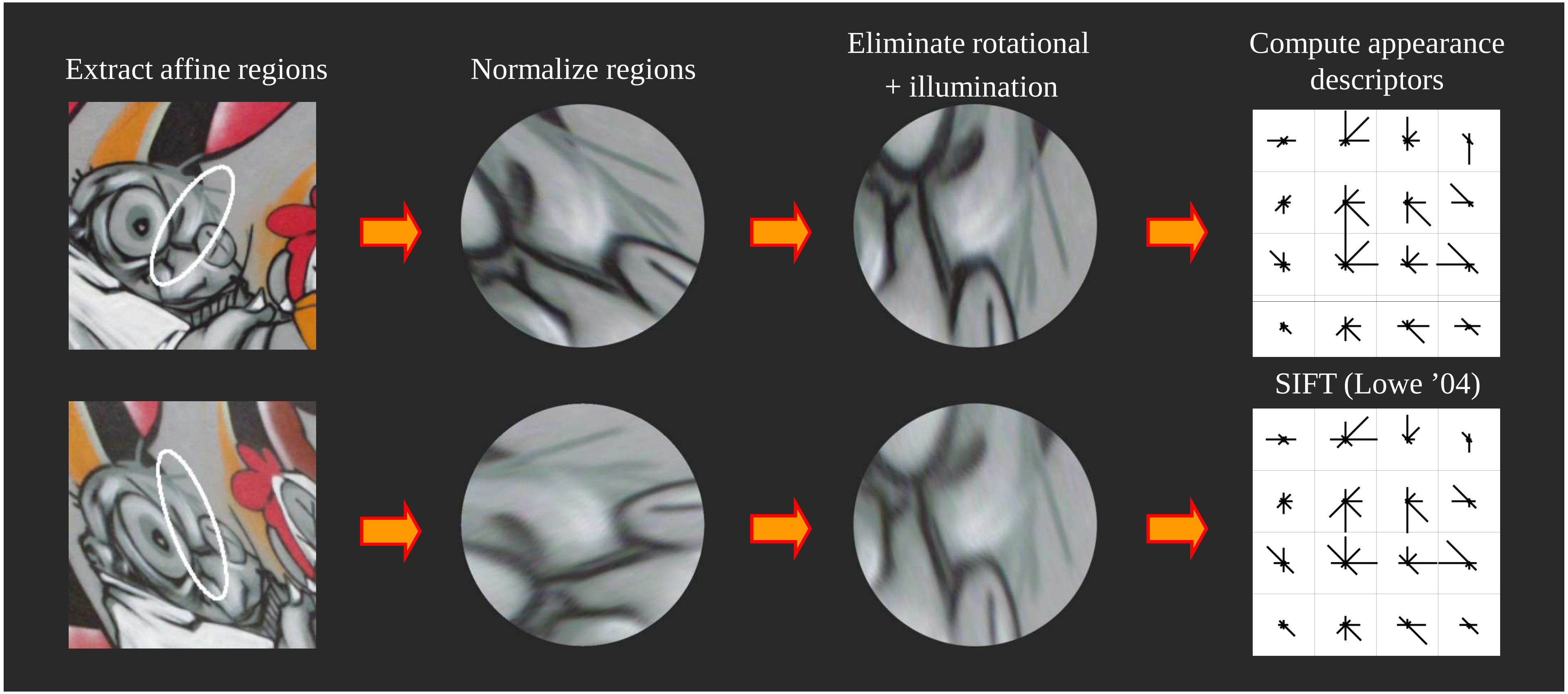
- Normalized regions are
  - invariant to geometric transformations except rotation
  - not invariant to photometric transformations

Descriptors

- Regions invariant to geometric transformations except rotation
  - rotation invariant descriptors
  - **normalization with dominant gradient direction**

- Regions not invariant to photometric transformations
  - invariance to affine photometric transformations
  - **normalization with mean and standard deviation of the image patch**

Descriptors



Descriptors

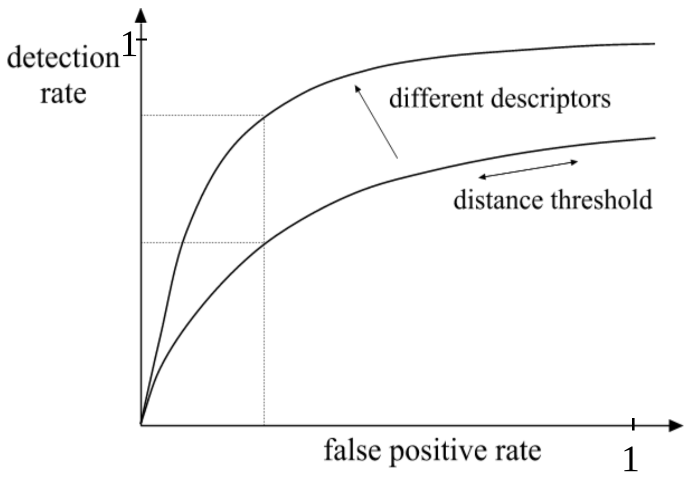
- Gaussian derivative-based descriptors
  - Differential invariants (*Koenderink and van Doorn'87*)
  - Steerable filters (*Freeman and Adelson'91*)

- SIFT (*Lowe'99)*
- Moment invariants [Van Gool et al.'96]
- Shape context [Belongie et al.'02]
- SIFT with PCA dimensionality reduction
- Gradient PCA [Ke and Sukthankar'04]
- SURF descriptor [Bay et al.'08]
- DAISY descriptor [Tola et al.'08, Windler et al'09]

Comparison criterion

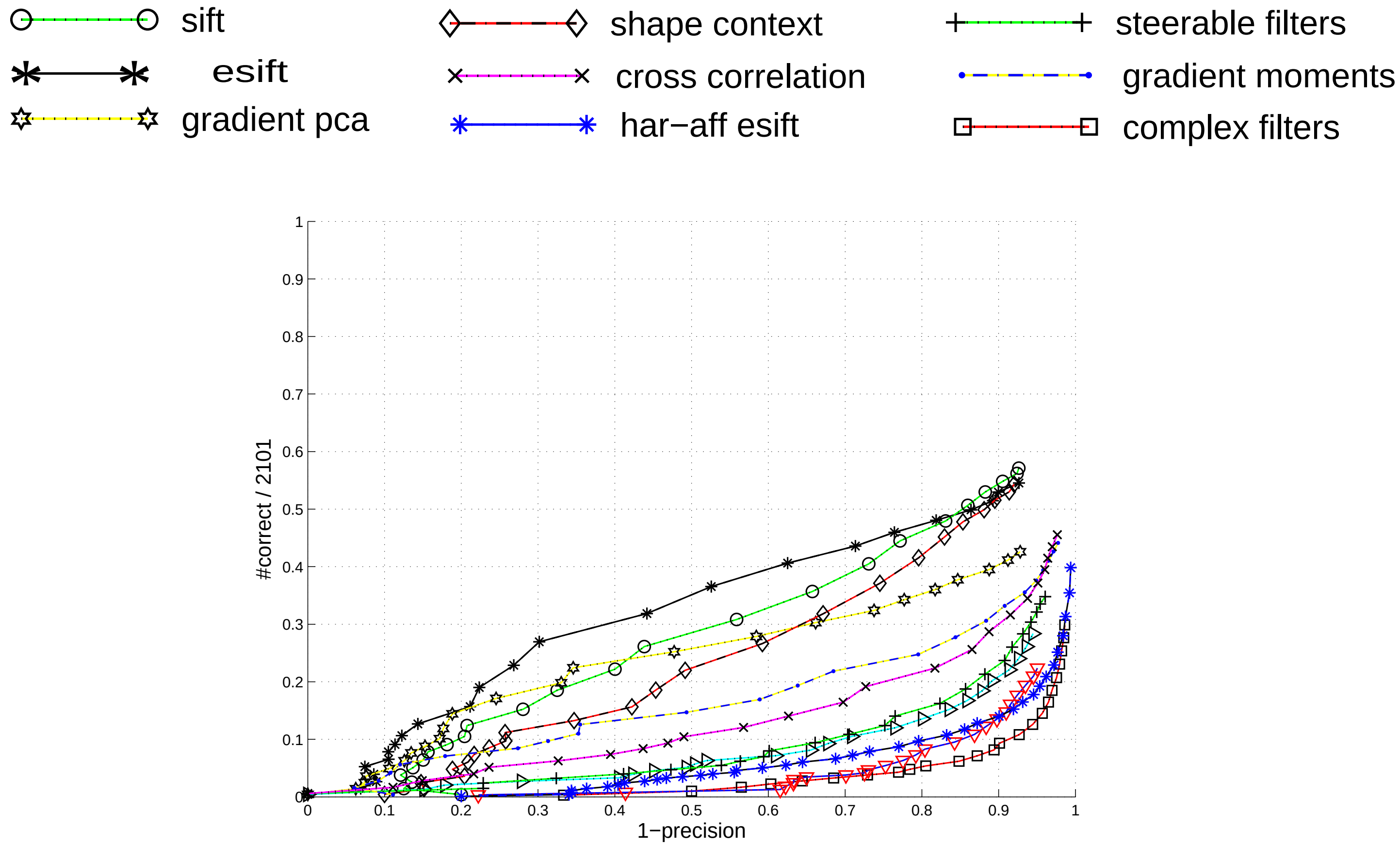
- Descriptors should be
  - Distinctive
  - Robust to changes on viewing conditions as well as to errors of the detector

- Detection rate (recall)
  - #correct matches / #correspondences
- False positive rate
  - #false matches / #all matches
- Variation of the distance threshold
  - distance (d1, d2) < threshold

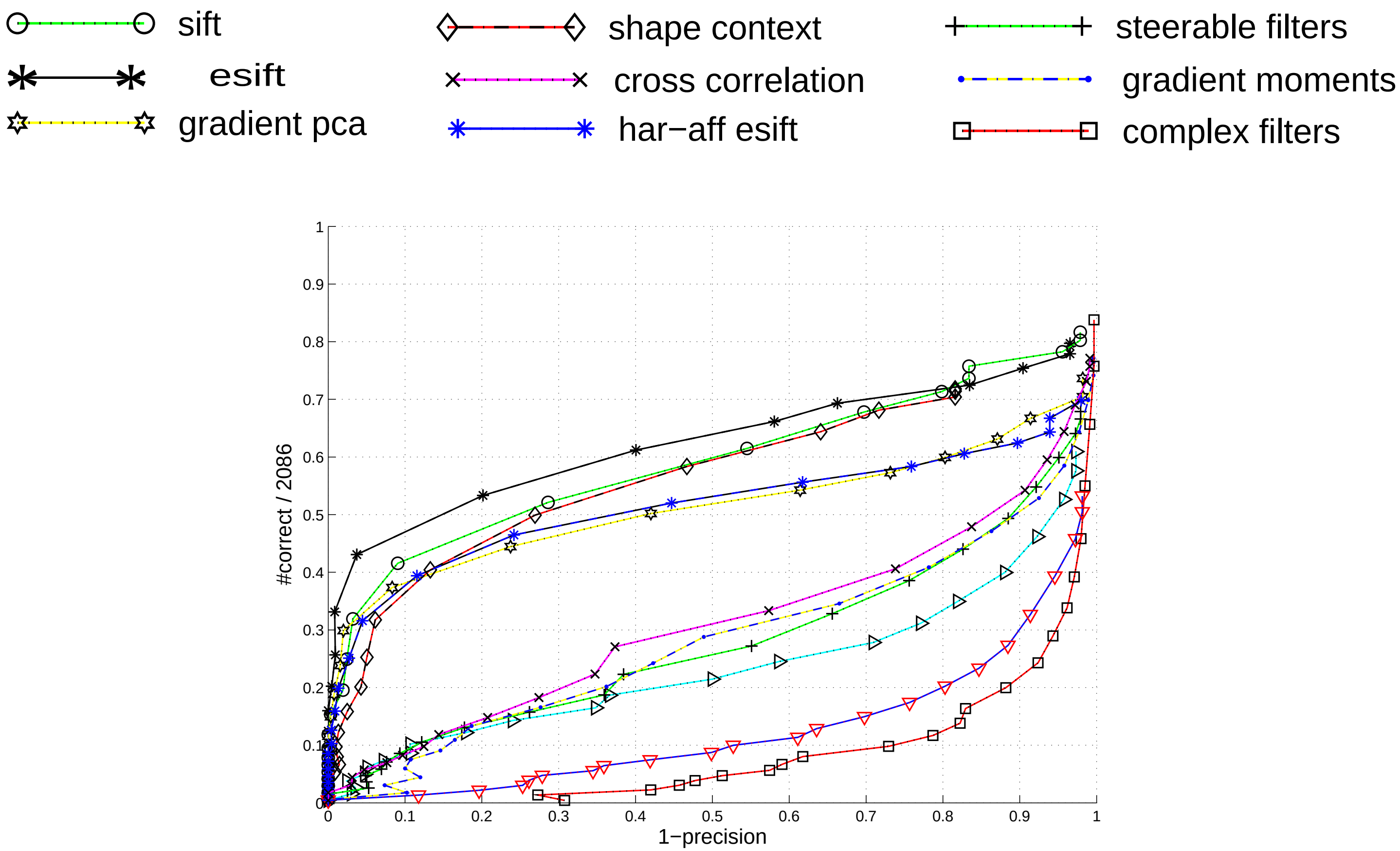


[K. Mikolajczyk & C. Schmid, PAMI'05]

Viewpoint change (60 degrees)



Scale change (factor 2.8)



Conclusion - descriptors

- SIFT based descriptors perform best
- Significant difference between SIFT and low dimension descriptors as well as cross-correlation
- Robust region descriptors better than point-wise descriptors
- Performance of the descriptor is relatively independent of the detector

Available on the internet

- Binaries for detectors and descriptors
  - *Building blocks for recognition systems*

- Carefully designed test setup
  - Dataset with transformations
  - Evaluation code in matlab
  - *Benchmark for new detectors and descriptors*

http://lear.inrialpes.fr/software