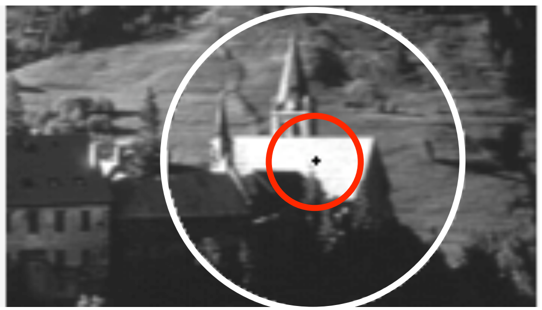
# Overview

- Harris interest points
- Comparing interest points (SSD, ZNCC, SIFT)
- **Scale & affine invariant interest points**
- Evaluation and comparison of different detectors
- Region descriptors and their performance

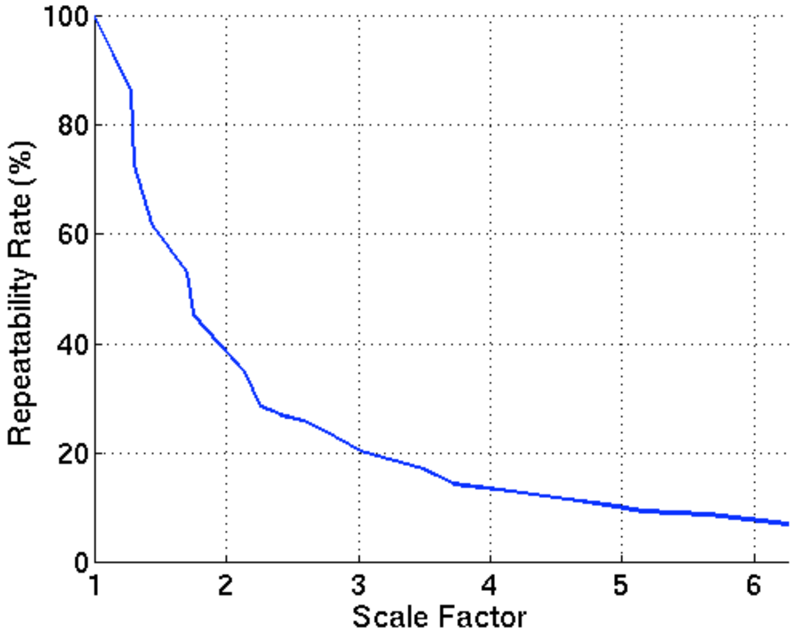
# Scale invariance - motivation

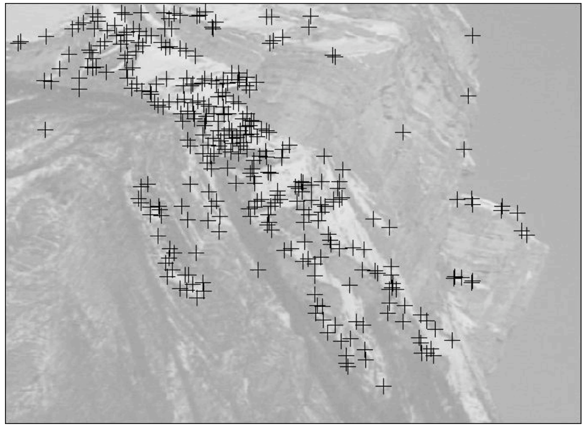
- Description regions have to be adapted to scale changes

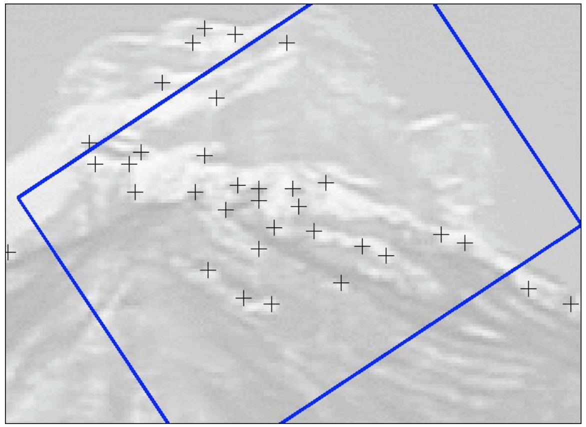


- Interest points have to be repeatable for scale changes

# Harris detector + scale changes





Repeatability rate

$$R(\varepsilon) = \frac{|\{(\mathbf{a}_i, \mathbf{b}_i) \mid dist(H(\mathbf{a}_i), \mathbf{b}_i) < \varepsilon\}|}{\max(|\mathbf{a}_i|, |\mathbf{b}_i|)}$$

# Scale adaptation

Scale change between two images

$$I_1\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = I_2\begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = I_2\begin{pmatrix} sx_1 \\ sy_1 \end{pmatrix}$$

Scale adapted derivative calculation

# Scale adaptation

Scale change between two images

$$I_1\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = I_2\begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = I_2\begin{pmatrix} sx_1 \\ sy_1 \end{pmatrix}$$

Scale adapted derivative calculation

$$I_1\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} \otimes G_{i_1 \dots i_n}(\sigma) = \boldsymbol{s}^{\boldsymbol{n}} I_2\begin{pmatrix} x_2 \\ y_2 \end{pmatrix} \otimes G_{i_1 \dots i_n}(\boldsymbol{s\sigma})$$

# Scale adaptation

$$G(\tilde{\sigma}) \otimes \begin{bmatrix} L_x^2(\sigma) & L_x L_y(\sigma) \\ L_x L_y(\sigma) & L_y^2(\sigma) \end{bmatrix}$$

where $L_i(\sigma)$ are the derivatives with Gaussian convolution

# Scale adaptation

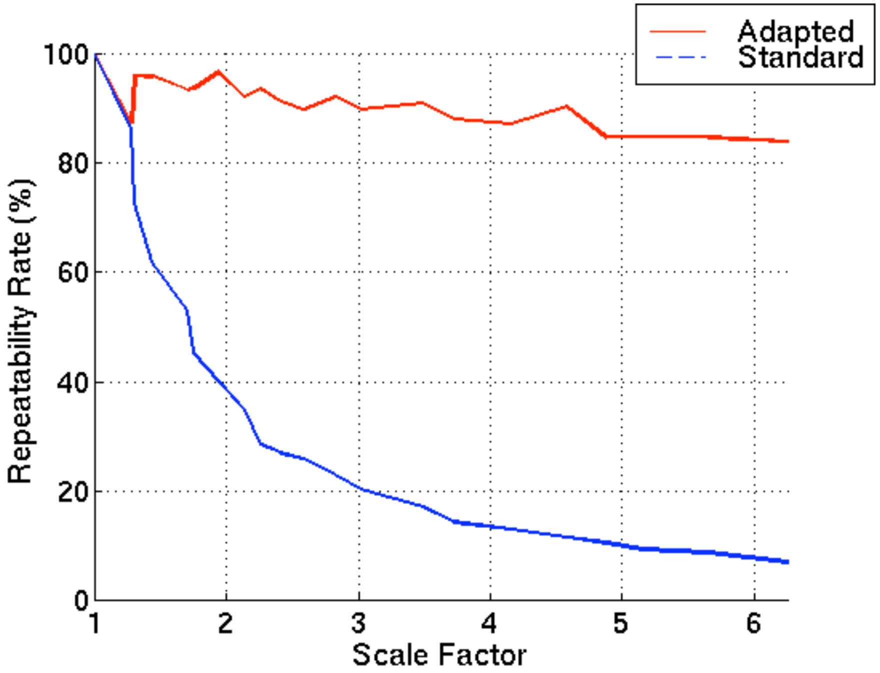
$$G(\widetilde{\sigma}) \otimes \begin{bmatrix} L_x^2(\sigma) & L_x L_y(\sigma) \\ L_x L_y(\sigma) & L_y^2(\sigma) \end{bmatrix}$$

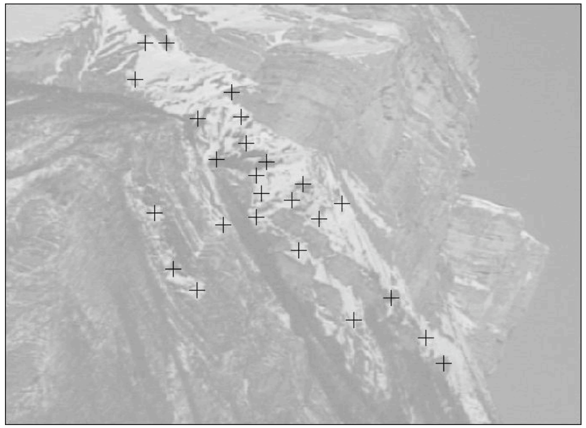
where $L_i(\sigma)$ are the derivatives with Gaussian convolution

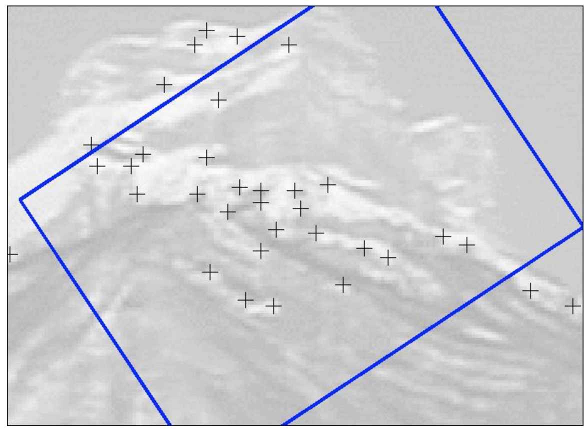
## Scale adapted auto-correlation matrix

$$s^2 G(s\widetilde{\sigma}) \otimes \begin{bmatrix} L_x^2(s\sigma) & L_x L_y(s\sigma) \\ L_x L_y(s\sigma) & L_y^2(s\sigma) \end{bmatrix}$$

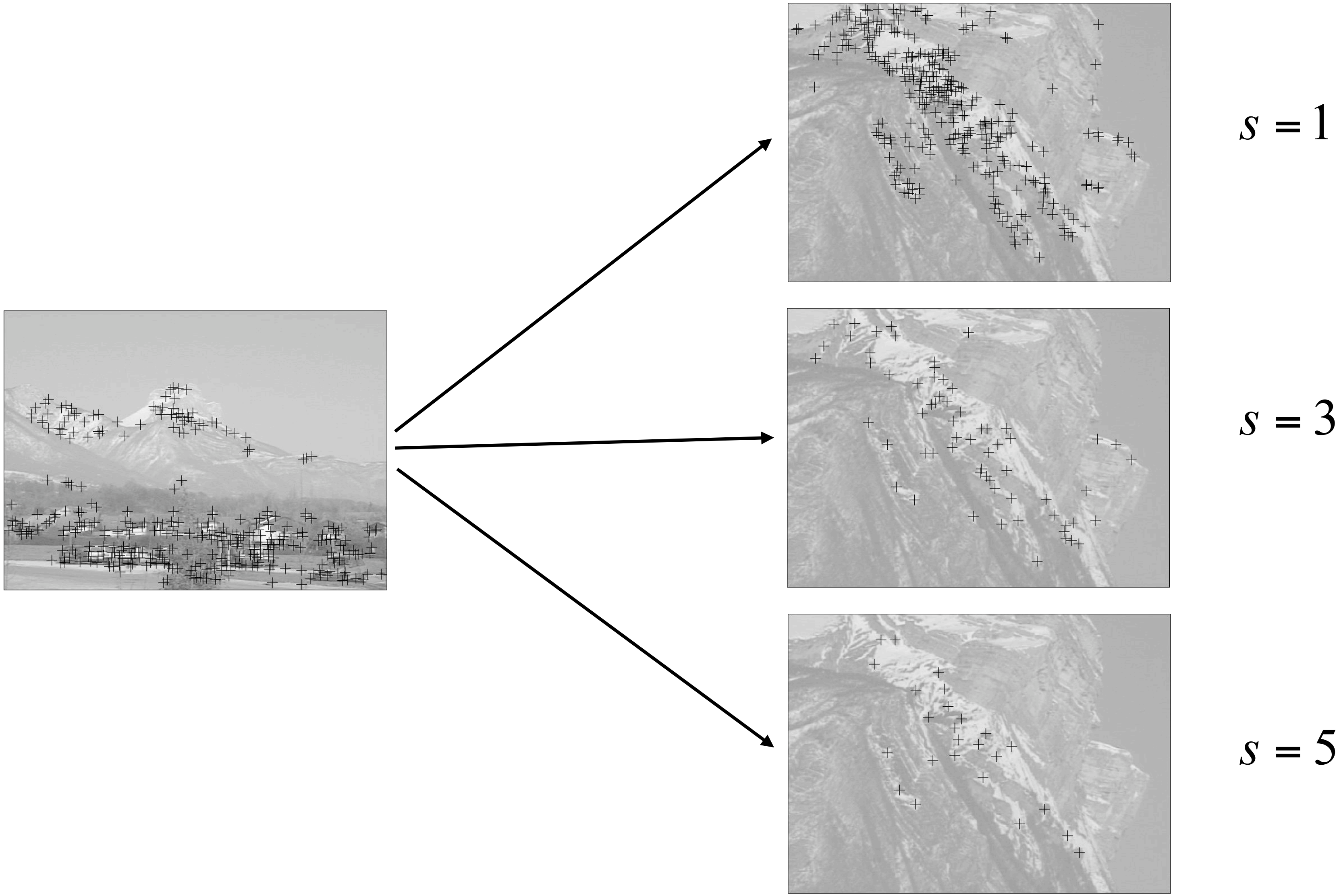
# Harris detector – adaptation to scale



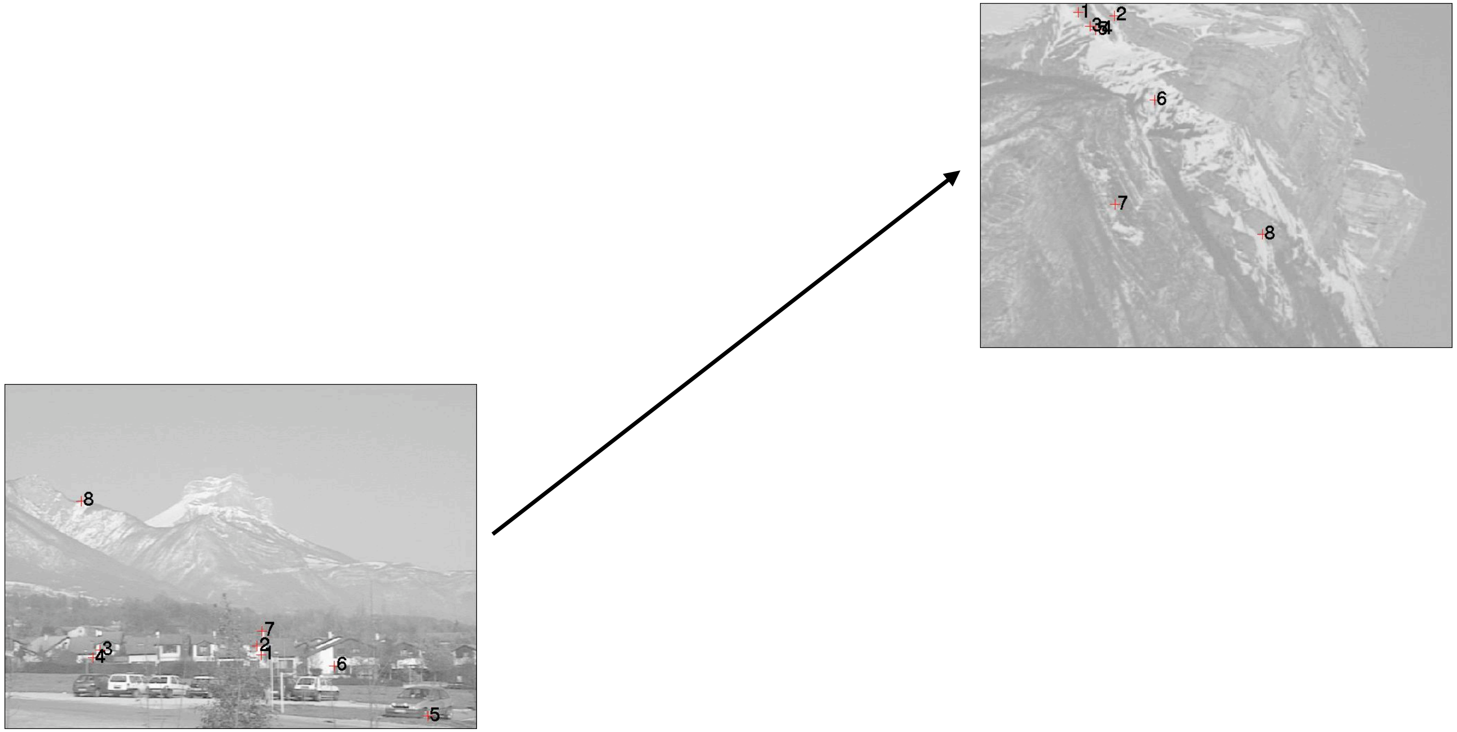





# Multi-scale matching algorithm



# Multi-scale matching algorithm

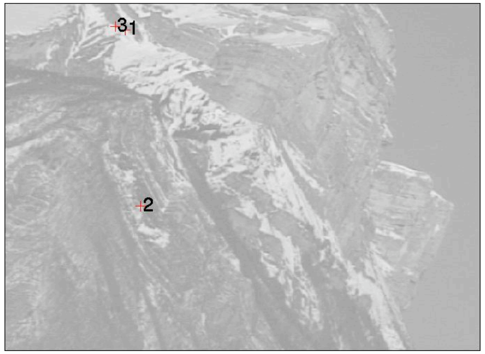


$s = 1$

8 matches

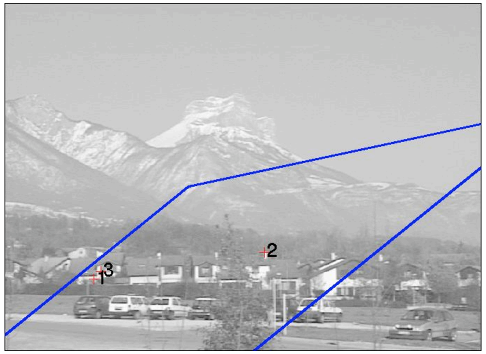
# Multi-scale matching algorithm

Robust estimation of a global affine transformation

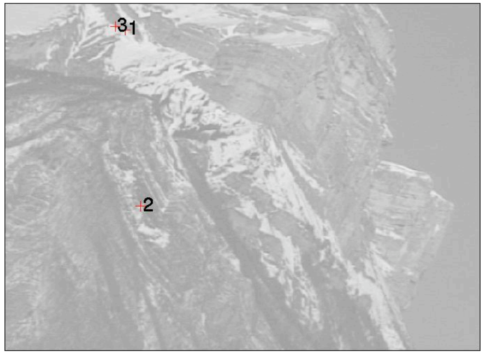



$s = 1$

3 matches

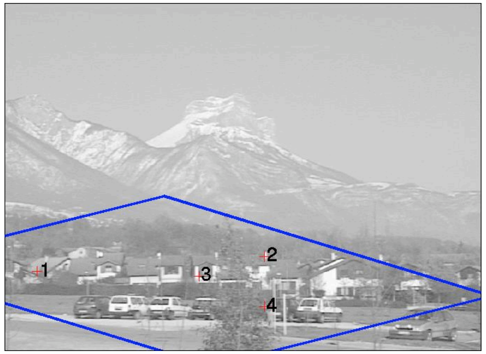


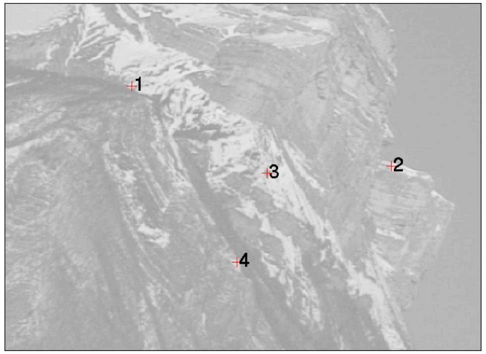
# Multi-scale matching algorithm




$s = 1$

3 matches

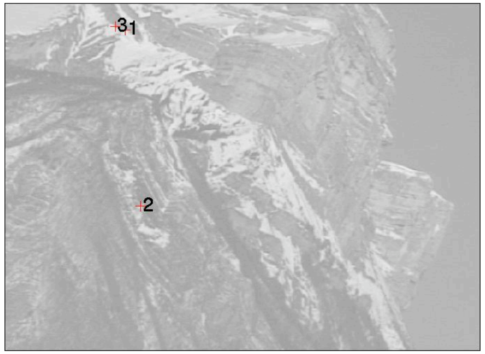



$s = 3$

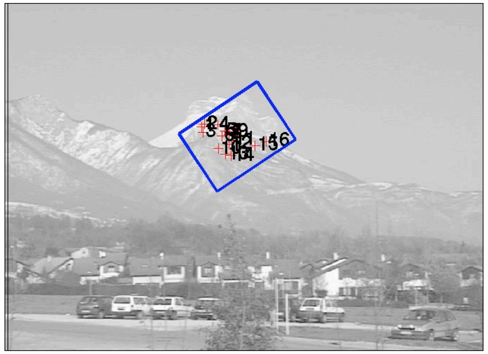
4 matches

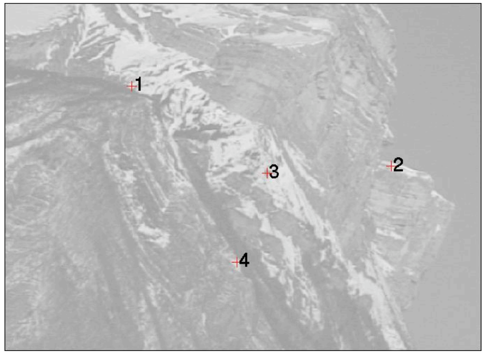
# Multi-scale matching algorithm



$s = 1$

3 matches



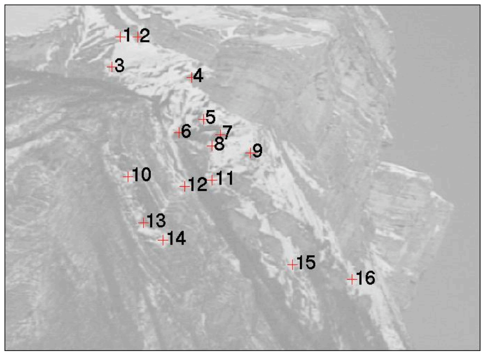


$s = 3$

4 matches

highest number of matches

correct scale

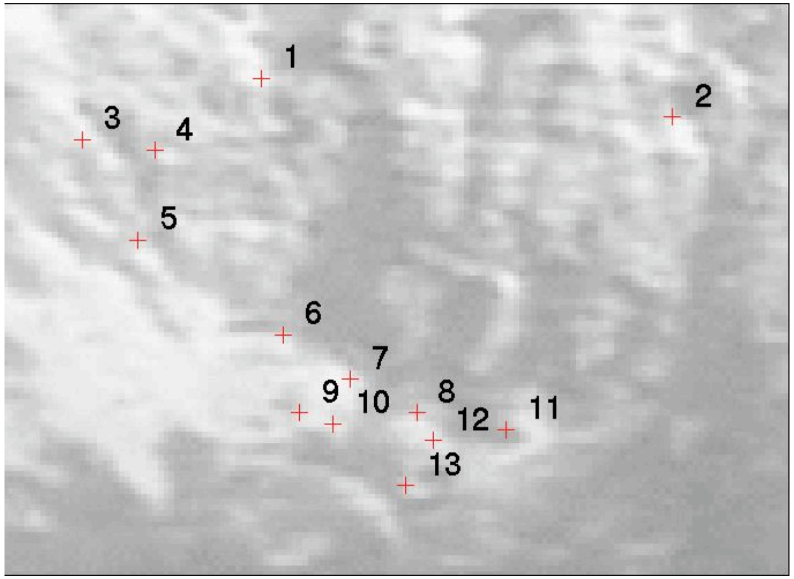


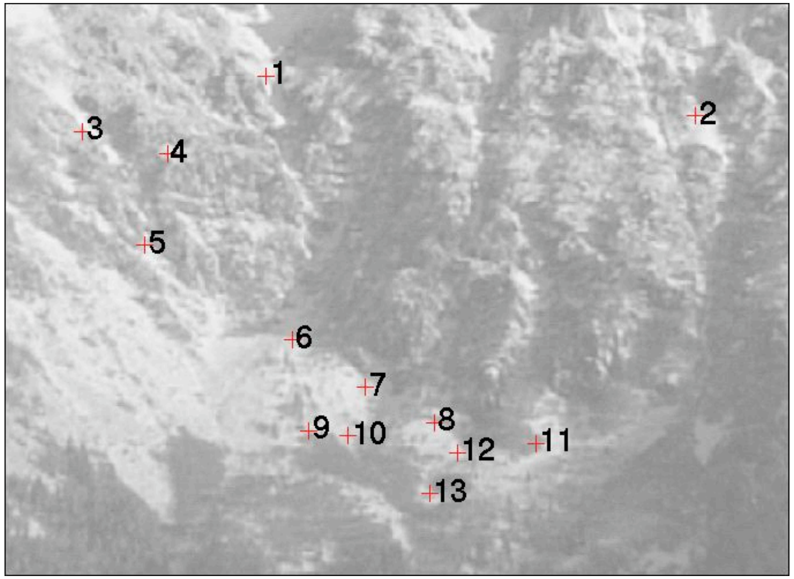
$\boxed{s = 5 \quad \text{16 matches}}$

# Matching results

Scale change of 5.7

# Matching results

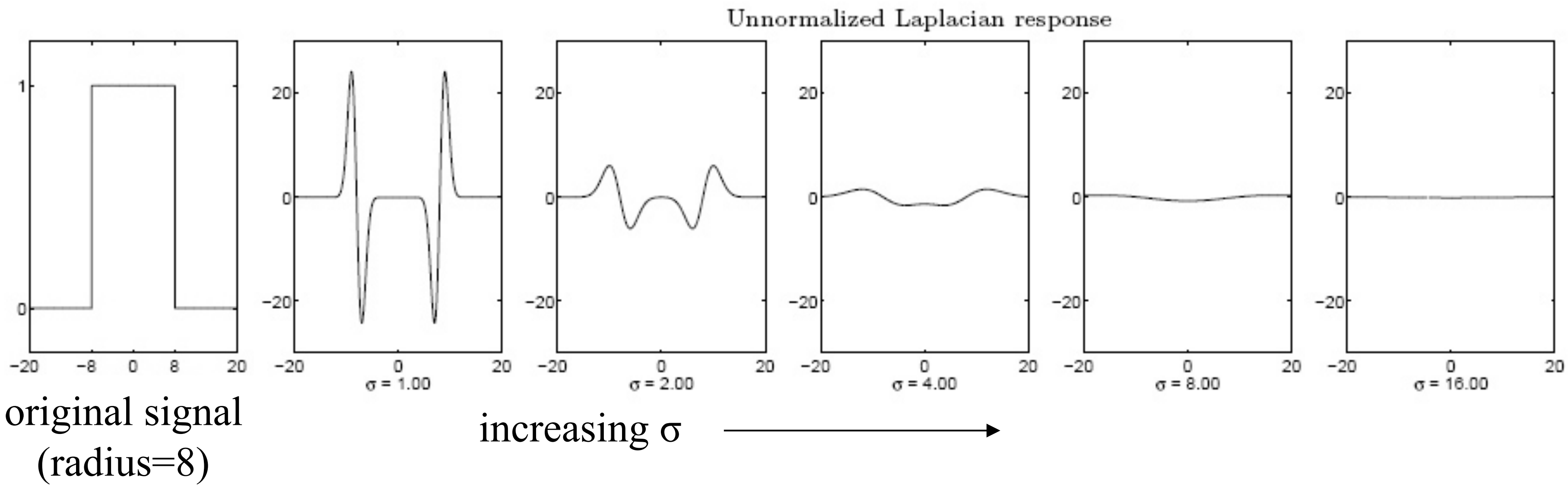



100% correct matches (13 matches)

# Scale selection

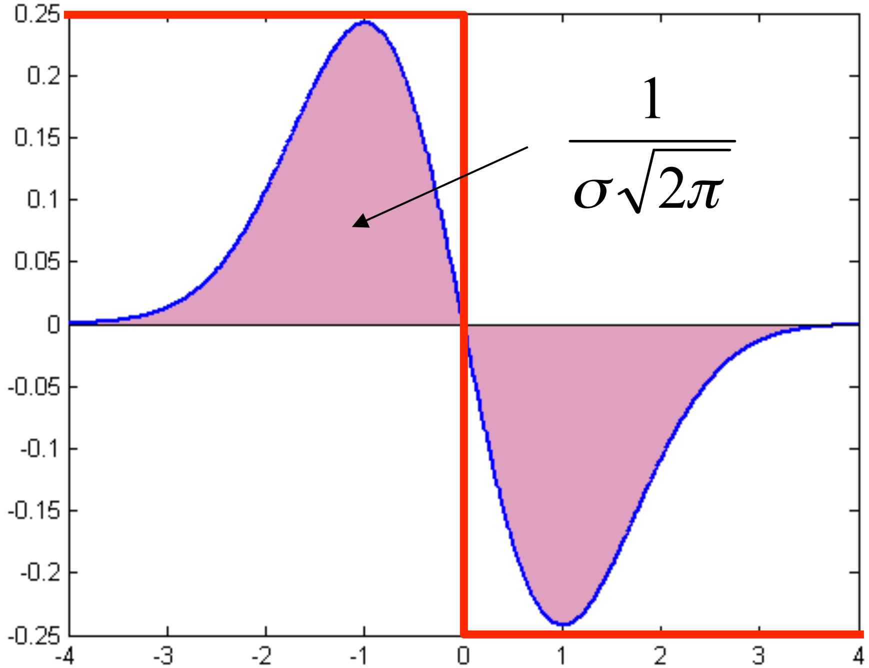
- We want to find the characteristic scale by convolving it with, for example, Laplacians at several scales and looking for the maximum response
- However, Laplacian response decays as scale increases:




Why does this happen?

# Scale normalization

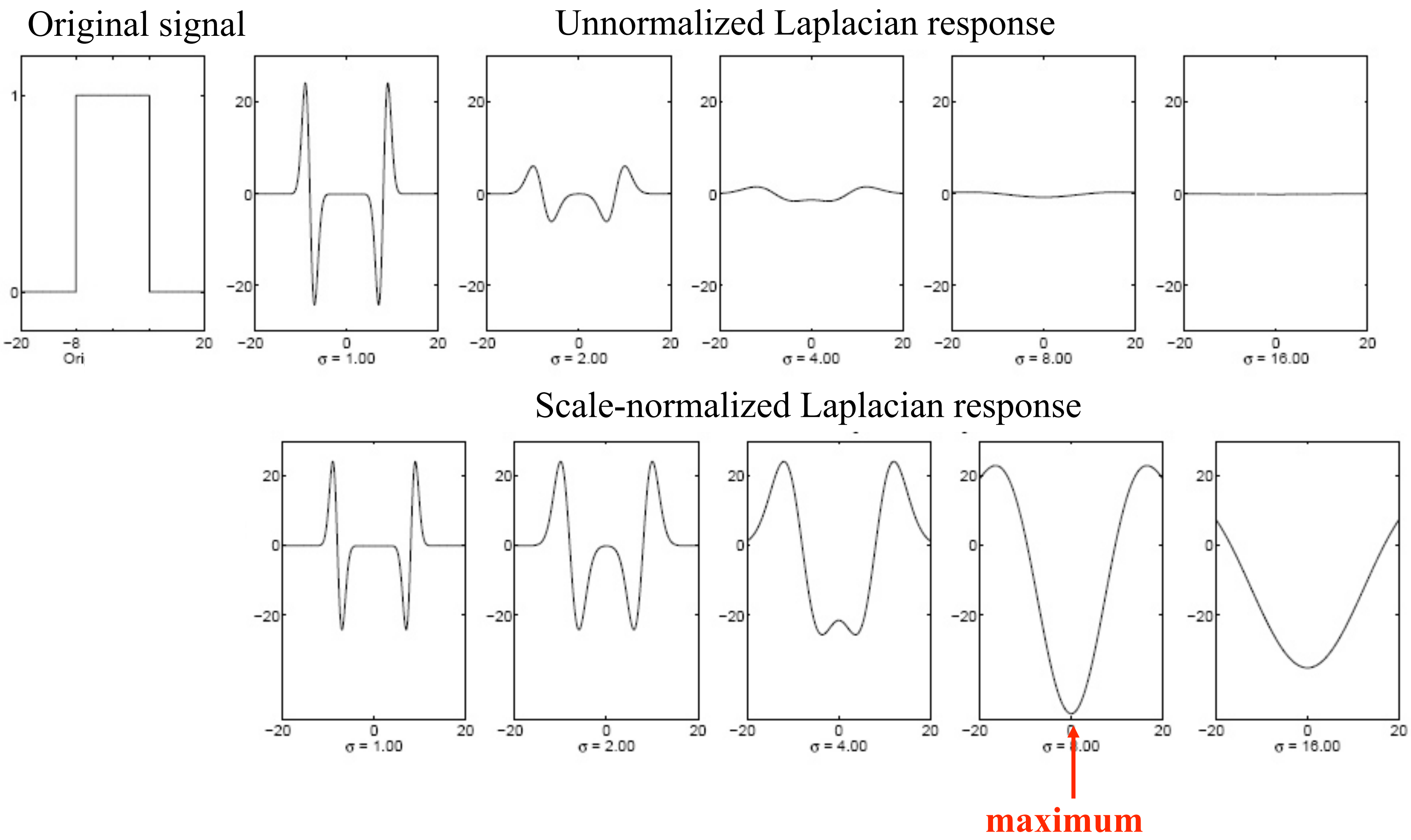
- The response of a derivative of Gaussian filter to a perfect step edge decreases as σ increases



# Scale normalization

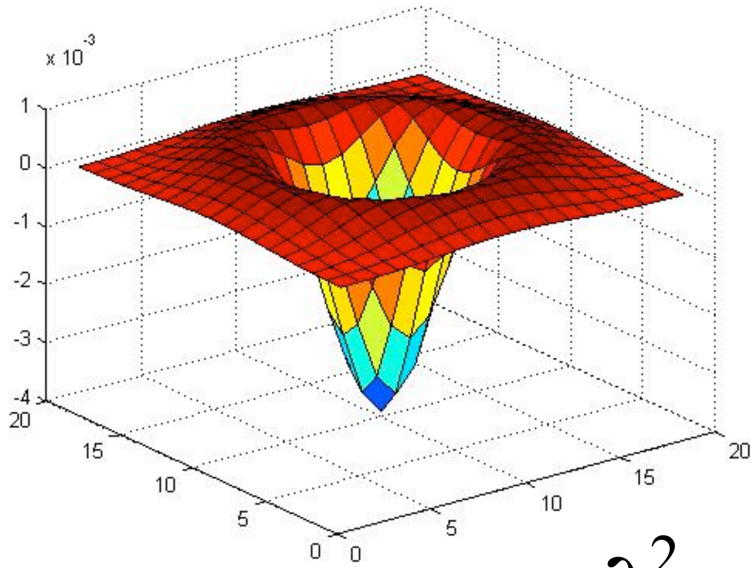
- The response of a derivative of Gaussian filter to a perfect step edge decreases as $\sigma$ increases
- To keep response the same (scale-invariant), must multiply Gaussian derivative by $\sigma$
- Laplacian is the second Gaussian derivative, so it must be multiplied by $\sigma^2$

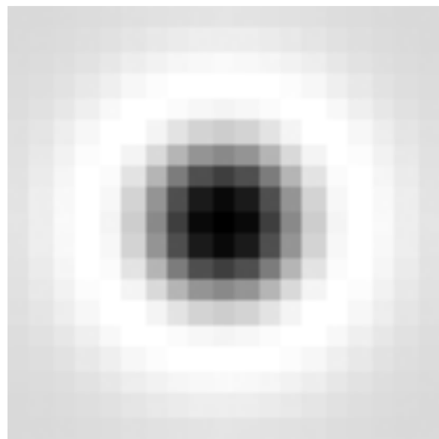
# Effect of scale normalization



# Blob detection in 2D

- Laplacian of Gaussian: Circularly symmetric operator for blob detection in 2D

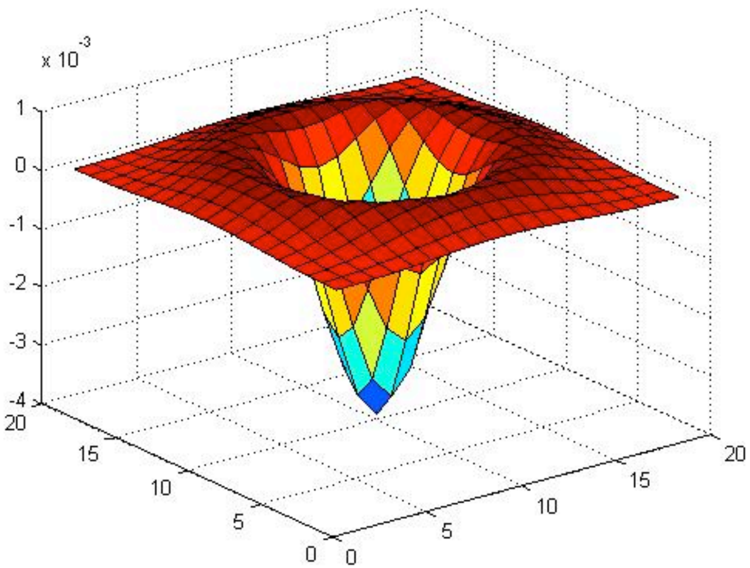


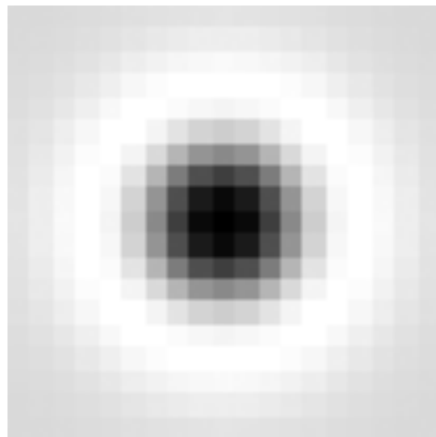



$$\nabla^2 g = \frac{\partial^2 g}{\partial x^2} + \frac{\partial^2 g}{\partial y^2}$$

# Blob detection in 2D

- Laplacian of Gaussian: Circularly symmetric operator for blob detection in 2D



Scale-normalized: $$\nabla^2_{\text{norm}} g = \sigma^2 \left( \frac{\partial^2 g}{\partial x^2} + \frac{\partial^2 g}{\partial y^2} \right)$$

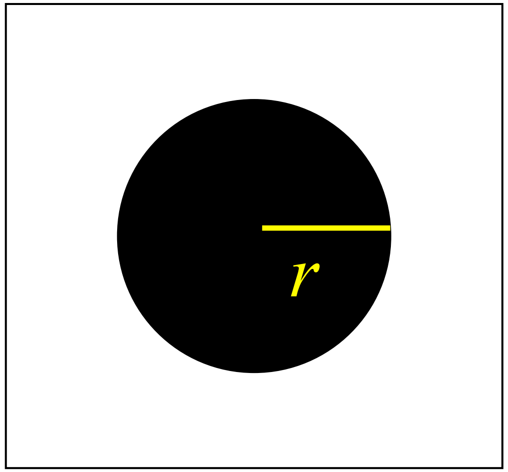
# Scale selection

- The 2D Laplacian is given by

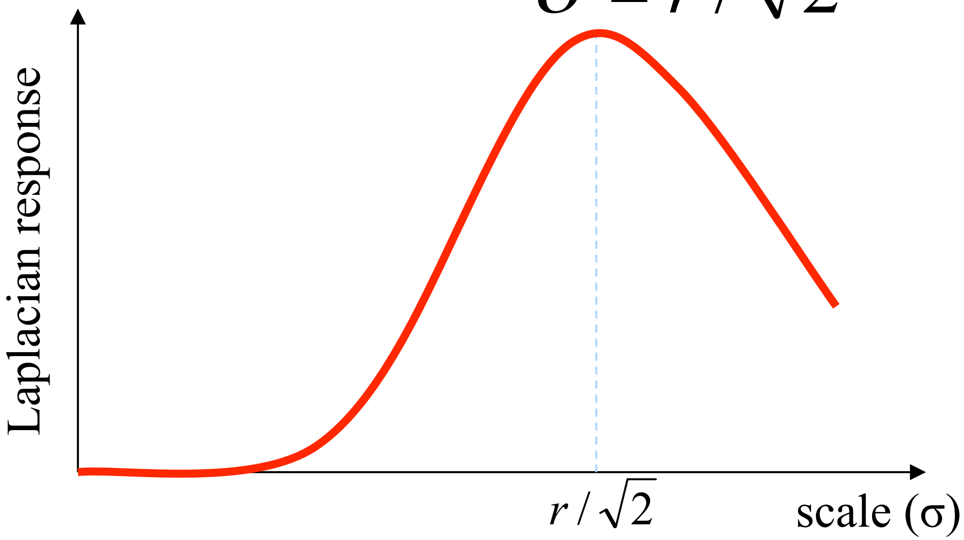
$$(x^2 + y^2 - 2\sigma^2)\, e^{-(x^2+y^2)/2\sigma^2} \quad \text{(up to scale)}$$

- For a binary circle of radius r, the Laplacian achieves a maximum at

$$\sigma = r / \sqrt{2}$$

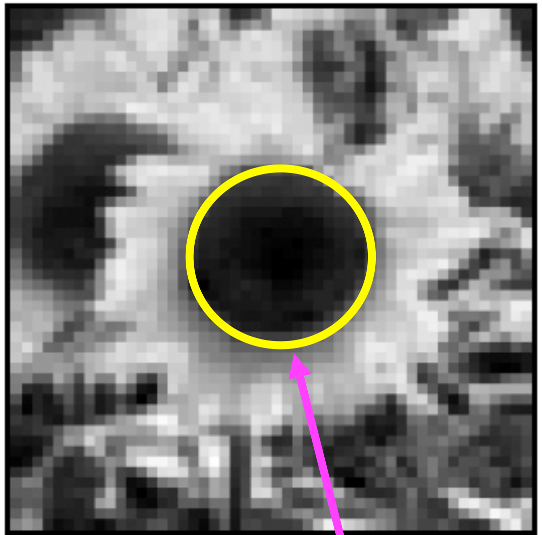


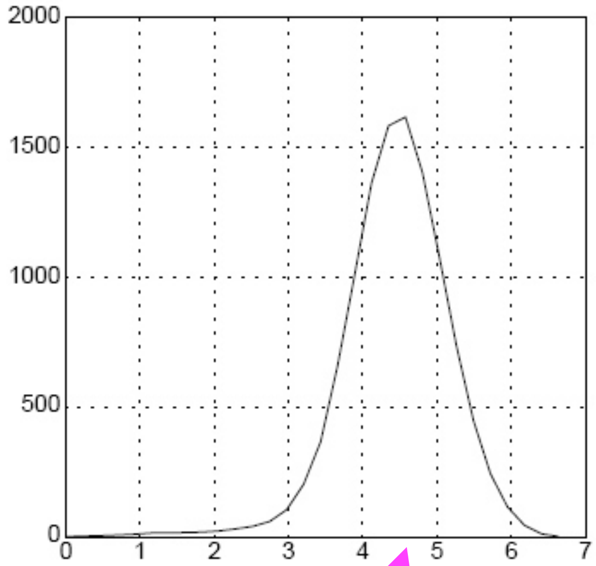

image



# Characteristic scale

- We define the characteristic scale as the scale that produces peak of Laplacian response






characteristic scale

T. Lindeberg (1998). Feature detection with automatic scale selection. *International Journal of Computer Vision* **30** (2): pp 77--116.

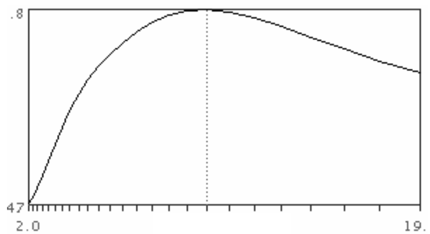
# Scale selection

- For a point compute a value (gradient, Laplacian etc.) at several scales
- Normalization of the values with the scale factor

  e.g. Laplacian $| s^2 (L_{xx} + L_{yy}) |$

- Select scale $s^*$ at the maximum → characteristic scale

$| s^2 (L_{xx} + L_{yy}) |$

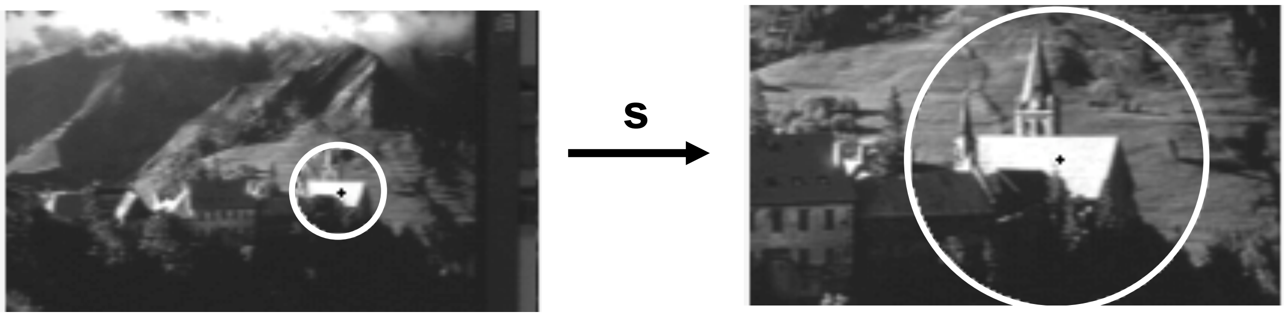


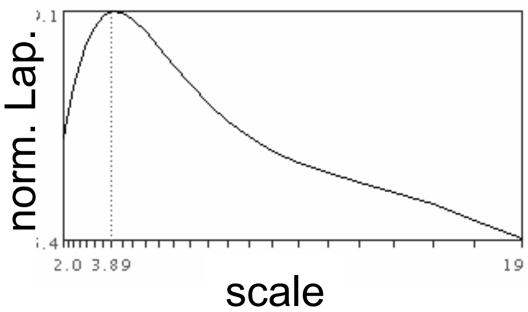

scale

- Exp. results show that the Laplacian gives best results

# Scale selection

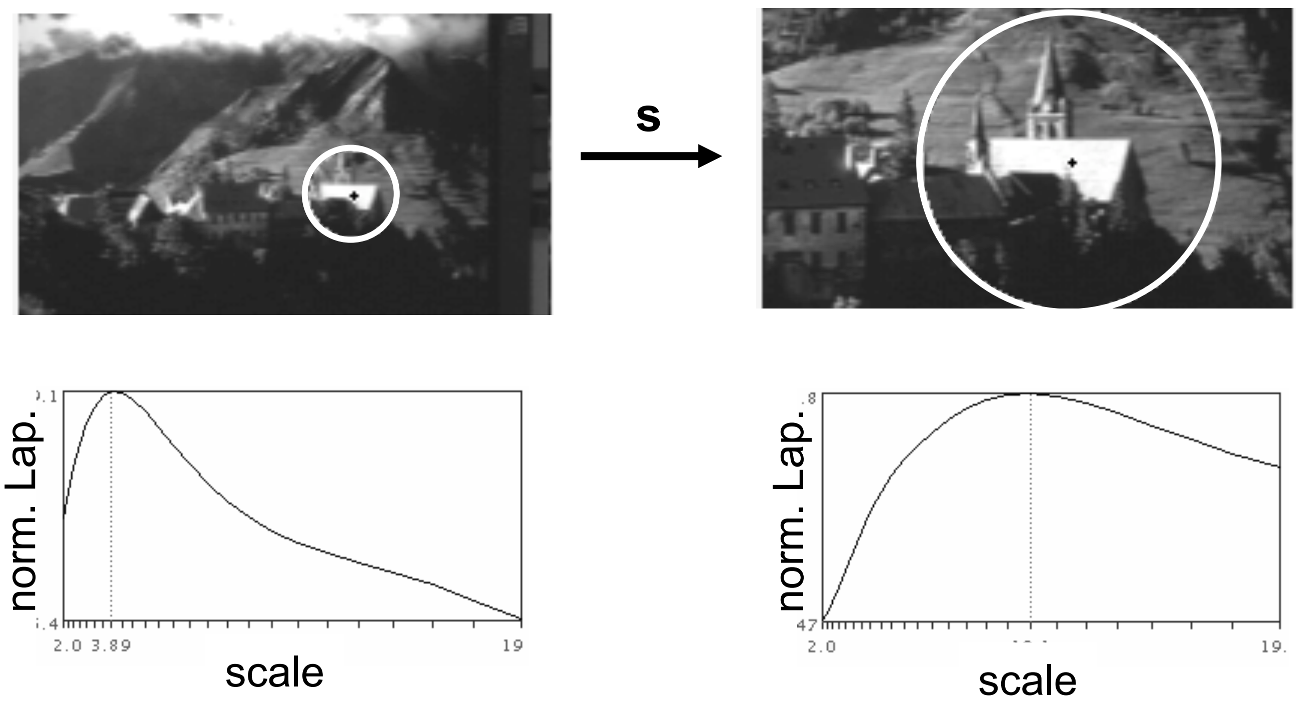
- Scale invariance of the characteristic scale





# Scale selection

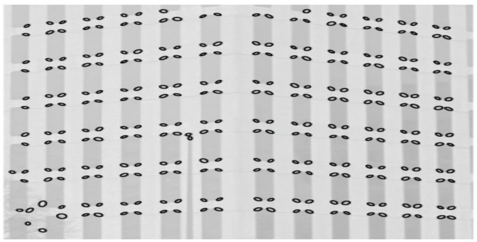
- Scale invariance of the characteristic scale



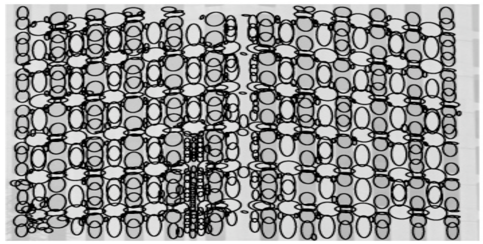

- Relation between characteristic scales $s \cdot s_1^* = s_2^*$

# Scale-invariant detectors

- Harris-Laplace (Mikolajczyk & Schmid'01)
- Laplacian detector (Lindeberg'98)
- Difference of Gaussian (Lowe'99)



Harris-Laplace

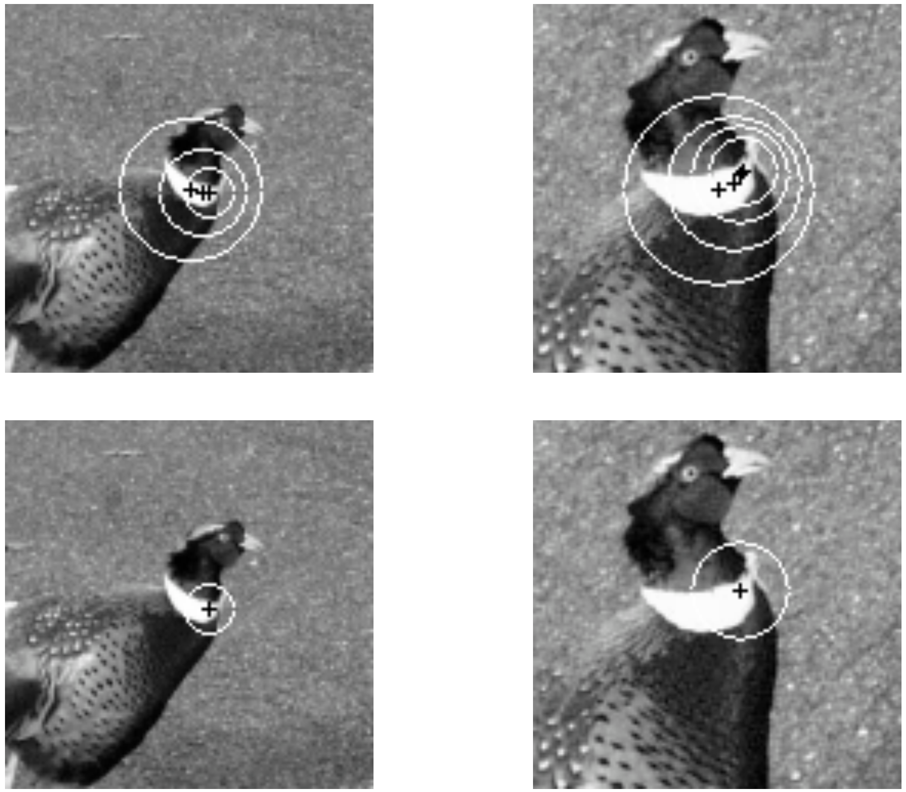


Laplacian

# Harris-Laplace

multi-scale Harris points

selection of points at
maximum of Laplacian



➡ invariant points + associated regions [Mikolajczyk & Schmid'01]

# Matching results

213 / 190 detected interest points

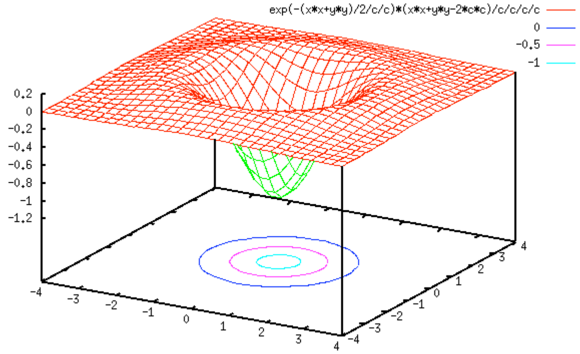
# Matching results

58 points are initially matched

# Matching results

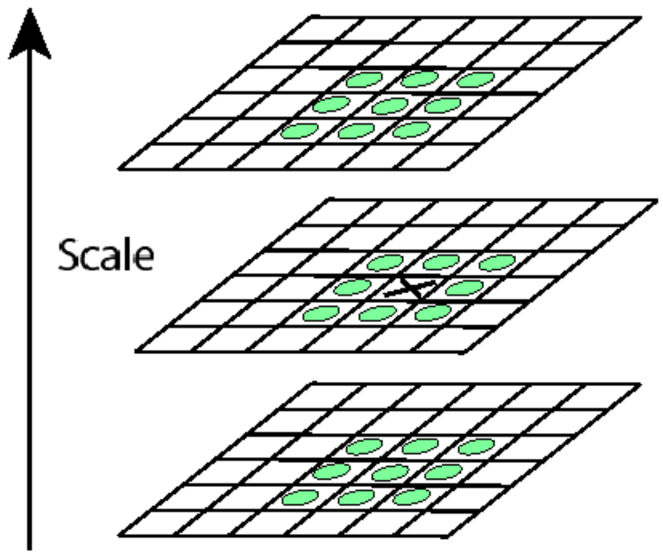
32 points are matched after verification – all correct

# LOG detector




Convolve image with scale-normalized Laplacian at several scales

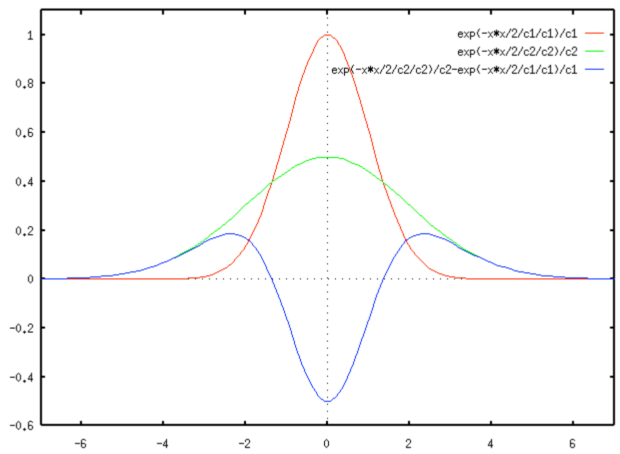
$$LOG = s^2(G_{xx}(\sigma) + G_{yy}(\sigma))$$



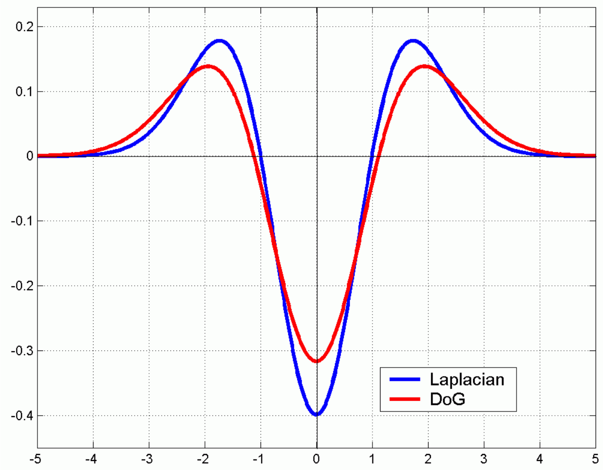

Detection of maxima and minima of Laplacian in scale space

# Efficient implementation

- Difference of Gaussian (DOG) approximates the Laplacian $DOG = G(k\sigma) - G(\sigma)$

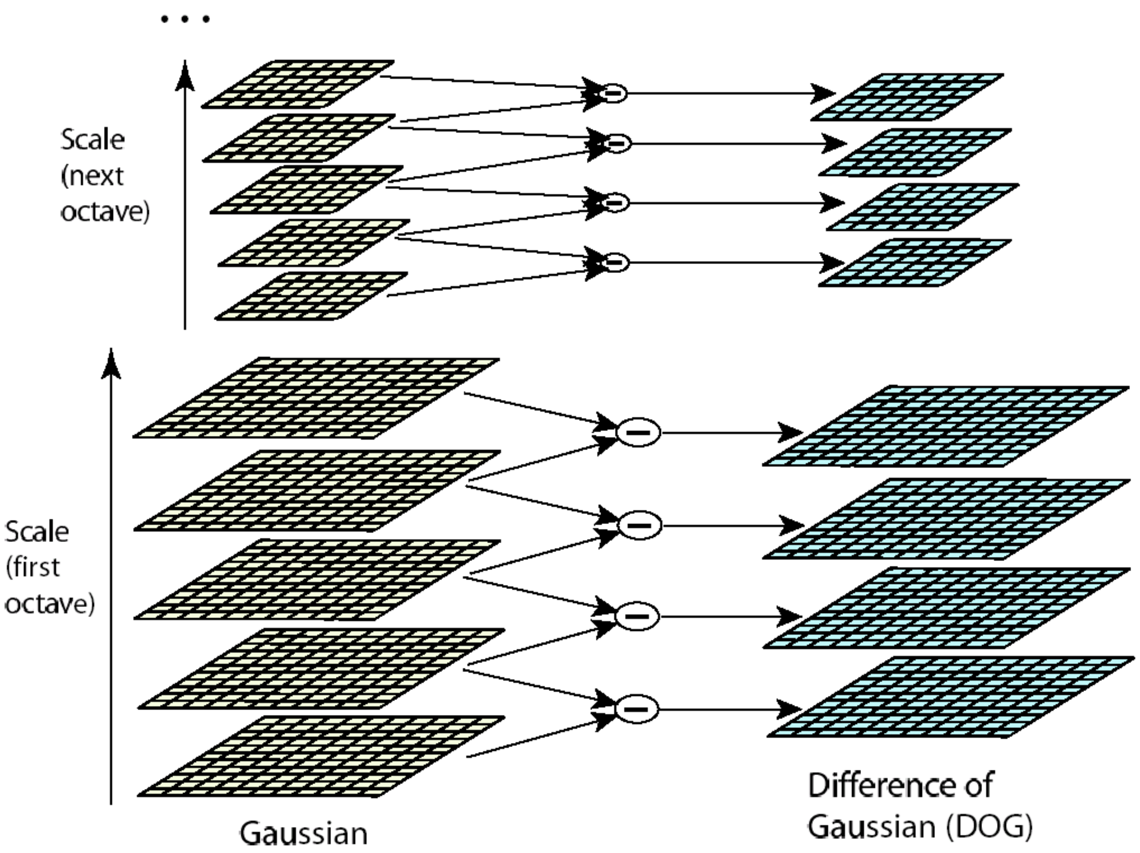



- Error due to the approximation



# DOG detector

- Fast computation, scale space processed one octave at a time




David G. Lowe. "Distinctive image features from scale-invariant keypoints."I*JCV* 60 (2).

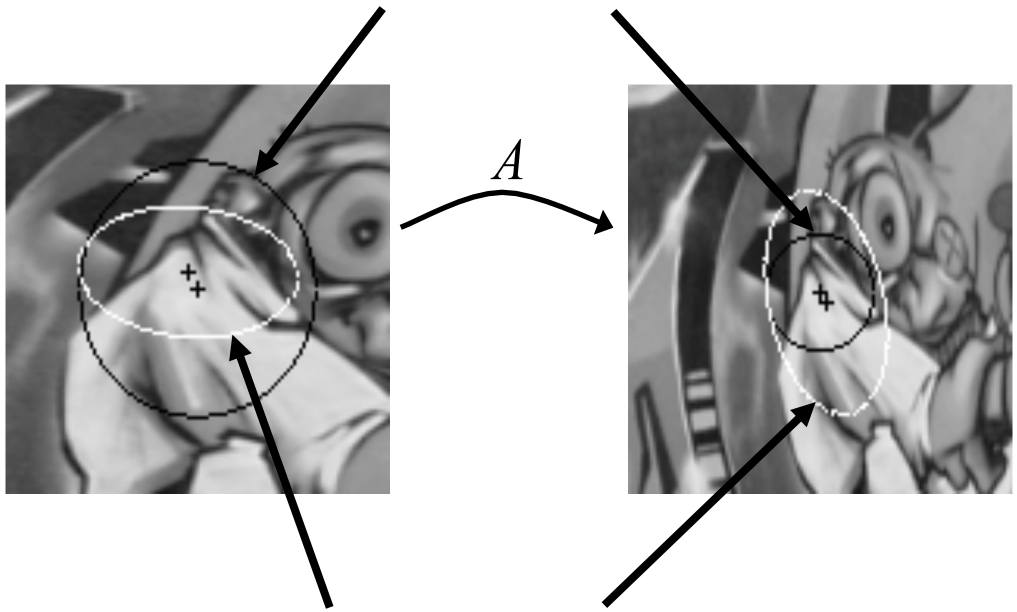
# Local features - overview

- Scale invariant interest points
- ***Affine invariant interest points***
- Evaluation of interest points
- Descriptors and their evaluation

# Affine invariant regions - Motivation

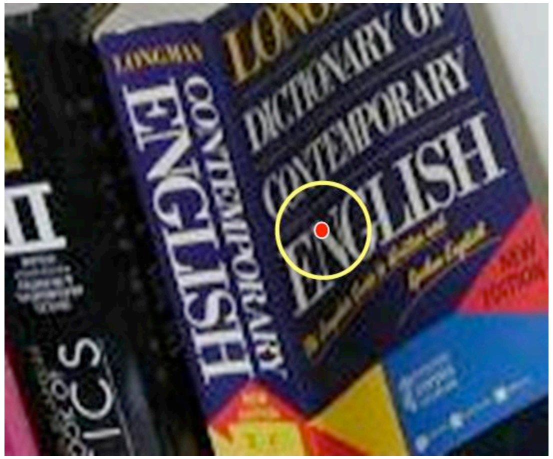
- Scale invariance is not sufficient for large baseline changes

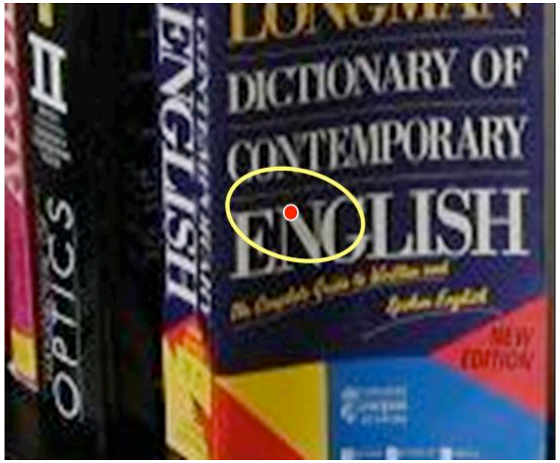
detected scale invariant region




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

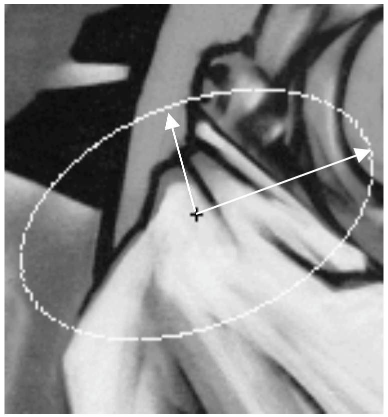
- Initialize with scale-invariant Harris/Hessian/Laplacian points
- Estimation of the affine neighbourhood with the second moment matrix [Lindeberg'94]
- Apply affine neighbourhood estimation to the scale-invariant interest points [Mikolajczyk & Schmid'02, Schaffalitzky & Zisserman'02]
- Excellent results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

# Affine invariant regions

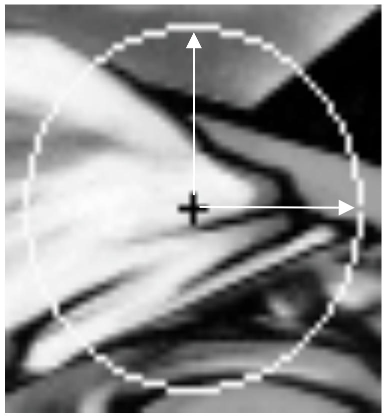
- Based on the second moment matrix (Lindeberg'94)

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

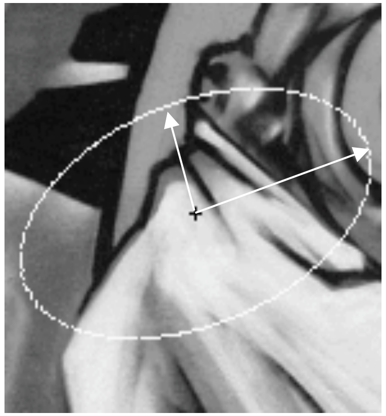
- Normalization with eigenvalues/eigenvectors



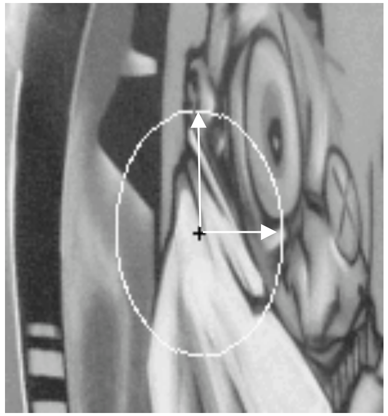
$$\mathrm{x}' = M^{\frac{1}{2}}\mathrm{x}$$



# Affine invariant regions

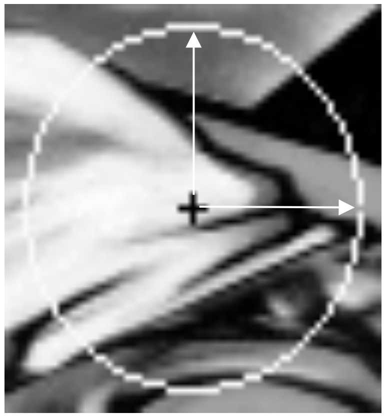


$$\mathrm{x}_\mathrm{R} = A\mathrm{x}_\mathrm{L}$$

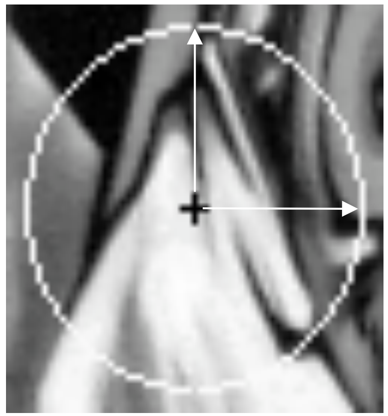


$$\mathrm{x}_\mathrm{L}{}' = M_L^{\frac{1}{2}}\mathrm{x}_\mathrm{L}$$

$$\mathrm{x}_\mathrm{R}{}' = M_R^{\frac{1}{2}}\mathrm{x}_\mathrm{R}$$



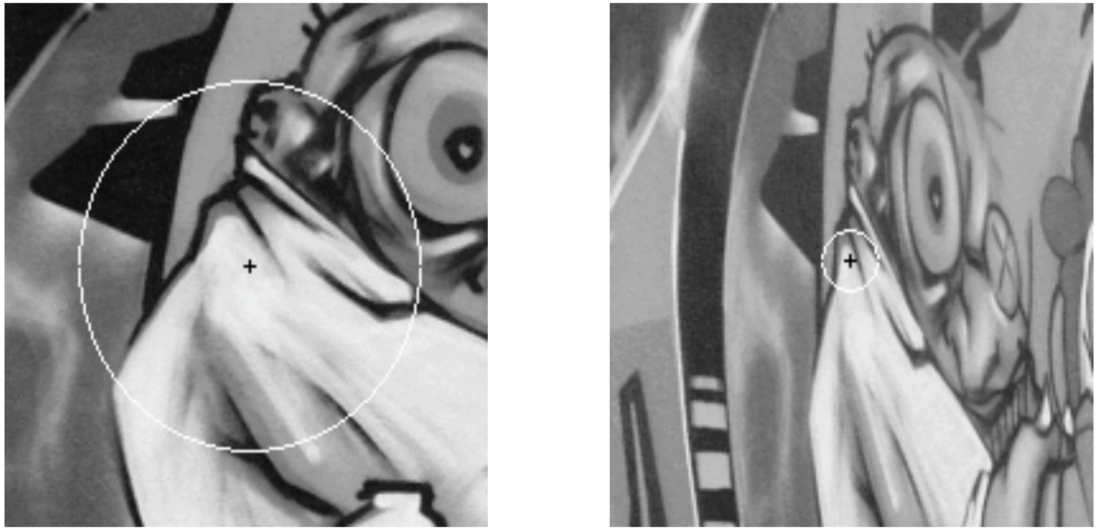
$$\mathrm{x}_R{}' = R\mathrm{x}_L{}'$$



Isotropic neighborhoods related by image rotation

# Affine invariant regions - Estimation

- Iterative estimation – initial points



# Affine invariant regions - Estimation

- Iterative estimation – iteration #1

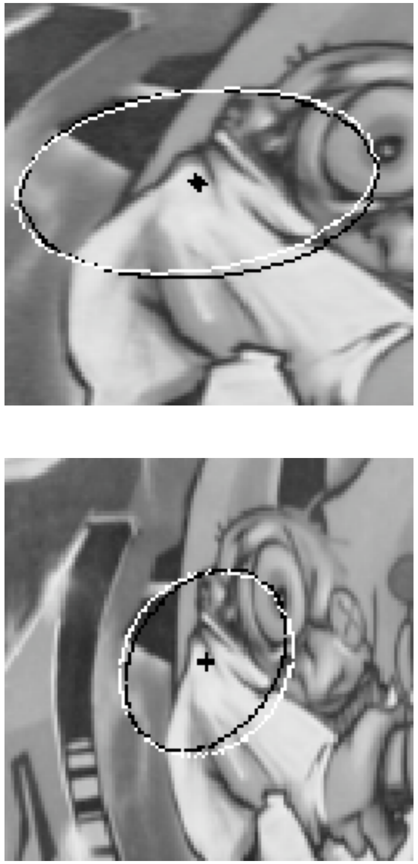
# Affine invariant regions - Estimation

- Iterative estimation – iteration #2

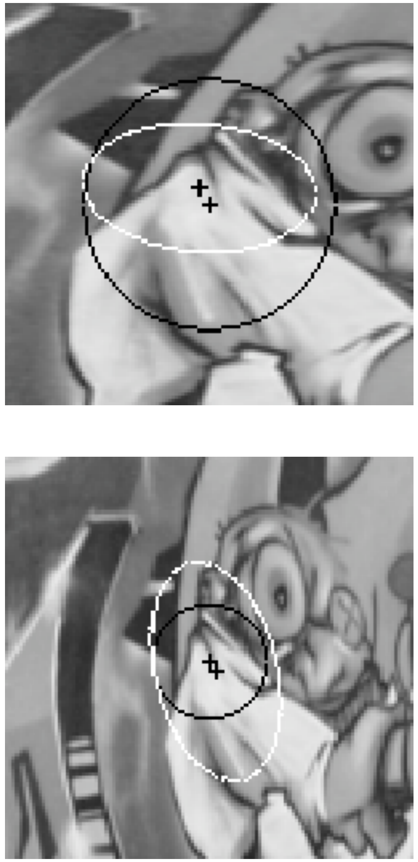

# Affine invariant regions - Estimation

- Iterative estimation – iteration #3, #4

# Harris-Affine versus Harris-Laplace

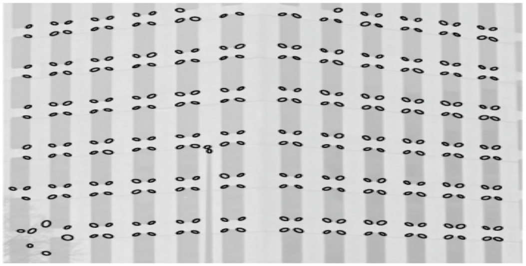


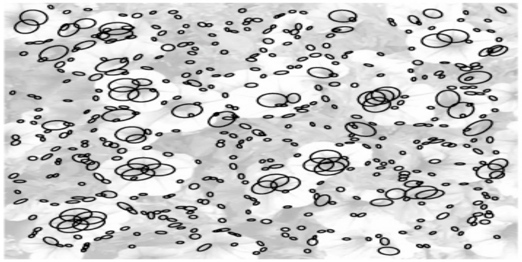
Harris-Affine



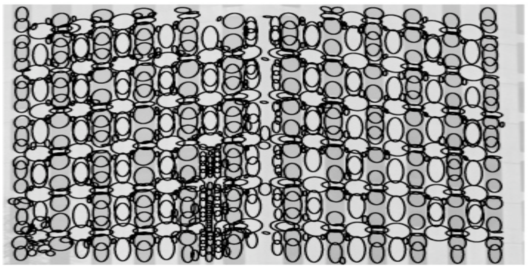
Harris-Laplace

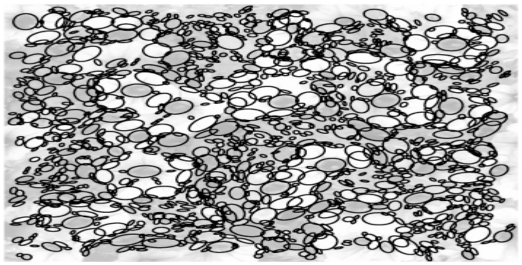
# Harris/Hessian-Affine





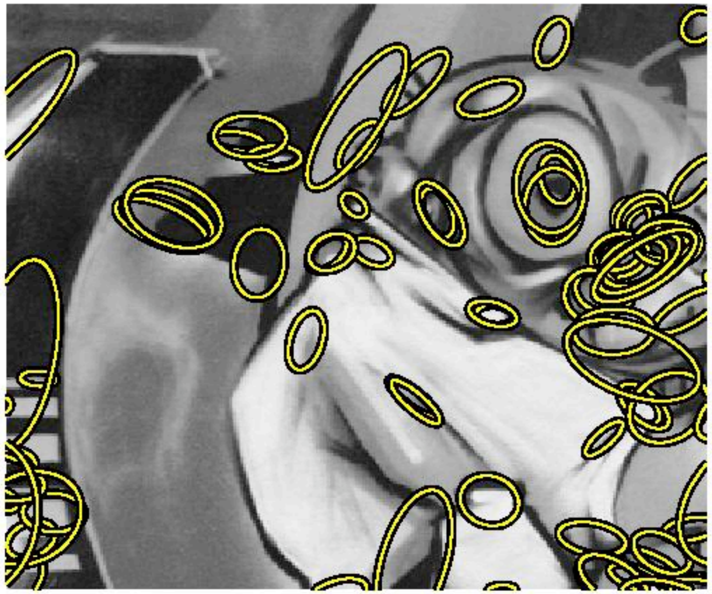
Harris-Affine





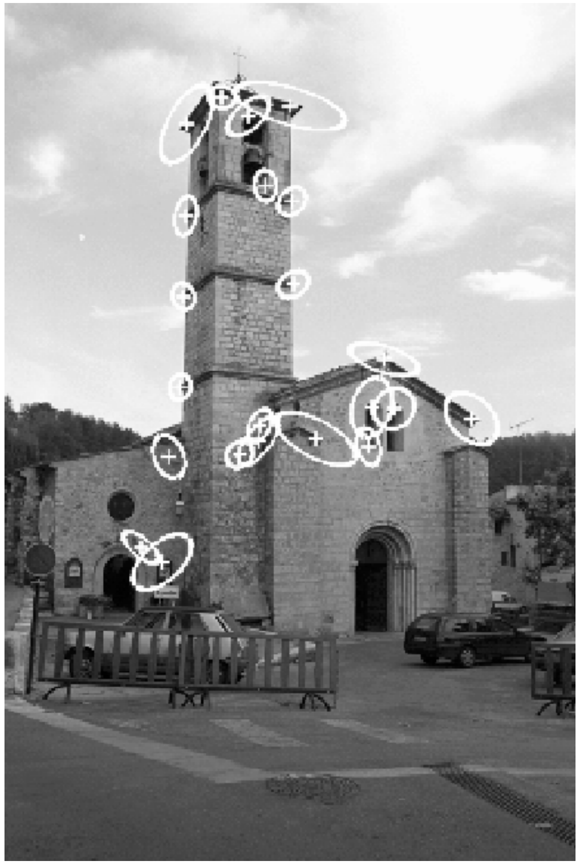
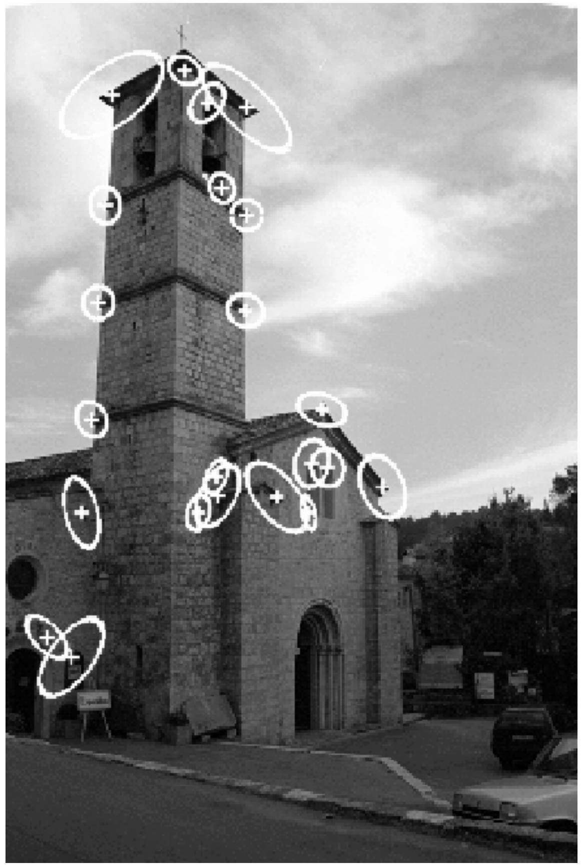
Hessian-Affine

# Harris-Affine



# Hessian-Affine

# Matches

22 correct matches

# Matches

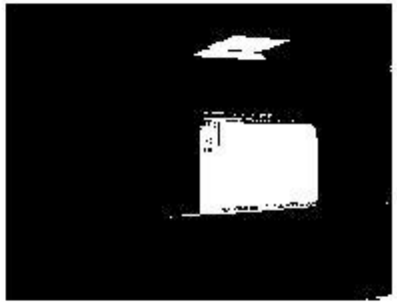
33 correct matches

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

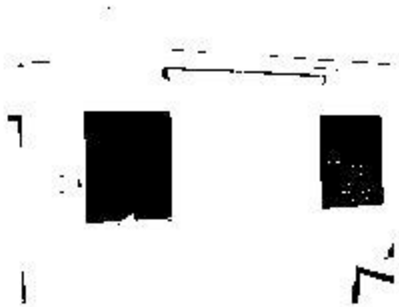
- Extremal regions: connected components in a thresholded image (all pixels above/below a threshold)
- Maximally stable: minimal change of the component (area) for a change of the threshold, i.e. region remains stable for a change of threshold
- Excellent results in a recent comparison

# Maximally stable extremal regions (MSER)

## Examples of thresholded images

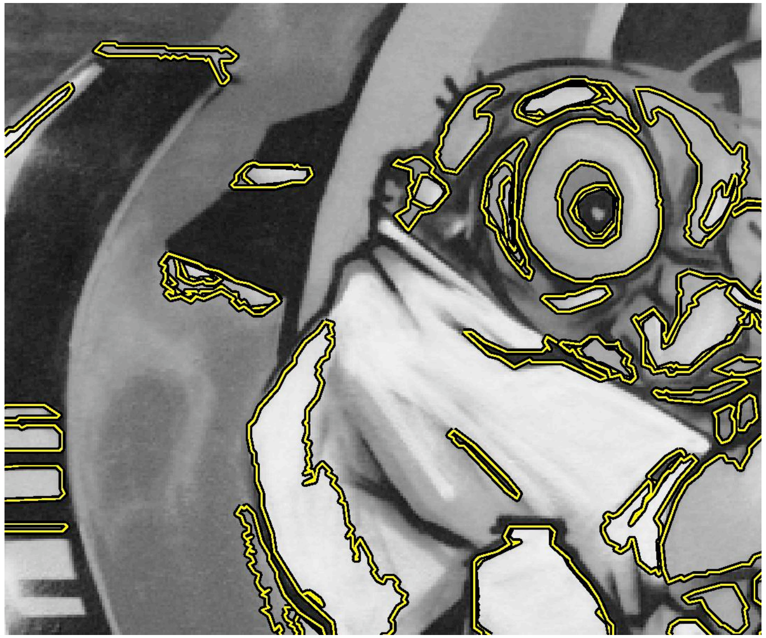


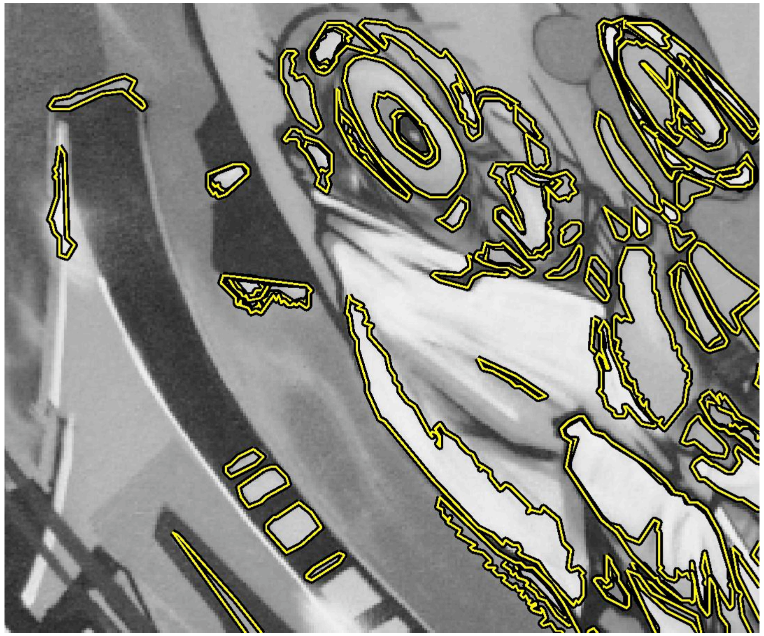
high threshold



low threshold

# MSER





# Overview

- Harris interest points
- Comparing interest points (SSD, ZNCC, SIFT)
- Scale & affine invariant interest points
- **Evaluation and comparison of different detectors**
- Region descriptors and their performance

# Evaluation of interest points

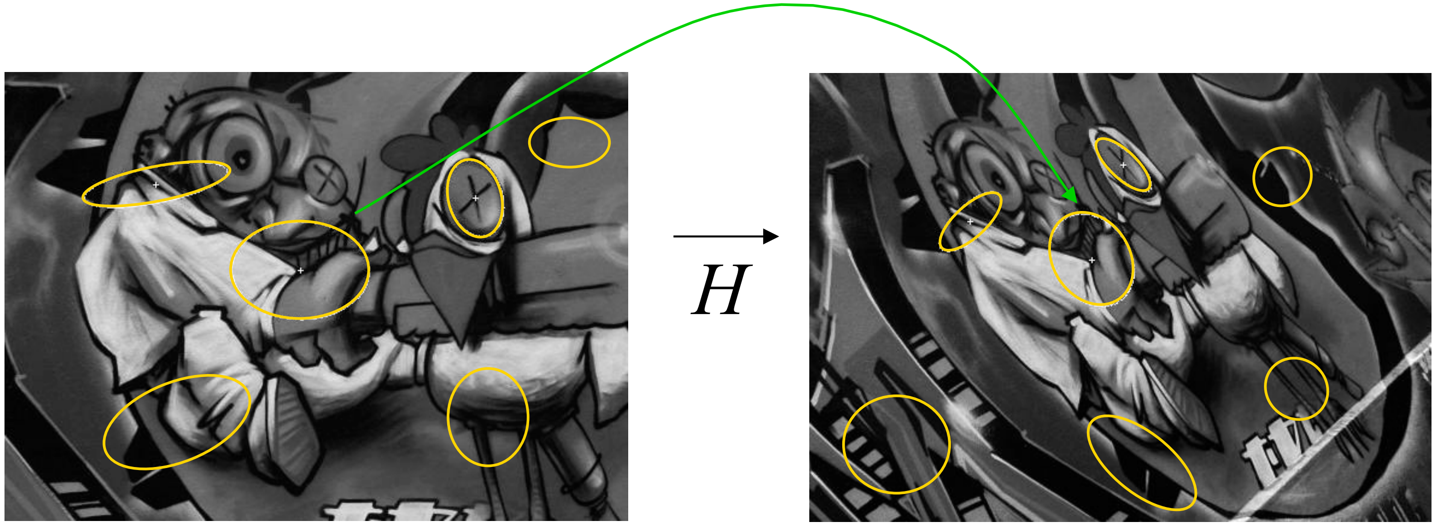
- Quantitative evaluation of interest point/region detectors
  - points / regions at the same relative location and area

- Repeatability rate : percentage of corresponding points

- Two points/regions are corresponding if
  - location error small
  - area intersection large

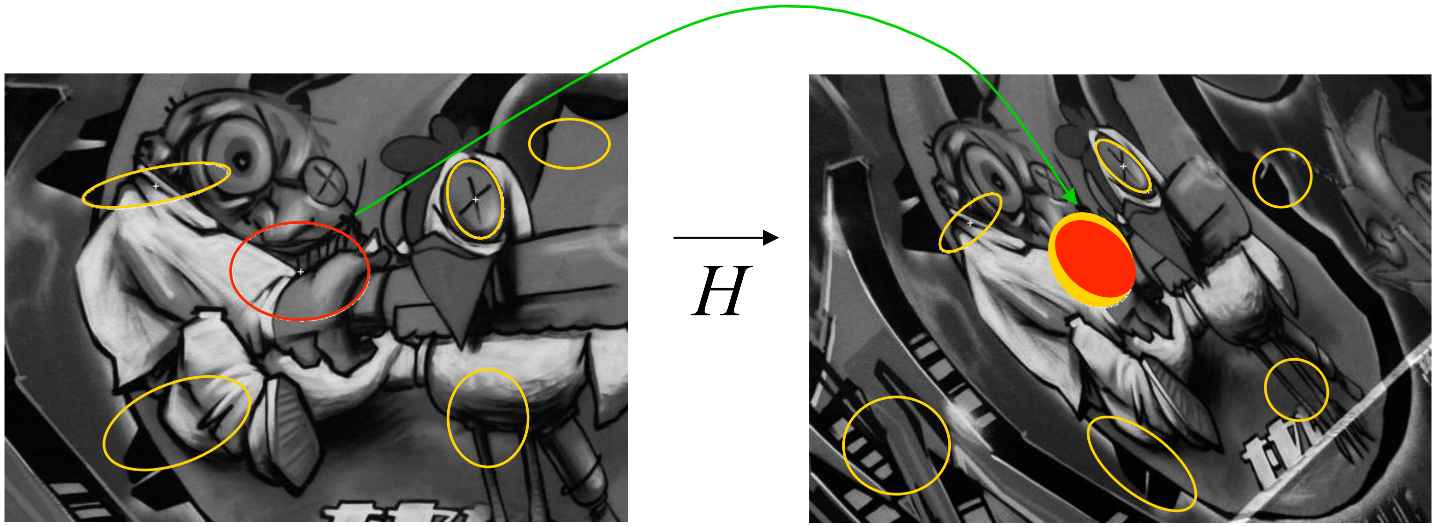
- [K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir & L. Van Gool ’05]

# Evaluation criterion



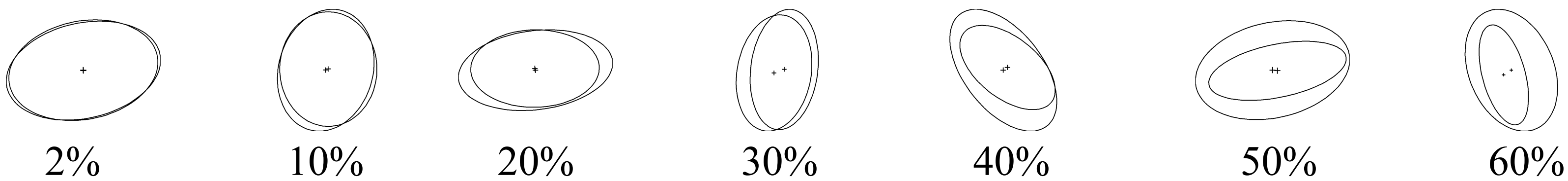
$$repeatability = \frac{\#\, corresponding\ regions}{\#\, detected\ regions} \cdot 100\%$$

# Evaluation criterion




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

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

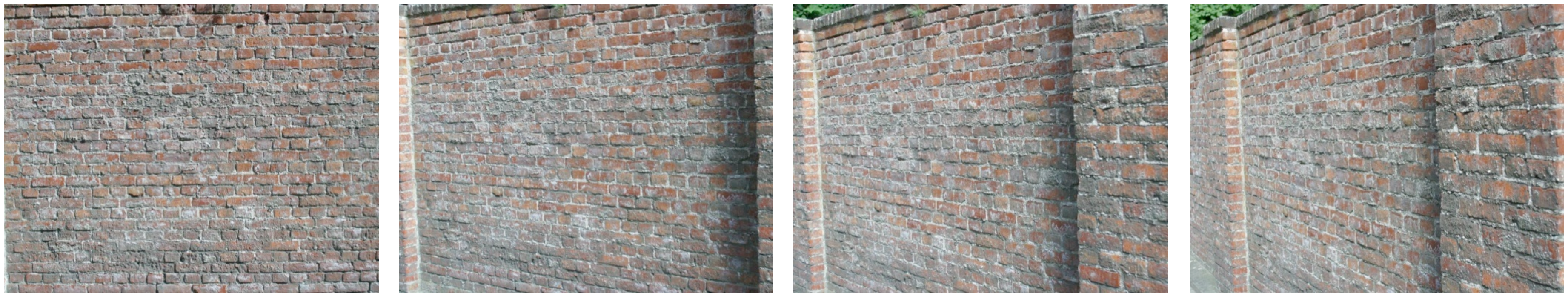


# Dataset

- Different types of transformation
  - Viewpoint change
  - Scale change
  - Image blur
  - JPEG compression
  - Light change
- Two scene types
  - Structured
  - Textured
- Transformations within the sequence (homographies)
  - Independent estimation

# Viewpoint change (0-60 degrees )

structured scene



textured scene

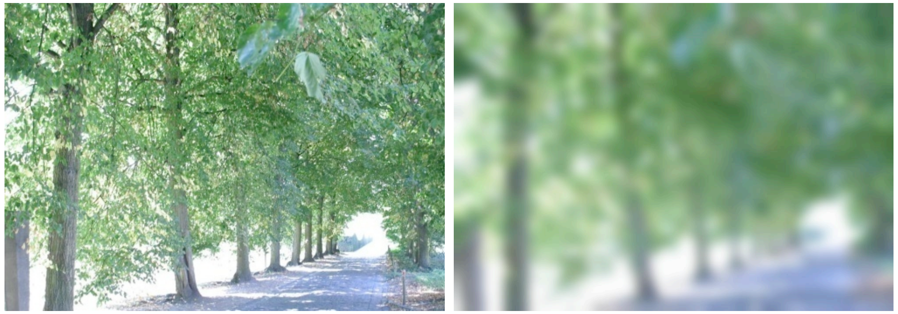
# Zoom + rotation (zoom of 1-4)

structured scene

textured scene

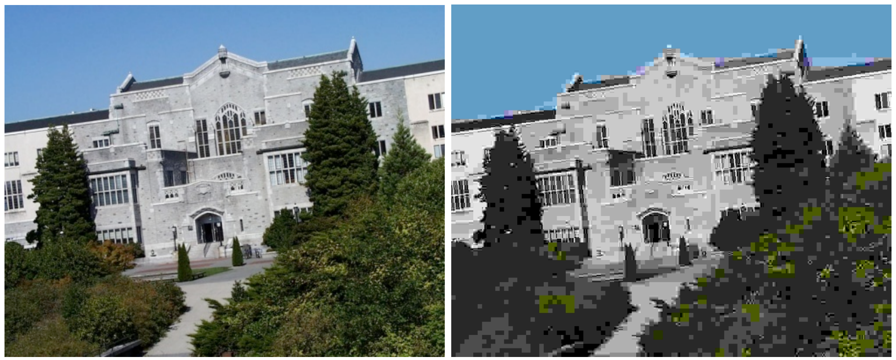
# Blur, compression, illumination

blur - structured scene



blur - textured scene

light change - structured scene

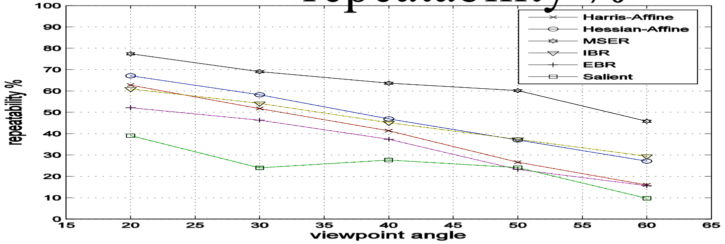


jpeg compression - structured scene

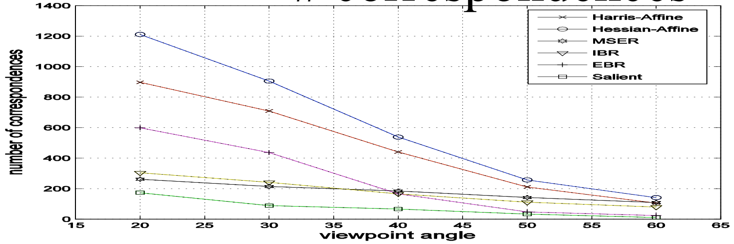
# Comparison of affine invariant detectors

## Viewpoint change - structured scene

repeatability %




# correspondences




reference image

20

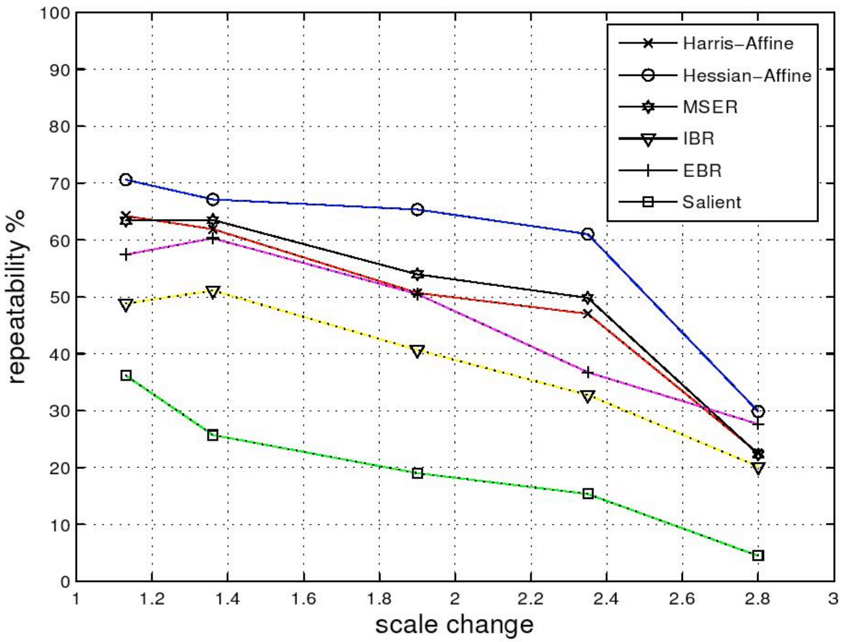
40

60

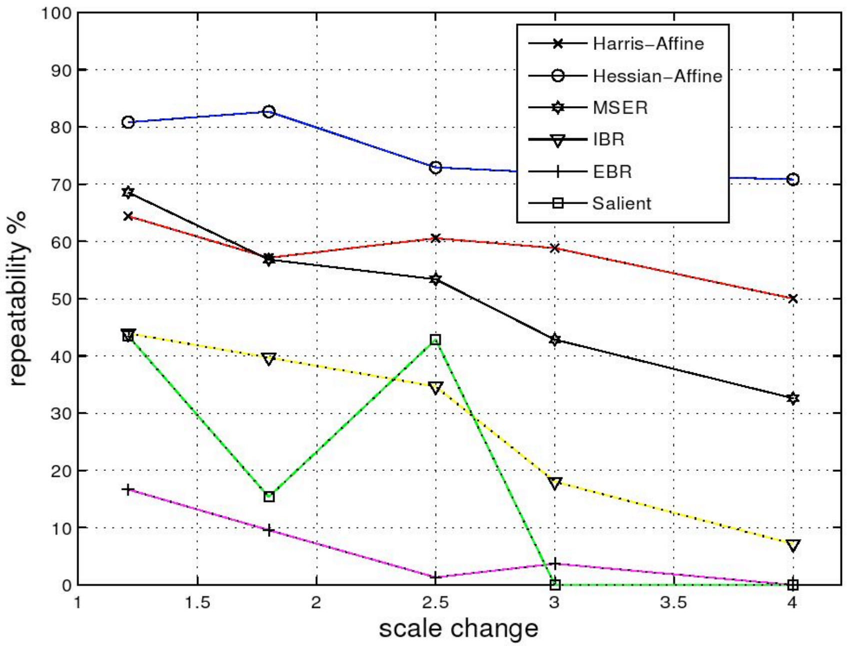
# Comparison of affine invariant detectors

## Scale change

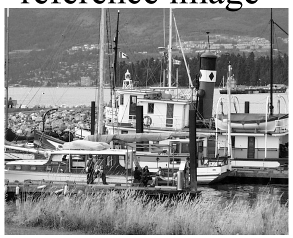
repeatability %




repeatability %




reference image



2.8

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

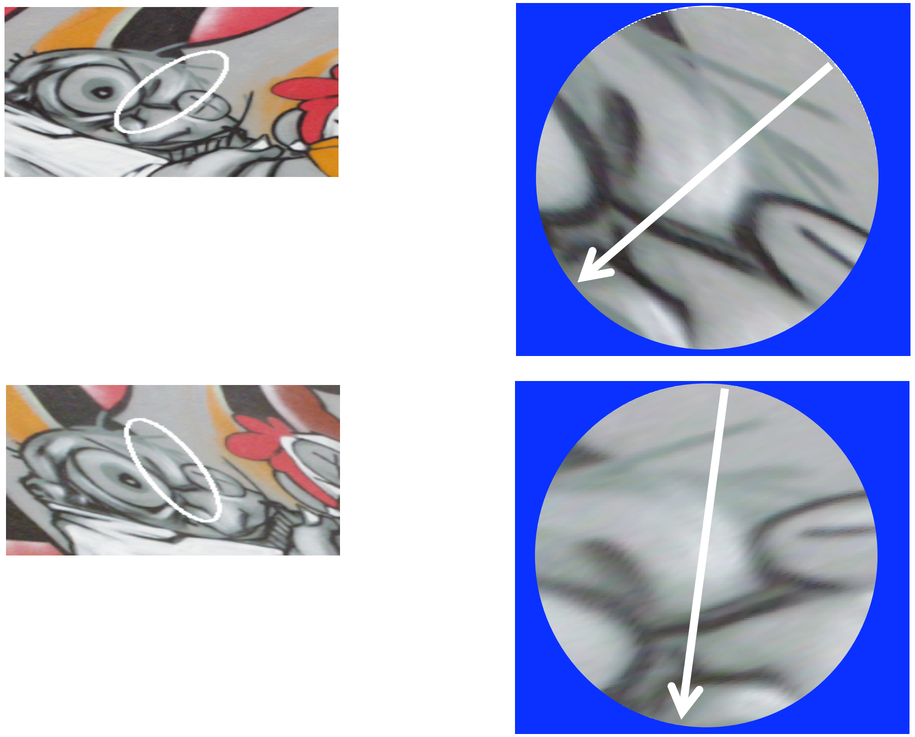
- Performance of the different scale invariant detectors is very similar (Harris-Laplace, Hessian, LoG and DOG)

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

# Overview

- Harris interest points
- Comparing interest points (SSD, ZNCC, SIFT)
- Scale & affine invariant interest points
- Evaluation and comparison of different detectors
- **Region descriptors and their performance**

# Region descriptors



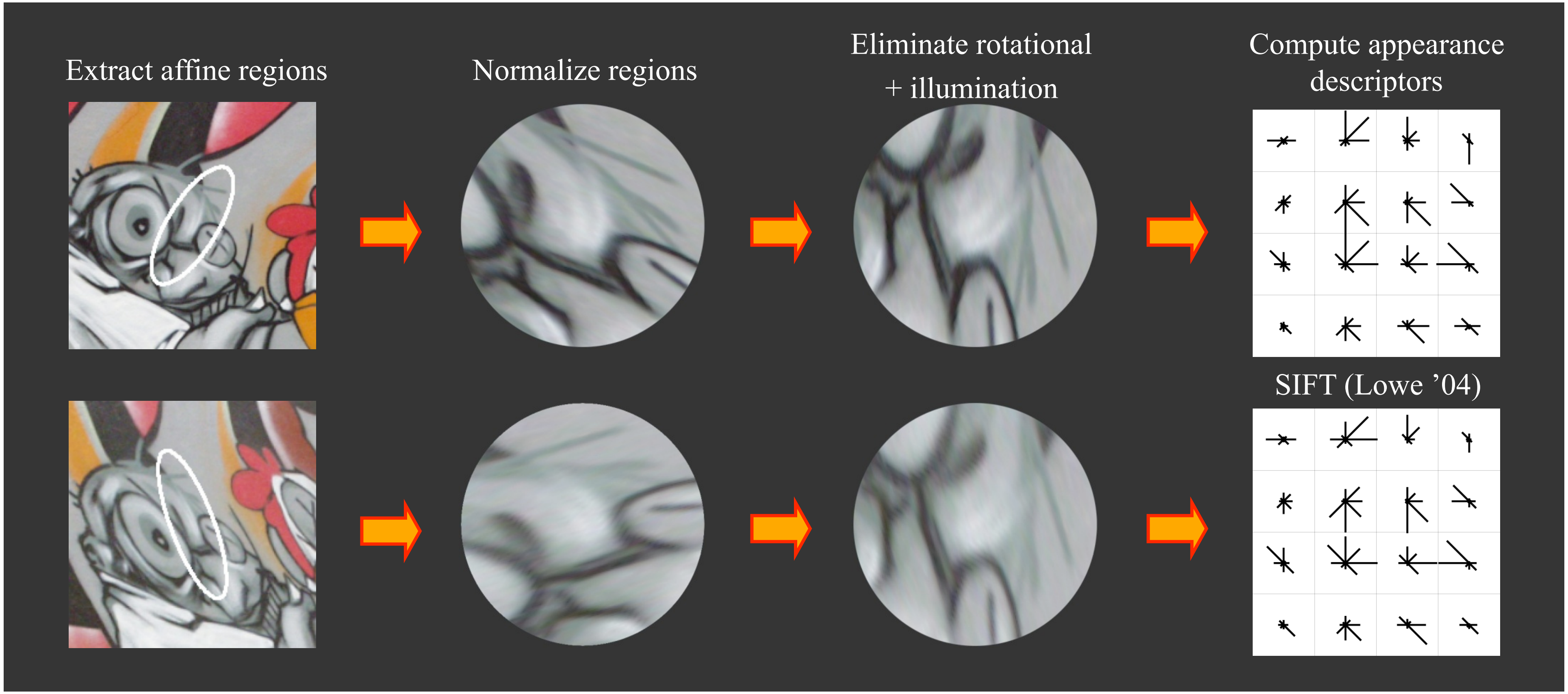
- Normalized regions are
  - invariant to geometric transformations except rotation
  - not invariant to photometric transformations

# Descriptors

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

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

# Descriptors



# Descriptors

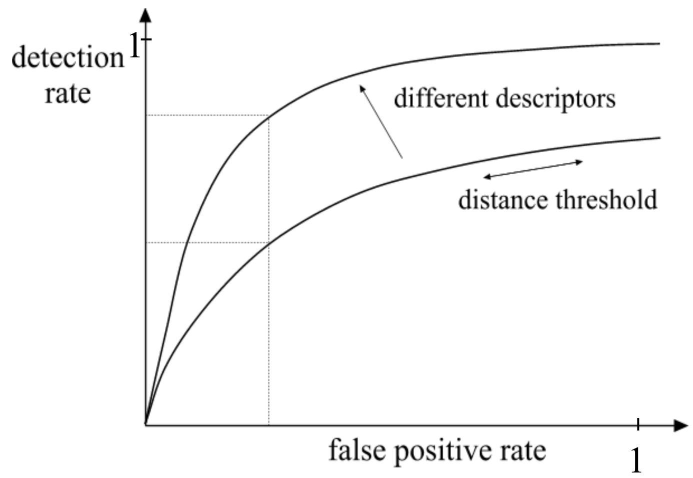
- Gaussian derivative-based descriptors
  - Differential invariants (*Koenderink and van Doorn'87*)
  - Steerable filters (*Freeman and Adelson'91*)

- Moment invariants [Van Gool et al.'96]
- SIFT (*Lowe'99)*
- 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

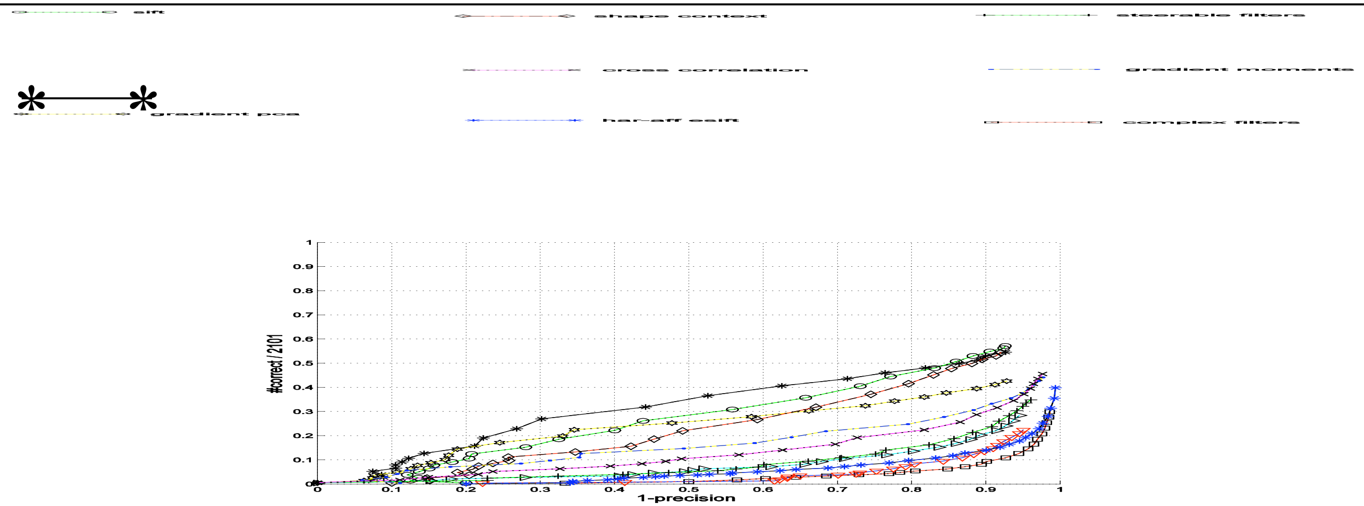
- Descriptors should be
  - Distinctive
  - Robust to changes on viewing conditions as well as to errors of the detector

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

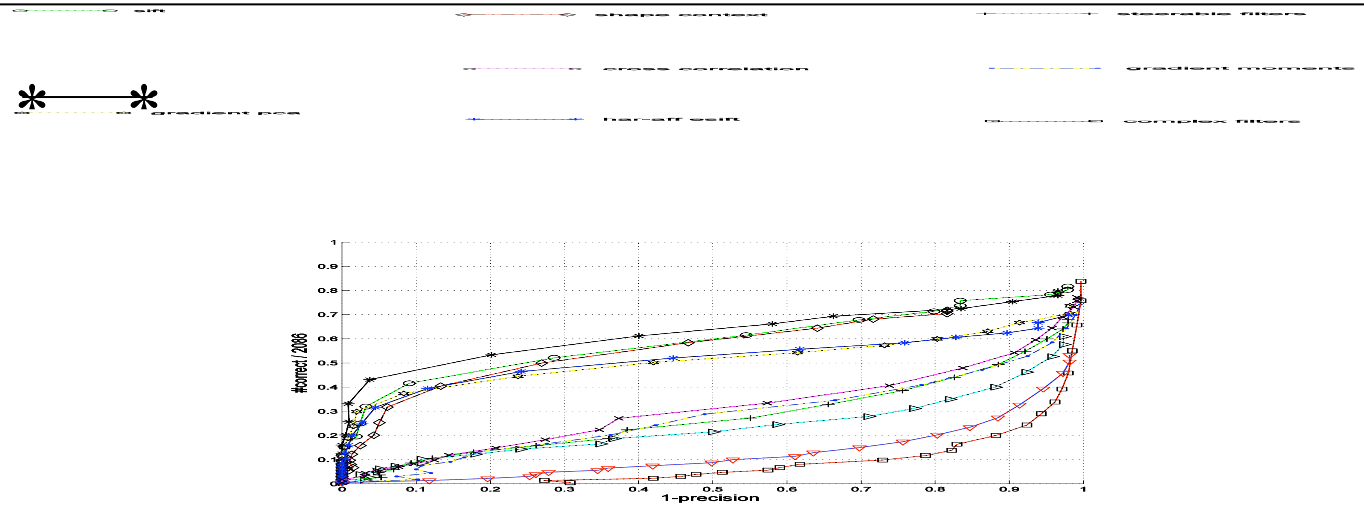



[K. Mikolajczyk & C. Schmid, PAMI'05]

# 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

http://lear.inrialpes.fr/software

- 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*