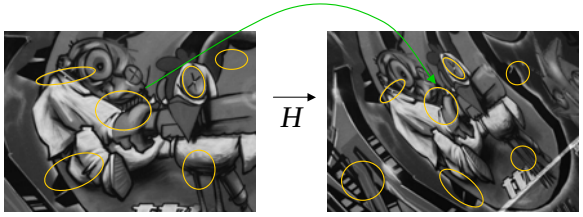
Evaluation and comparison of interest points/regions

Introduction

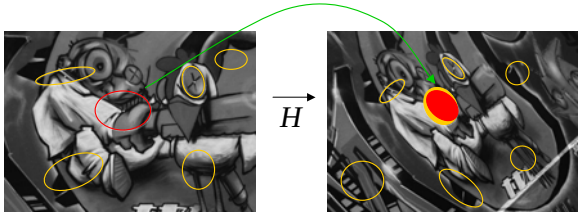
- Quantitative evaluation of interest point/region detectors
  - points / regions at the same relative location and area
- Repeatability rate : percentage of corresponding points
- Two points/regions are corresponding if
  - location error small
  - area intersection large

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

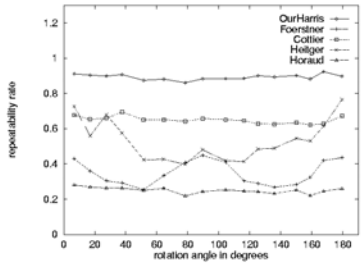
blur – textured scene

light change – structured scene

jpeg compression – structured scene

Comparison of different detectors

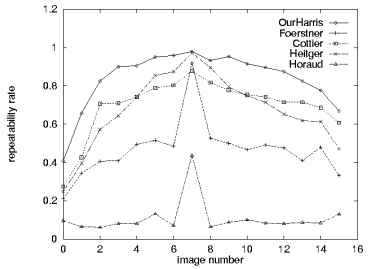
repeatability - image rotation



[Comparing and Evaluating Interest Points, Schmid, Mohr & Bauckhage, ICCV 98]

Comparison of different detectors

repeatability – perspective transformation



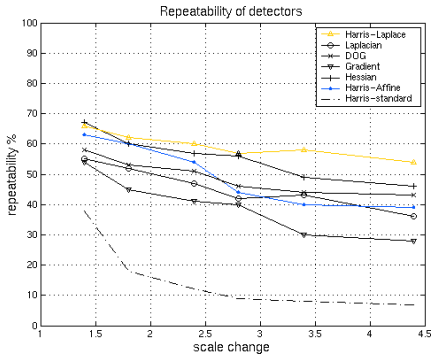
[Comparing and Evaluating Interest Points, Schmid, Mohr & Bauckhage, ICCV 98]

Conclusion – different detectors

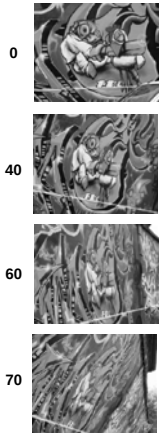
- Harris – best performance, more accurate than other detectors
- Biologically inspired detector performs worse
- Edge based detector gives low performance
  - inaccuracy of line detection + intersection decrease performance

Comparison of scale invariant detectors

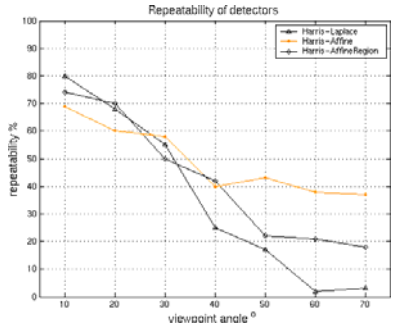
repeatability – scale changes



Evaluation of an affine invariant detector



repeatability – perspective transformation

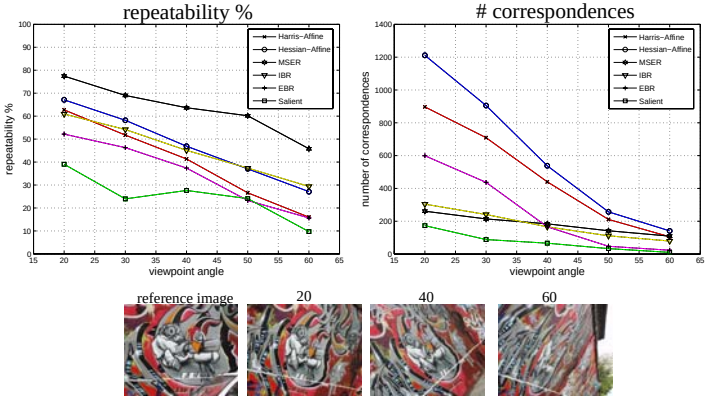


Conclusion – scale-invariant detectors

- Harris-Laplace, Hessian-Laplace, LoG and DOG give good results
- Scale-invariant detector sufficient up to 40 degrees of viewpoint change

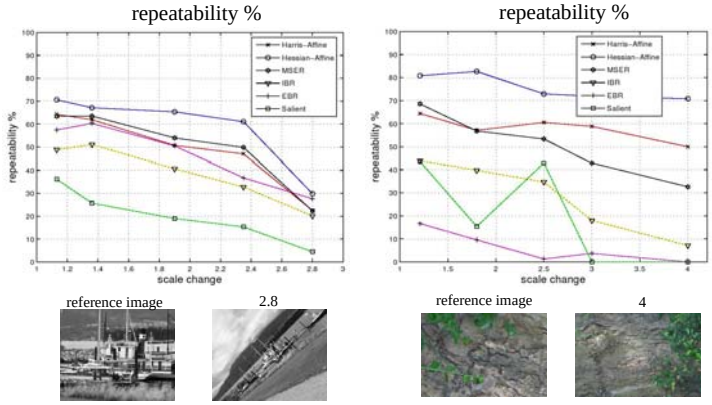
Comparison of affine invariant detectors

Viewpoint change - structured scene

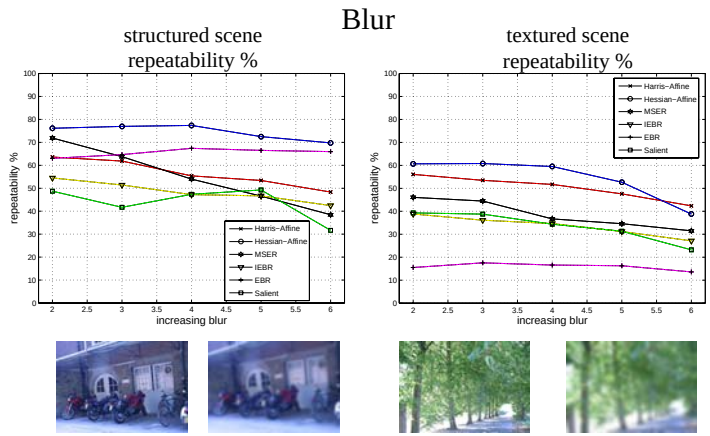


Comparison of affine invariant detectors

Scale change



Comparison of affine invariant detectors



Conclusion - affine invariant detectors

- MSER – best performance, more accurate than other detectors
- Hessian-Affine – second best
- Harris-Affine and IBR – average
- Edge based regions fail for texture scenes
- Salient regions – low performance
- Hessian-Affine and Harris-Affine – provide more regions than other detectors

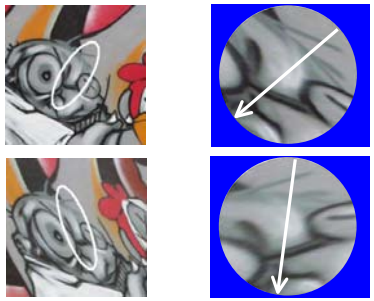
Conclusion - affine invariant detectors

- Good performance for large viewpoint and scale changes
- Results depend on transformation and scene type, no one best detector
- Detectors are complementary
  - MSER and EBR adapted to structured scenes
  - Harris-Affine and Hessian-Affine adapted to textured scenes

[A comparison of affine region detectors, K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir and L. Van Gool, IJCV'05]

Region descriptors and their performance

Region descriptors



- 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

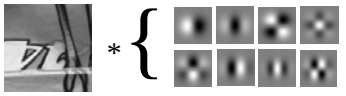
- Sampled image patch
  - descriptor dimension is 81

- Moment invariants (*Van Gool'96*)
  - descriptor dimension is 10

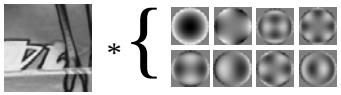
$$\rightarrow \iint_{x,y} x^p y^q [I(x,y)]^a \, dxdy$$

Descriptors

- Gaussian derivative-based descriptors
  - Differential invariants (*Koenderink and van Doorn'87*) (dim. 8)
  - Steerable filters (*Freeman and Adelson'91*) (dim. 13)

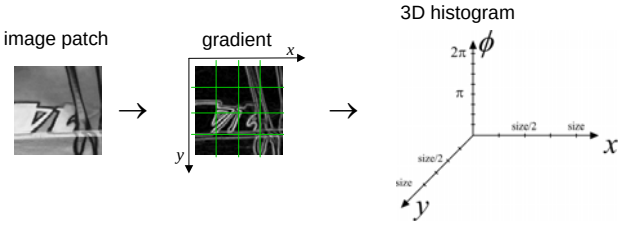


- Complex filters (*Baumberg'00, Schaffalitzky and Zisserman'02 - code*)
  - modulus of complex filter responses (dim. 15)



Descriptors

- SIFT (*Lowe'99 - code*)
  - 8 orientations of the gradient (dim. 128)
  - 4x4 spatial grid
  - normalisation of the descriptor to norm one

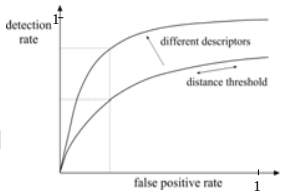


Descriptors

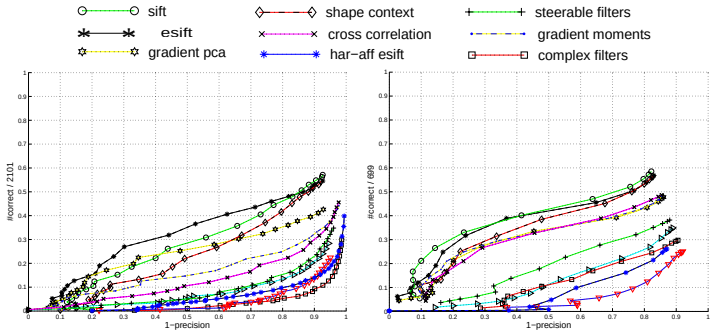
- Shape context (*Belongie et al.'02*)
- Extended SIFT, SIFT with PCA dimensionality reduction
-
- Gradient PCA (*Ke and Sukthankar'04*)

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



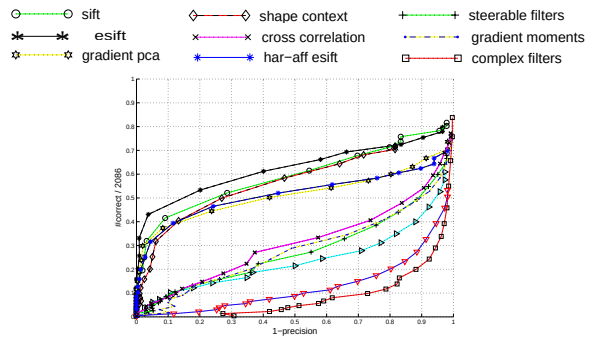
Viewpoint change (60 degrees)



Similarity matching

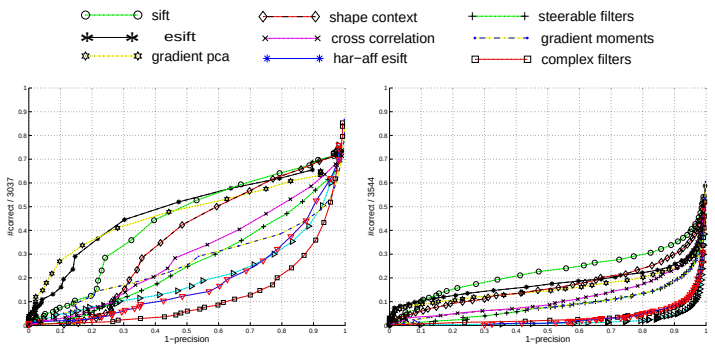
Nearest neighbor ratio matching

Scale change (factor 2.8)



Similarity matching

Image blur



Conclusion - descriptors

- Performance of the descriptor is relatively independent of the detector
- Results similar for different matching strategies
- Dimension can be chosen optimally
- Region overlap does not affect ranking, but higher recall for small overlap errors

Conclusion - descriptors

- SIFT based descriptors perform best (high dimensional) ESIFT > SIFT > shape context
- Low dimensional descriptors : good results for gradient moment and steerable filters
- Cross-correlation gives unstable results
- Robust region descriptors better than point-wise descriptors

[A performance evaluation of local descriptors, K. Mikolajczyk and C. Schmid, CVPR'03 and PAMI'05]

Conclusion

- A large set of good region detectors and descriptors exist
  - extensions are possible, for example to deal with shape
- Good performance for recognizing an object/scene observed under different viewpoints and in a different environment
  - invariance, occlusion, clutter
  - evaluation criteria tuned to this context
- Well adapted for object categorization?
  - good building blocks?
  - design of an appropriate model
  - Example : Fergus, Schiele, Lazebnik

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*
- Reports on the detector & descriptor evaluation