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
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)

\[ M = \mu(x, \sigma_x, \sigma_y) = \sigma_x^2 G(\sigma_x) \otimes \begin{bmatrix} L_x^2(x, \sigma_x) & L_x L_y(x, \sigma_y) \\ L_x L_y(x, \sigma_y) & L_y^2(x, \sigma_y) \end{bmatrix} \]

- Normalization with eigenvalues/eigenvectors

\[ x' = \frac{1}{2} M^{\frac{1}{2}} x \]
Affine invariant regions

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/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

Evaluation criterion

\[ \text{repeatability} = \frac{\text{#corresponding regions}}{\text{#detected regions}} \times 100\% \]
Evaluation criterion

\[ \text{repeatability} = \frac{\# \text{corresponding regions}}{\# \text{detected regions}} \cdot 100\% \]

\[ \text{overlap error} = (1 - \frac{\text{intersection}}{\text{union}}) \cdot 100\% \]
Comparison of affine invariant detectors

Viewpoint change - structured scene

<table>
<thead>
<tr>
<th>viewpoint angle</th>
<th>repeatability %</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>100</td>
</tr>
<tr>
<td>20</td>
<td>90</td>
</tr>
<tr>
<td>25</td>
<td>80</td>
</tr>
<tr>
<td>30</td>
<td>70</td>
</tr>
<tr>
<td>35</td>
<td>60</td>
</tr>
<tr>
<td>40</td>
<td>50</td>
</tr>
<tr>
<td>45</td>
<td>40</td>
</tr>
<tr>
<td>50</td>
<td>30</td>
</tr>
<tr>
<td>55</td>
<td>20</td>
</tr>
<tr>
<td>60</td>
<td>10</td>
</tr>
</tbody>
</table>

reference image: [Images of structured scenes at different viewpoint angles]
Comparison of affine invariant detectors

Scale change – textured scene

![Graph showing repeatability percentage for different detectors with scale change on the x-axis and repeatability percentage on the y-axis. The graph includes lines for Harris-Affine, Hessian-Affine, MSER, IBR, EBR, and Saliency. The reference image is shown at the bottom left.](image-url)
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

1. Extract affine regions
2. Normalize regions
3. Eliminate rotational + illumination
4. Compute appearance descriptors

SIFT (Lowe '04)
Descriptors

- Gaussian derivative-based descriptors
  - Differential invariants \((\text{Koenderink and van Doorn'87})\)
  - Steerable filters \((\text{Freeman and Adelson'91})\)
- SIFT \((\text{Lowe'99})\)
- Moment invariants \([\text{Van Gool et al.'96}]\)
- Shape context \([\text{Belongie et al.'02}]\)
- SIFT with PCA dimensionality reduction
- Gradient PCA \([\text{Ke and Sukthankar'04}]\)
- SURF descriptor \([\text{Bay et al.'08}]\)
- DAISY descriptor \([\text{Tola et al.'08, Windler et al'09}]\)
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 \((d1, d2) < \text{threshold}\)

[K. Mikolajczyk & C. Schmid, PAMI’05]
Viewpoint change (60 degrees)

- sift
- esift
- gradient pca
- shape context
- cross correlation
- har-aff esift
- steerable filters
- gradient moments
- complex filters
Scale change (factor 2.8)

- sift
- esift
- gradient pca
- shape context
- cross correlation
- har-aff esift
- steerable filters
- gradient moments
- complex filters

Graph showing the change in correct matches with different descriptors and methods.
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