Scale and Affine Invariant Interest Point Detectors

Krystian Mikolajczyk and Cordelia Schmid

*** Sources: Schmid (CVPR'03), Tuytelaars (ECCV'06).*

Why Local Features?

- Robust to noise, occlusion and clutter.
- Distinctive and repeatable.
- No explicit segmentation required – represent objects (classes).
- Invariance to image transformations + robust to illumination changes.
- Applications: SLAM, object (class) recognition, matching…

Some keywords…

- Harris corner detector.
  - Scale sensitive…
- Difference-of-Gaussian (DoG).
  - Lowe's paper: approx. to normalized LoG.
- Laplacian-of-Gaussian (LoG).
  - Normalized => Extrema in scale-space.
- Related to *second moment matrix* (SMM): second-order derivates of kernel-convolved image.

The scale-adapted SMM

- Terms: differentiation scale, integration scale, based on variance of kernel.

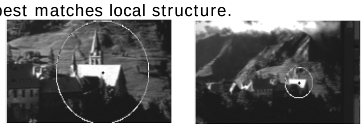
$$\mu(\mathbf{x},\sigma_I,\sigma_D)=\sigma_D^2 g(\sigma_I)\begin{bmatrix} L_x^2(\mathbf{x},\sigma_D) & L_xL_y(\mathbf{x},\sigma_D) \\ L_xL_y(\mathbf{x},\sigma_D) & L_y^2(\mathbf{x},\sigma_D)\end{bmatrix}$$

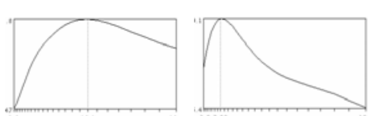
$$\sigma_D = s\sigma_I$$

- Ref: Elimination of edge responses in Lowe's paper using eigen values…

Characteristic Scale – scale invariance

- Apply *local operator* at scales: scale where operator best matches local structure.

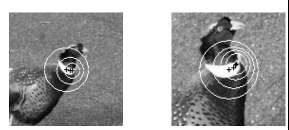




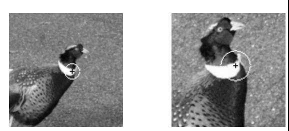
- LoG better than scale-adapted Harris.

Characteristic scale selection

- Multi-scale Harris.



- Characteristic scale with Laplacian.



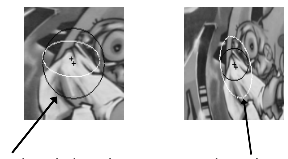
Scale invariant feature selection

- Harris-Laplace (HL) detector:
  - Harris measure, 8-neighborhood IP – larger scale ratio.
  - Iterate using LoG until convergence – smaller scale ratio.
- Simplified HL:
  - Reduce scale diff, find IP, keep those with LoG extremum.
- Simplified HL almost as good as HL.

7

Affine transforms – why?

- Viewpoint changes ~ affine transform.
- Scale changes by *different* amounts.
- Harris, HL *not* affine invariant.
- Operate in affine Gaussian scale-space: *ellipses* as point neighborhoods.



detected scale invariant region

projected region

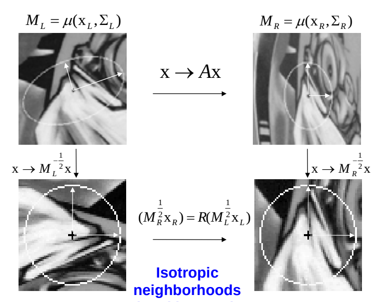
8

Affine Invariance – linear algebra 101 ☺

- Basis: Anisotropy is affine-transformed isotropy. High-dim search space.
- Constraints on Σ of Gaussian kernels:
  - recover affine shape,
  - reduce to orthogonal transform in normalized frames.
- Patterns in normalized frames are isotropic with respect to SMM.
- Estimation of Σ_L, Σ_R - iterative algorithm.

9

Affine Invariance – a picture



10

Affine Invariance – how?

- Step 1: Detect presence of affine transformation.
- Step 2: Transform IPs to normalized frames, get to circular point neighborhoods, achieve affine invariance…

11

Affine Transformation detection

- Eigen values – yes, again!
- Ratio of eigen values of SMM: eigen values equal=> isotropy.

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

- Once more, a measure of the skew/stretch.
- Ref: Lowe's feature rejection based on the *r-factor*.

12

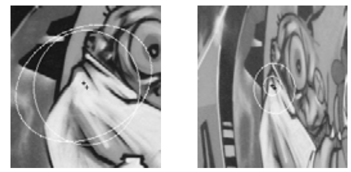
Algorithm (Σ_L, Σ_R) – iterate until convergence

- Shape adaptation – normalize window using a function of SMM.
- Select σ_I - remember *characteristic scale*.
- Select σ_D - equalize eigen values.
- Spatial localization of IP (interest point) – Harris detectors.
- Compute SMM and update normalization matrix.

13

Affine invariant Harris points

- Iterative estimation of localization, scale, neighborhood

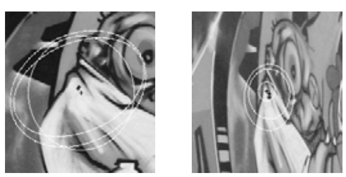


Iteration #1

14

Affine invariant Harris points

- Iterative estimation of localization, scale, neighborhood



Iteration #2

15

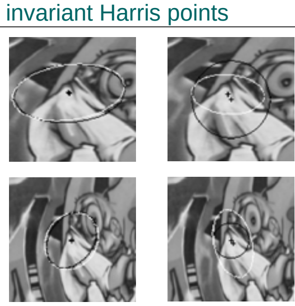
Affine invariant Harris points

- Iterative estimation of localization, scale, neighborhood

Iteration #3, #4, ...

16

Affine invariant Harris points



affine Harris   Harris-Laplace

17

Notes

- Convergence based on reasonable choice of scales and initial estimates.
  - *Initial estimates of IPs not affine invariant*.
- Averaging of similar features.
- Only (20-30)% of initial IPs used.
- Repeatability criterion.
- More robust to large viewpoint changes.
- *Smallest number of features found*.
- *Largest time complexity*.

18

Descriptors and Matching

- Normalized Gaussian gradient descriptors – *weak*!
  - Cause of matching failure – use SIFT descriptors (ref: Moreels + Perona evaluation)…
- Matching based on Mahalanobis distance and filters.
- Comparable performance under scale changes and localization errors.
- Performance much better under significant viewpoint changes.
- Next, some 'lab made' image results ☺

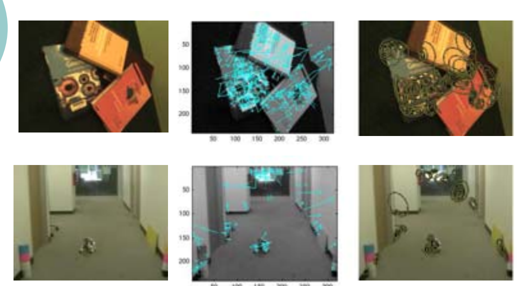
Matches – HarAff, large change in viewpoint

33 correct matches

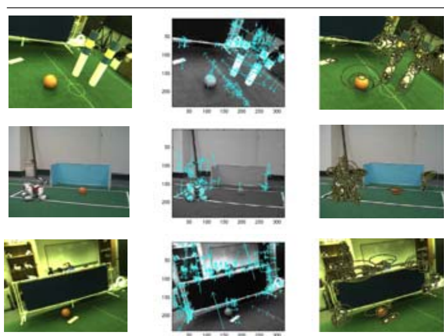
Matches – SIFT, large change in viewpoint

12 correct matches – hmm...

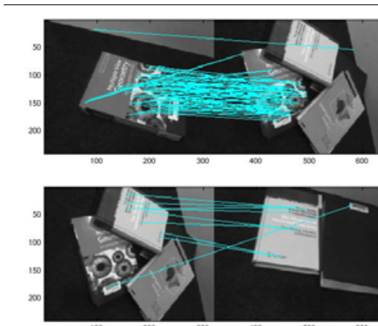
Images – difference in feature selection…



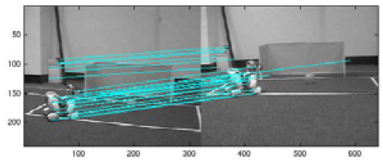
Images – difference in feature selection…



Some SIFT matching – good…

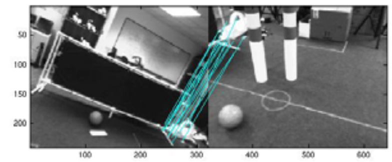


SIFT Matching – not so good…



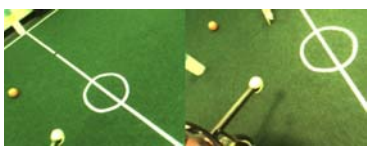
25

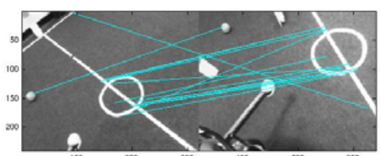
SIFT Matching – not so good…



26

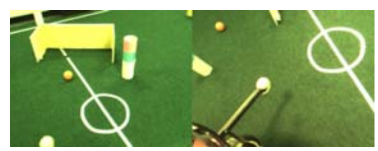
SIFT Matching – not so good…

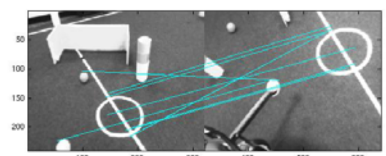




27

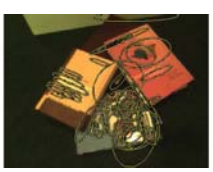
SIFT Matching – not so good…

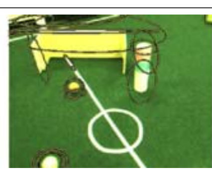


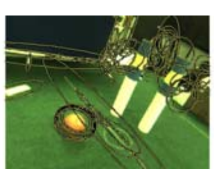


28

Some other methods – MSER

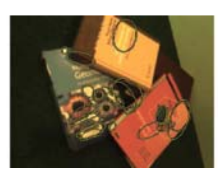


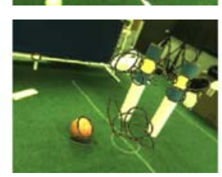




29

Some other methods – IBR





30

Observations…

- Local features intuitively appealing – *a lot of open questions still*.
- Scale, rotation, affine invariance, robust to viewpoint and illumination changes.
- Depend on *texture* in images – absence of texture can make it unreliable.
- Can add other features – Color? Texture? Structure?
- Can combine with feature-learning approaches?

That's all folks ☺