Estimation of Localization Uncertainty for Scale Invariant Feature Points

BMVC 2009

8.9.2009

Bernhard Zeisl[1]
bernhard.zeisl@mytum.de

Pierre Fite Georgel[1]
georgel@in.tum.de

Florian Schweiger[2]
florian.schweiger@tum.de

Eckehard Steinbach[2]
eckehard.steinbach@tum.de

Nassir Navab[1]
navab@cs.tum.edu

[1]Chair for Computer Aided Medical Procedures and Augmented Reality
[2]Institute for Media Technology
Technische Universität München, Munich, Germany

Introduction
Motivation and Problem Statement

Local features are state-of-the-art for a number of computer vision problems, e.g.:

- Object detection and localization
- Object recognition and Image retrieval
- Wide baseline matching and 3D reconstruction

Common assumptions for detected local features:

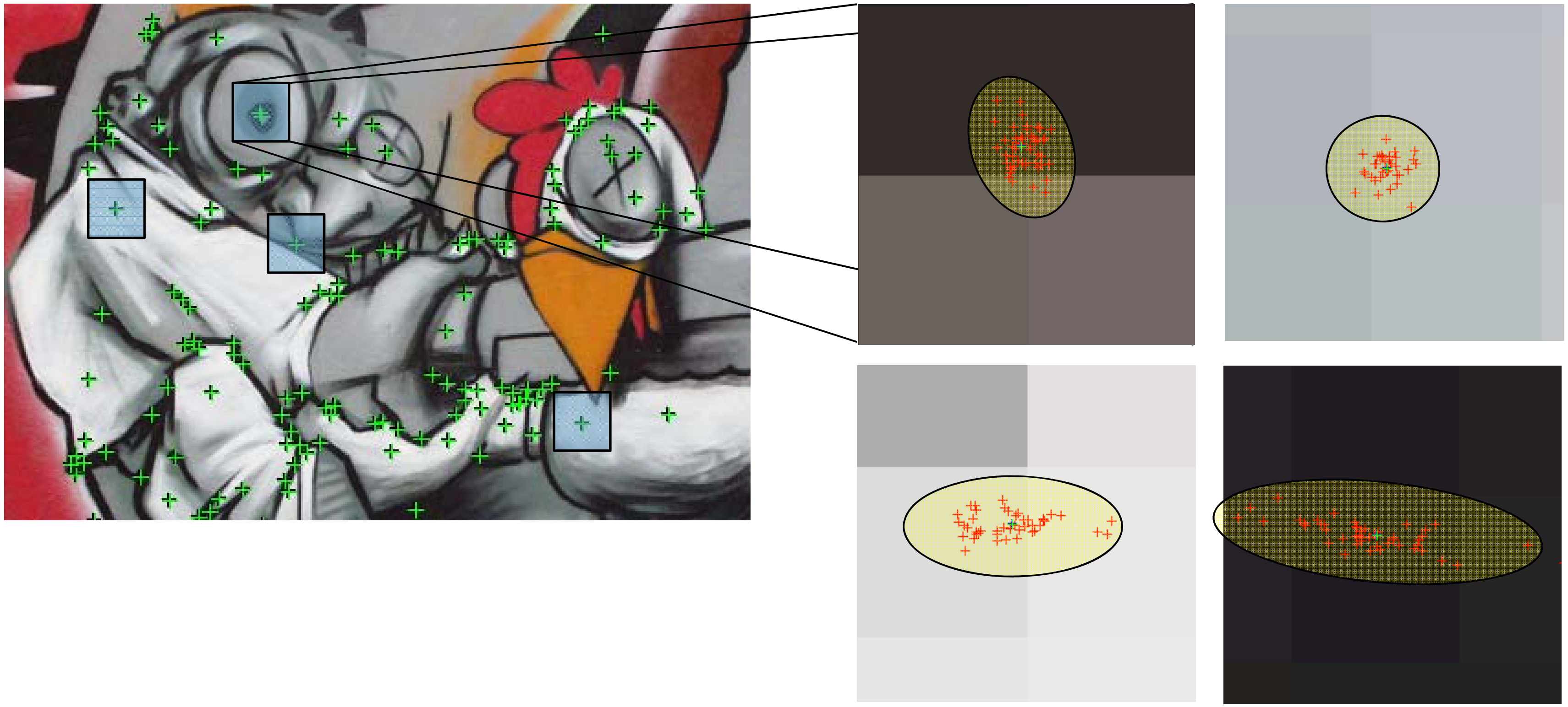
- Accurately detected or same deviation in localization error ($N(\mu, \sigma\mathbf{I})$)

→ Does not hold for image detectors searching in scale space.

Introduction

Motivation and Problem Statement

Repeated detection of same local feature under noise in the image:



Our method: Estimation of individual localization error for each feature found parameterized by a covariance matrix.

Agenda

Invariant Local Feature Detection

Uncertainty Estimation Framework

Experiments and Results

Conclusion

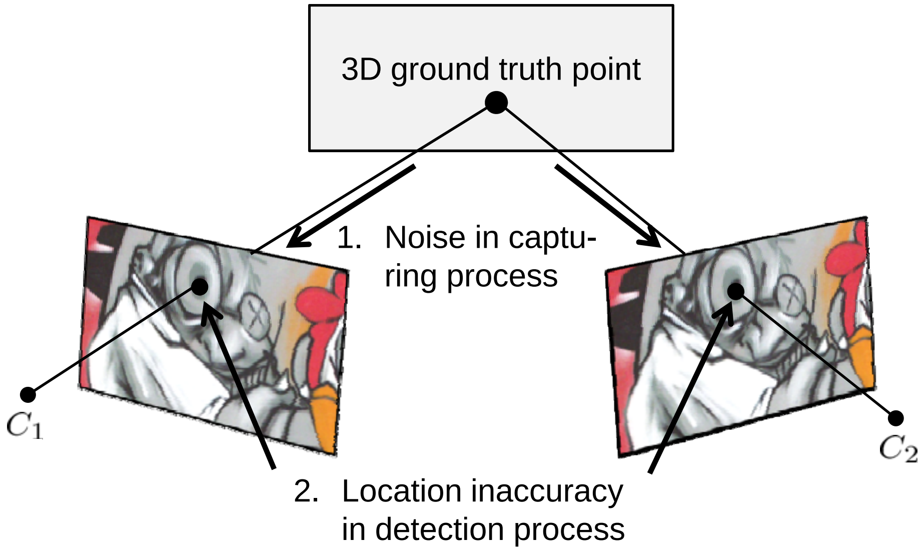
Localization Error

Inaccuracy is caused by pixel noise and the detection algorithm itself

Pixel Intensity Noise

Noise in pixel intensity values results from the image capturing process.

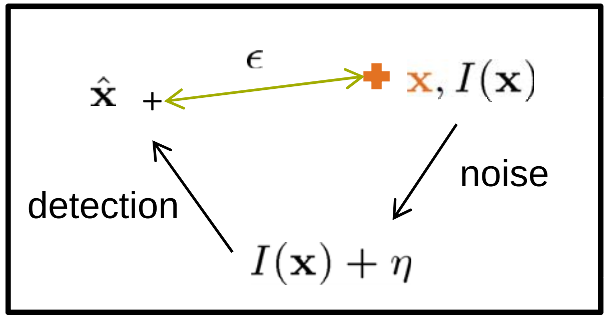
→ In different images a ground truth point will be mapped to different points $\hat{\mathbf{x}}$.



Detection Algorithm

Feature point detection algorithms use approximations in their calculation for complexity reasons.

→ Additional error introduced for the feature point $\hat{\mathbf{x}}$ depending on the algorithmic noise.



$$\epsilon = \hat{\mathbf{x}} - \mathbf{x}$$
$$\hat{\mathbf{x}} = dect(I(\mathbf{x}) + \eta)$$

$\mathbf{x} \ldots$ *ground truth feature point*

$I(\mathbf{x}) \ldots$ *intensity value at* $\mathbf{x}$

Scale Invariant Feature Detection

The same feature can be detected at different scales

Scale Space Representation

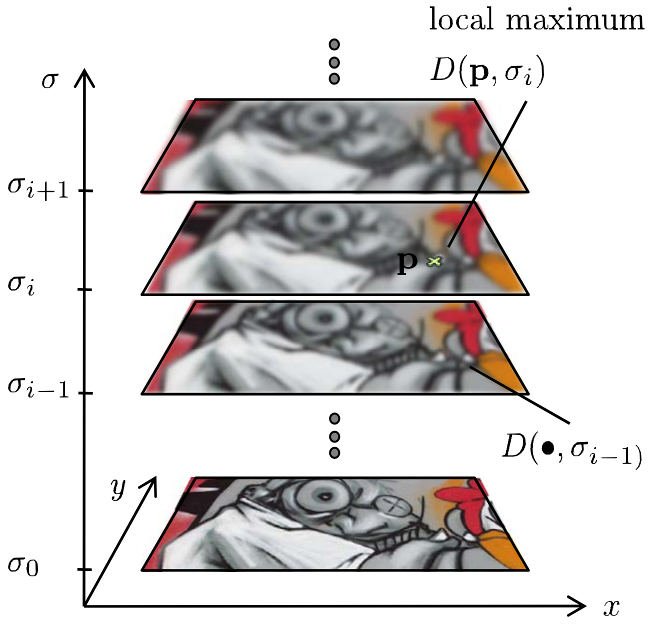
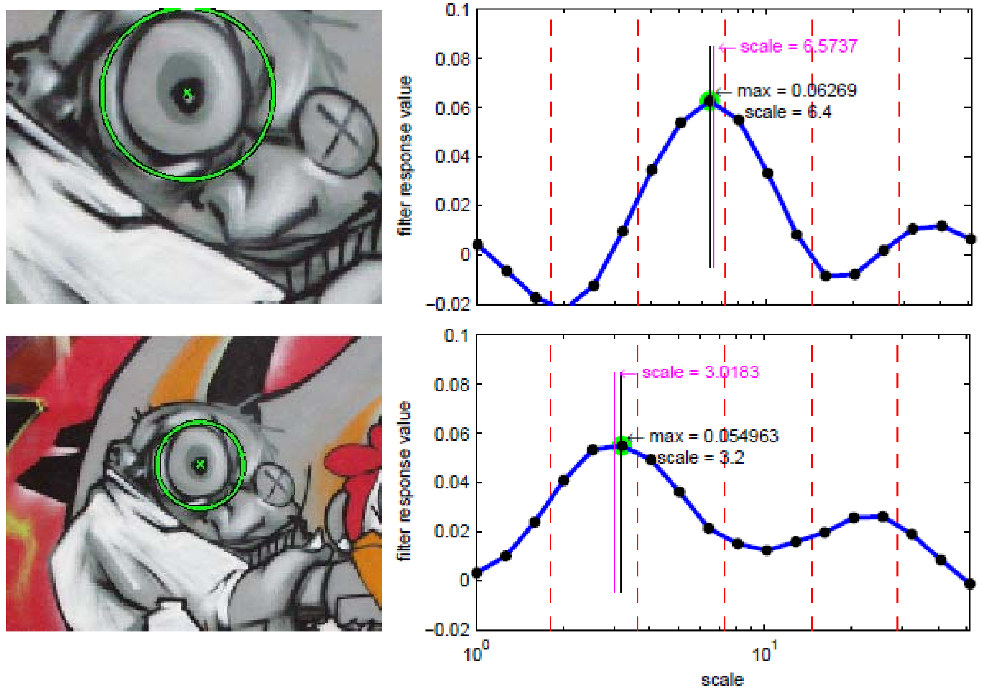


Image / detector response stack

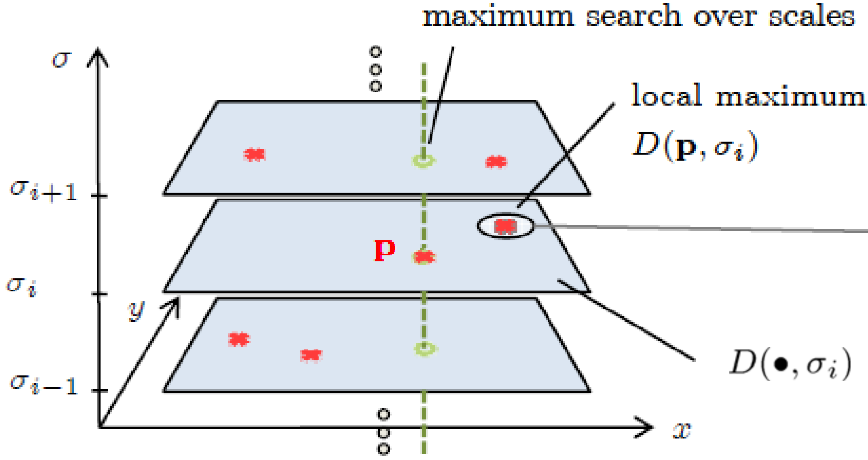
Characteristic Scale Selection

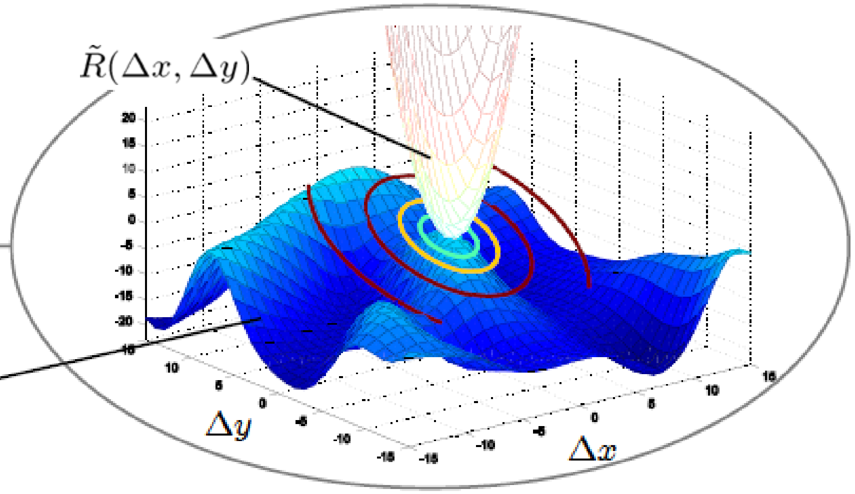


Mikolajczky, K., Schmid, C., *Scale & Affine Invariant Interest Point Detectors*, 2004

Uncertainty Evaluation Framework

Covariance is estimated from the detector response curvature





Residual at feature point:

$$R(\Delta\mathbf{p}) = |D(\mathbf{p}, \sigma) - D(\mathbf{p} + \Delta\mathbf{p}, \sigma)|$$

$$R(\Delta\mathbf{p}) \approx \tilde{R}(\Delta\mathbf{p}) = \frac{1}{2}\Delta\mathbf{p}^\top \mathbf{H} \Delta\mathbf{p}$$

Covariance based on Hessian:

$$\Sigma = \mathbf{H}^{-1} = \begin{bmatrix} D_{xx}(\mathbf{p}, \sigma) & D_{xy}(\mathbf{p}, \sigma) \\ D_{xy}(\mathbf{p}, \sigma) & D_{yy}(\mathbf{p}, \sigma) \end{bmatrix}^{-1}$$

low curvature → error due to the missing discriminative behavior of $D(\bullet, \sigma_i)$ in $\mathcal{N}_{\mathbf{p}}$.
high curvature → detection process more accurate

Kanazawa, Y., Kanatani, K., *Do we really have to consider covariance matrices for image features?*, ICCV 2001

Framework Application

Application is identical for SIFT and SURF

	SIFT	SURF
Detector function	$D(\mathbf{x},\sigma_i) = \underbrace{(G(\mathbf{x},\sigma_{i+1}) - G(\mathbf{x},\sigma_i))}_{\approx\nabla^2 G(\mathbf{x},\sigma_i)} * I(\mathbf{x})$	$D(\mathbf{x},\sigma_i) = \det\begin{bmatrix} L_{xx}(\mathbf{x},\sigma_i) & L_{xy}(\mathbf{x},\sigma_i) \\ L_{xy}(\mathbf{x},\sigma_i) & L_{yy}(\mathbf{x},\sigma_i) \end{bmatrix}$

Covariance calculation

$$\Sigma = \left(\sum_{i,j \in \mathcal{N}_\mathbf{p}} w(i,j) \cdot \begin{bmatrix} D_{xx}(i,j,\sigma) & D_{xy}(i,j,\sigma) \\ D_{xy}(i,j,\sigma) & D_{yy}(i,j,\sigma) \end{bmatrix} \right)^{-1}$$

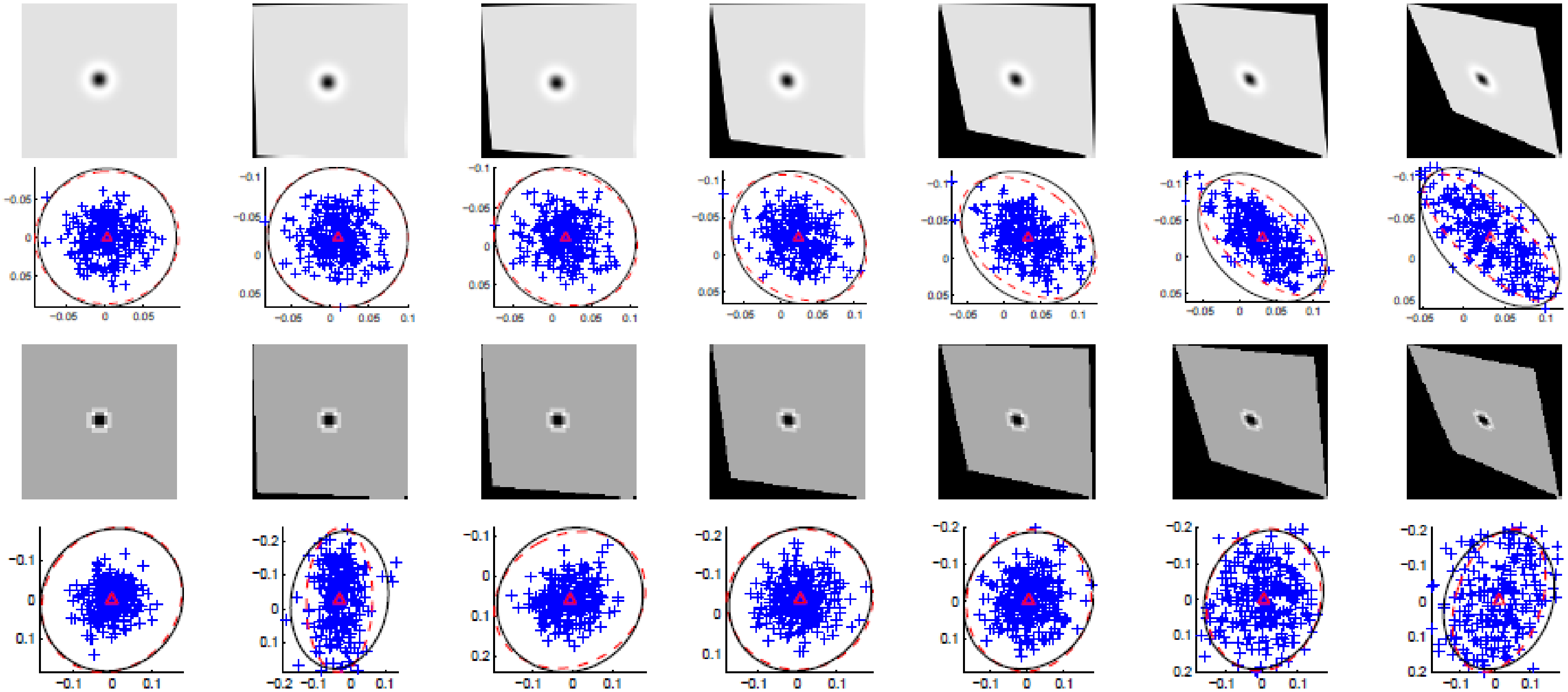
$$D_{xx} = d_{xx} \cdot D(\mathcal{N}_\mathbf{p}, \sigma_i)$$

Back projection

$$\Sigma^{(0)} = \Sigma \cdot (2^{octave})^2$$

Statistical Error Modeling

Maximum likelihood estimate and our covariance coincide



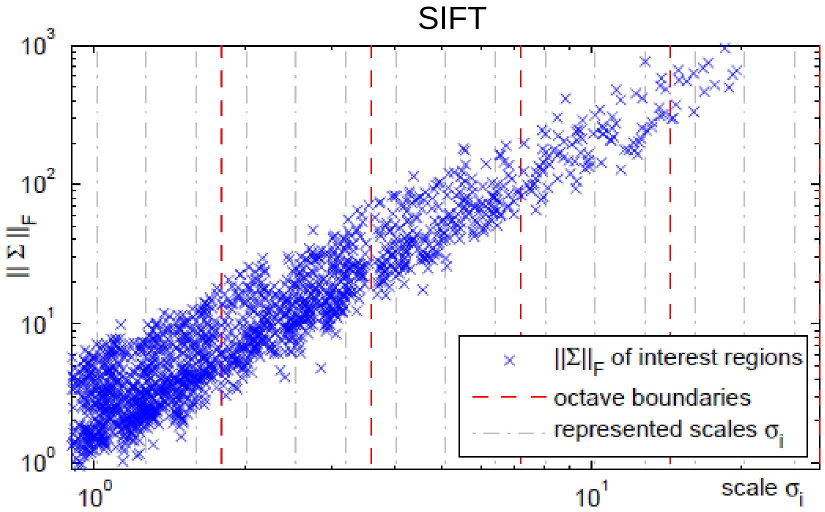
(+) distribution of location error (--) our covariance estimate (- -) maximum likelihood estimate

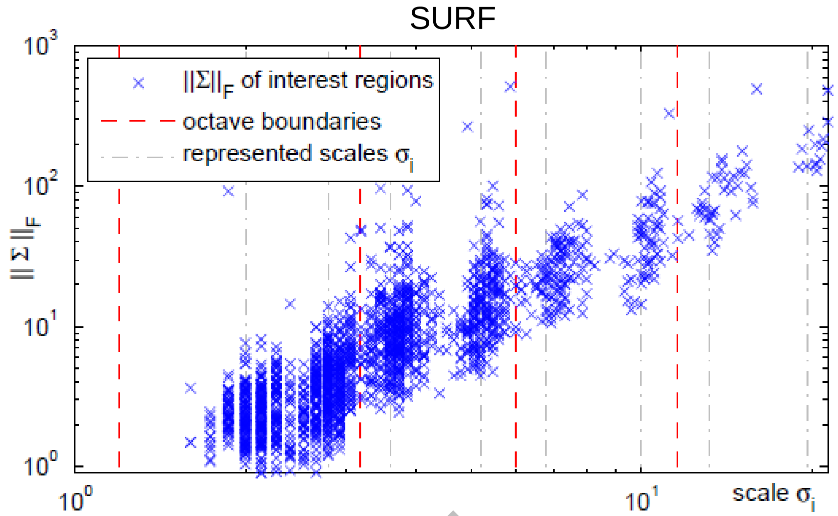
The covariance estimates fit the modeled error distribution

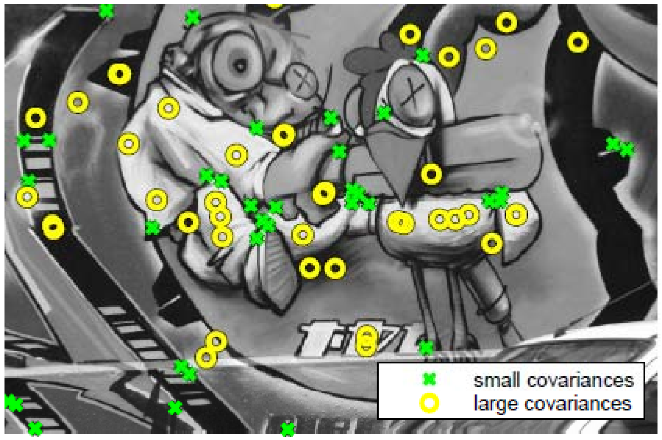
Schweiger, F. et. al., *Maximum Detector Response Markers for SIFT and SURF*, VMV 2009

Covariance Dependence on Scale

Feature points are localized better on smaller scales







Change of Frobenius norm over detection scale for feature points detected in real images.

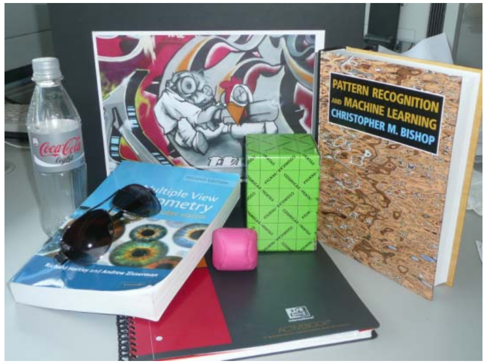
Feature points with small ($\sigma < 2.1$) and large ($\sigma > 8$) covariances.

Blobs are worse localized than distinctive image points.

Covariance Dependence on Scale

Covariances imply automatic scale normalization

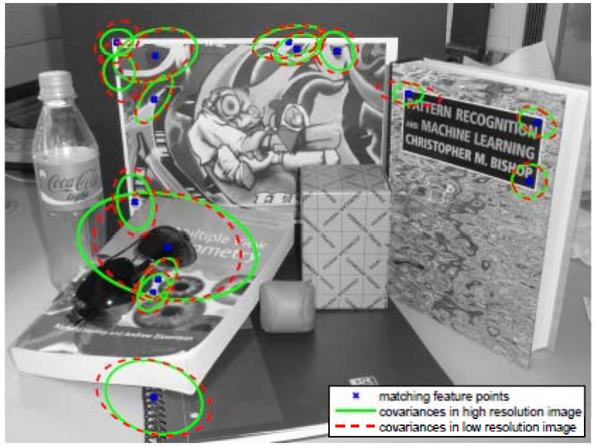
High and low resolution images:



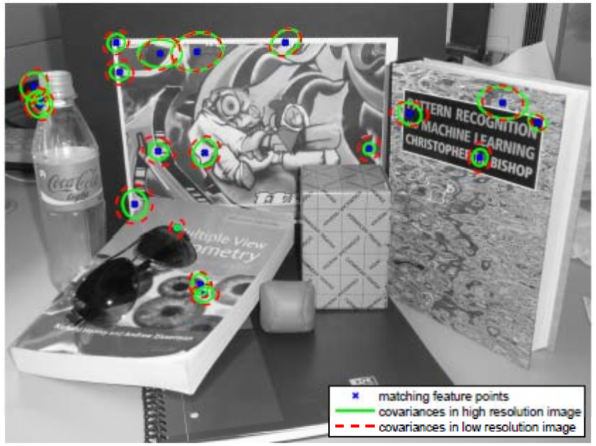
3072x2304 pixel

800x600 pixel

Covariances of matching feature points in the two images:
(covariances are projected with the underlying homography)



SIFT



SURF

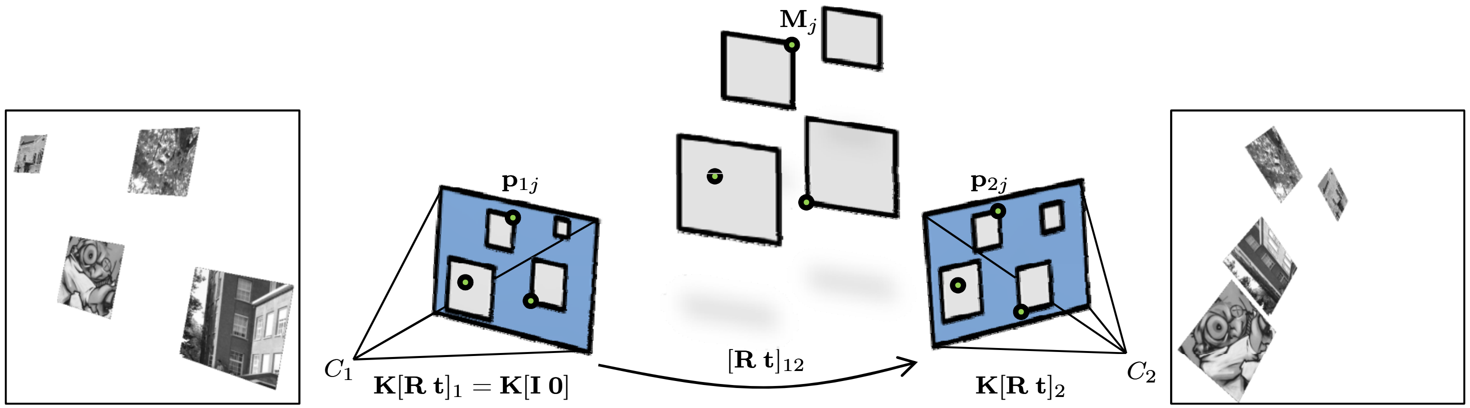
Corresponding feature points are detected at different scales;
but (projected) covariances of features are almost identically

→ Localization error is similar in both images *in relation to their size*

Covariances normalize and weight the error in an optimization and thus differently sized images can be used

Results for Model Fitting
Bundle Adjustment

Bundle adjustment simultaneously refines the 3D coordinates describing the scene geometry as well as camera poses and intrinsic camera parameters.



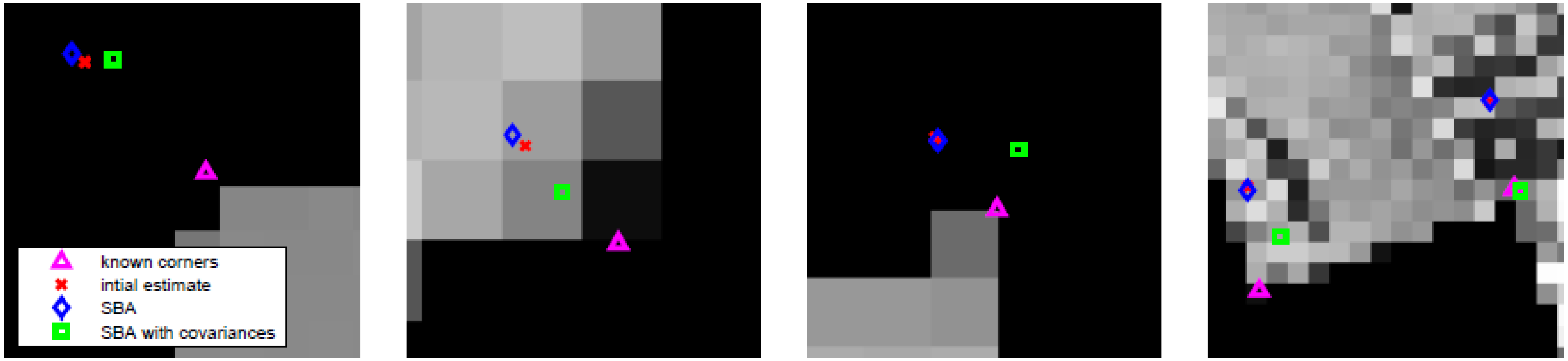
Euclidian distance: $$\arg \min_{[\mathbf{R}\ \mathbf{t}]_i, \mathbf{M}_j} \sum_{i=1}^{N} \sum_{j=1}^{m} v_{ij} \cdot (\mathbf{p}_{ij} - w(\mathbf{K}[\mathbf{R}\ \mathbf{t}]_i \mathbf{M}_j))^\top (\mathbf{p}_{ij} - w(\mathbf{K}[\mathbf{R}\ \mathbf{t}]_i \mathbf{M}_j))$$

Mahalanobis distance: $$\arg \min_{[\mathbf{R}\ \mathbf{t}]_i, \mathbf{M}_j} \sum_{i=1}^{N} \sum_{j=1}^{m} v_{ij} \cdot (\mathbf{p}_{ij} - w(\mathbf{K}[\mathbf{R}\ \mathbf{t}]_i \mathbf{M}_j))^\top \mathbf{\Sigma}_{\mathbf{ij}}^{-1} (\mathbf{p}_{ij} - w(\mathbf{K}[\mathbf{R}\ \mathbf{t}]_i \mathbf{M}_j))$$

Triggs, B. et. al., *Bundle Adjustment - A Modern Synthesis*, Lecture Notes in Computer Science 1999

Bundle Adjustment

Performance is evaluated with the reprojection error of corner points



Reprojection error of 3D corner points:

$$e = \frac{1}{4} \sum_{i=1}^{4} \left\| \bar{\mathbf{c}}_i - w(\hat{\mathbf{T}}\bar{\mathbf{C}}_i) \right\|$$

$\bar{\mathbf{c}}_i$ *... ground truth 2D corner point*

$\bar{\mathbf{C}}_i$ *... ground truth 3D corner point*

$\hat{\mathbf{T}}$ *... estimated projection*

Mean performance as pixel offset for about 100 different image pairs:

	mean all patches	
covariance usage	no	yes
SIFT	2.031	1.759
SURF	2.554	2.363

We get a performance improvement for the reconstruction with bundle adjustment using our feature point covariances.

Conclusion

Main Contributions

- Derivation of general formulation for feature detection in scale space
- Computation of stable covariances for scale invariant image features
- Justification of correctness for our covariance estimates
- Inherent scale normalization
- Performance improvement for bundle adjustment

We would like to encourage you to test and use our results:

Code and binaries for SIFT and SURF local feature detection and covariance estimation are available at: http://campar.in.tum.de/Main/CovarianceEstimator