



# Photo-based Augmented Reality

## Registration And Navigation Methods

*Application to Discrepancy Check*

*Imperial College London*

*11/09/09*

*Pierre Georgel*

---

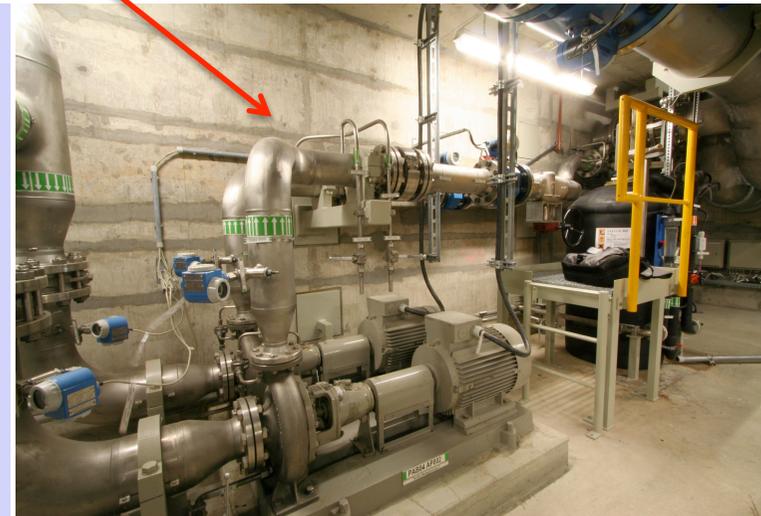
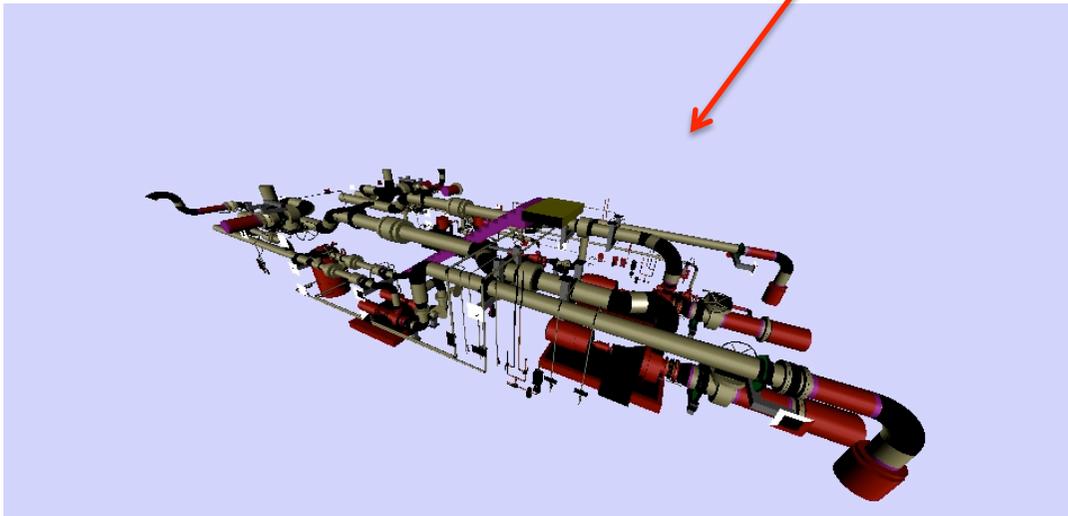


## Agenda

- Problem statement and application
- Software features
- Estimation of uncertainty localization for multi-scale local features.
- Outline of future work

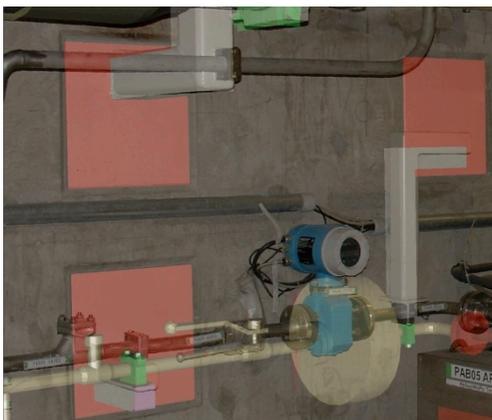
## Project Overview

- Discrepancy check between CAD data and built items

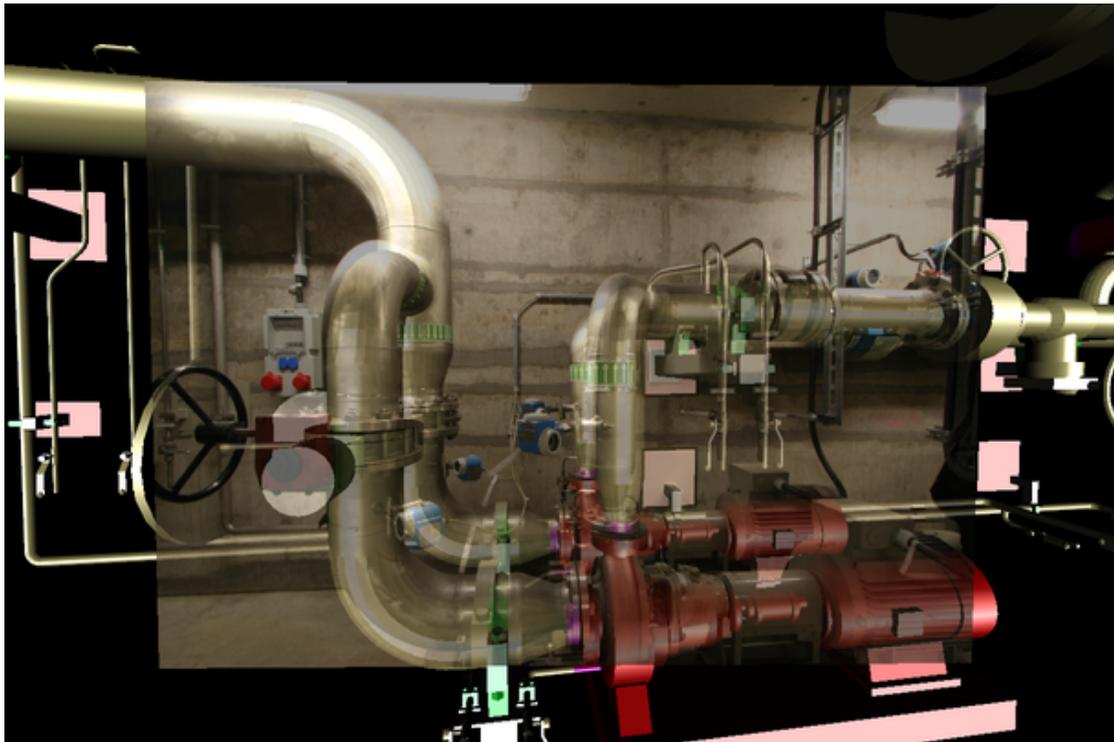


## Project Overview

- Discrepancy check between CAD data and built items
- Project started in February 2006
- Consortium of CAMP – Siemens CT – Areva NP
- Yearly release - version 3.5 was ship last month



## Photo-based Augmented Reality



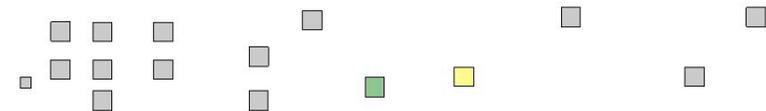
*Single frame augmentation*

## Requirements

1. Registration
2. Interaction

## Landmark-based Registration

- Goal
  - Register images to the CAD coordinate system
- Approach
  - Extract landmarks (Anchor-Plates) from images
  - Find corresponding 3D landmarks
  - Compute the Pose (Rotation, Translation)



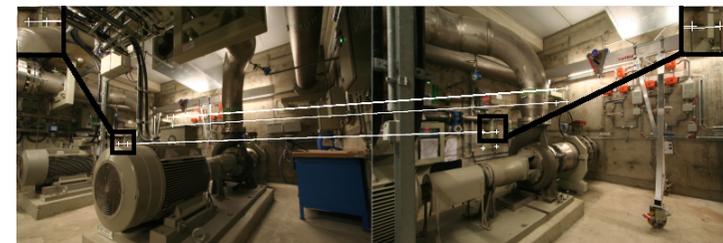
- Limitation
  - Need of 2d visible Landmarks

## Automatic 6 DoF Stereo Registration

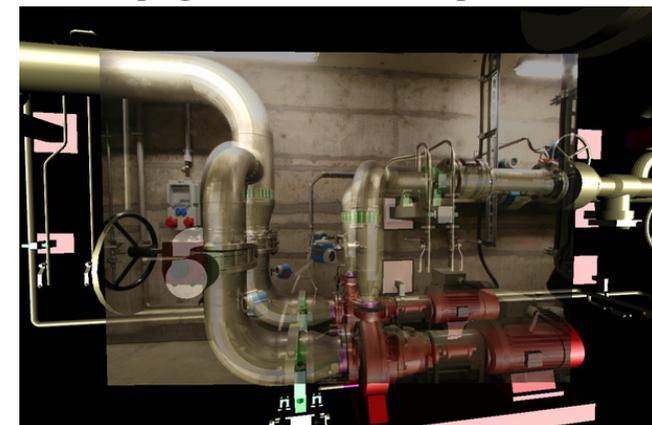
- Goal
  - Register images using a unique keyframe
- Approach
  - Compute epipolar geometry using keypoints
  - Extend the relative pose to a full pose using
    - Planar structure
    - 2d-3d correspondences
  - Non-linear estimation
    - Bundle adjustment with planarity constraint
    - Hybrid pose estimation that includes intensities



*Extracted planar structure*

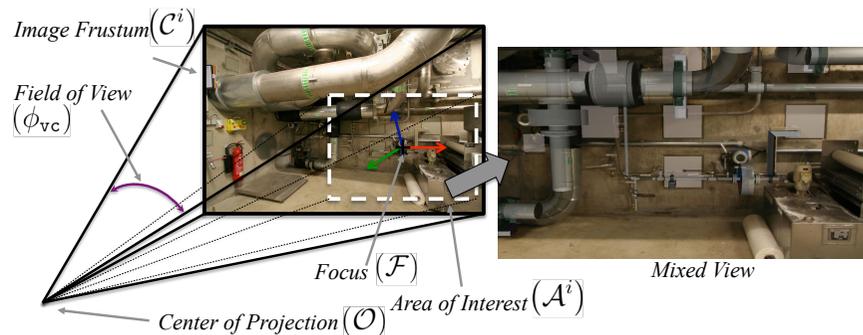


*Propagated 2d-3d correspondences*

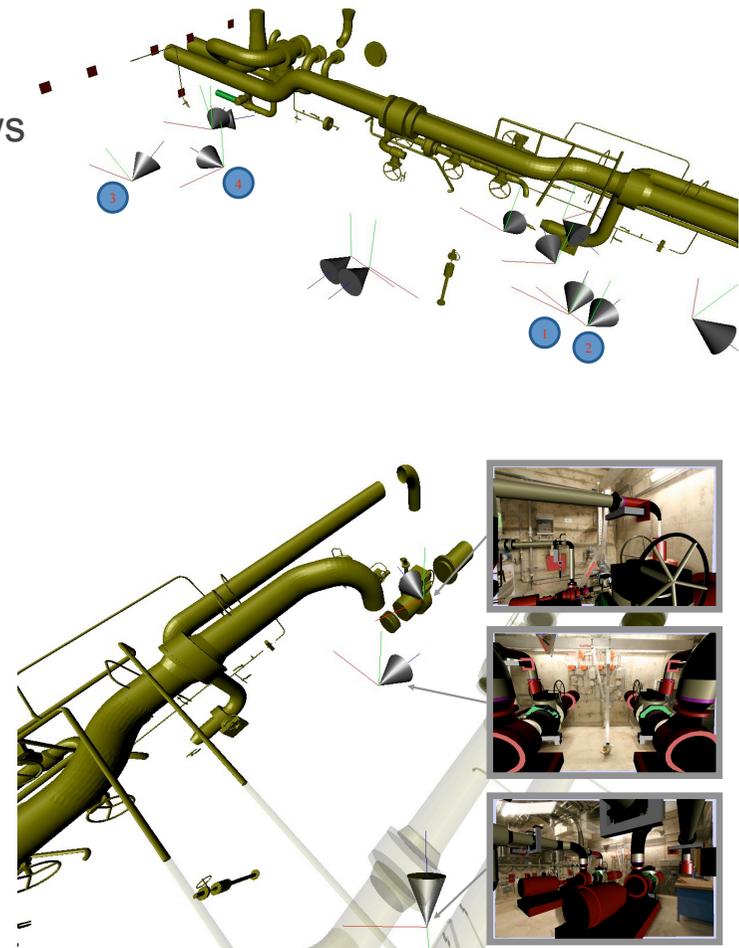


# 3D User Interface for an Augmented CAD Software

- New paradigm requires new interactions
  - Extension of 2D zoom/pan interface to Mixed views



- Intuitive 3D image navigation
  - No points track are required
  - No Model is considered



# Estimation and Exploitation of Localization Uncertainty for Scale Invariant Feature Points

BMVC 2009

**Bernhard Zeisl,  
Florian Schweiger\*  
Eckehard Steinbach\*  
Nassir Navab**

*I stole these slides from Bernhard Zeisl*

\* Institute for Media Technology, see <http://www.lmt.ei.tum.de>



# Agenda

**1. Motivation**

**2. Image Feature Detection and Related Work**

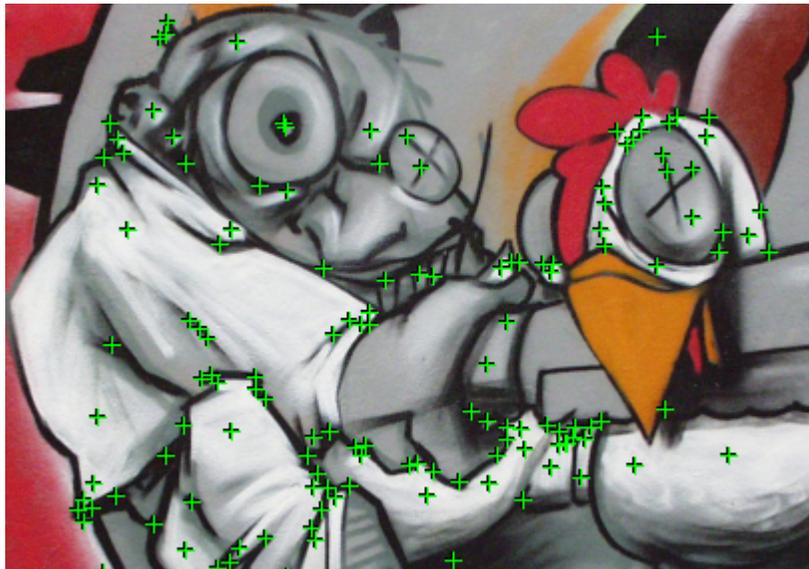
**3. Uncertainty Estimation Framework and its Application**

**4. Experiments**

**5. Results for Model Fitting**

# 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



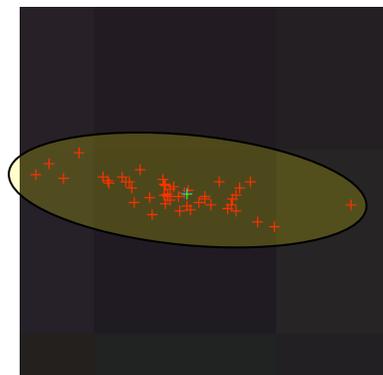
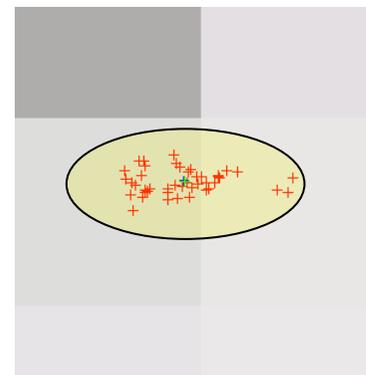
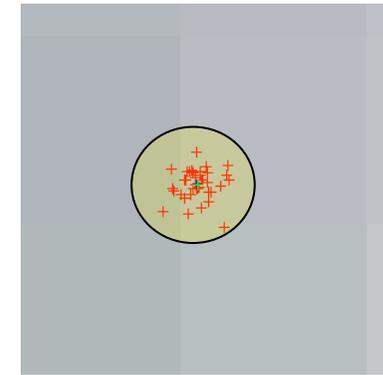
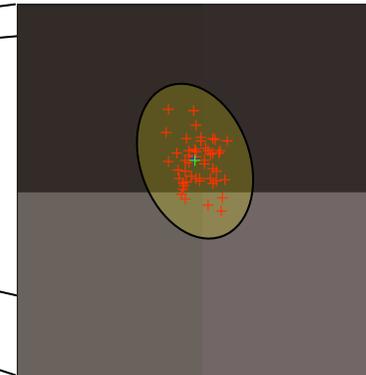
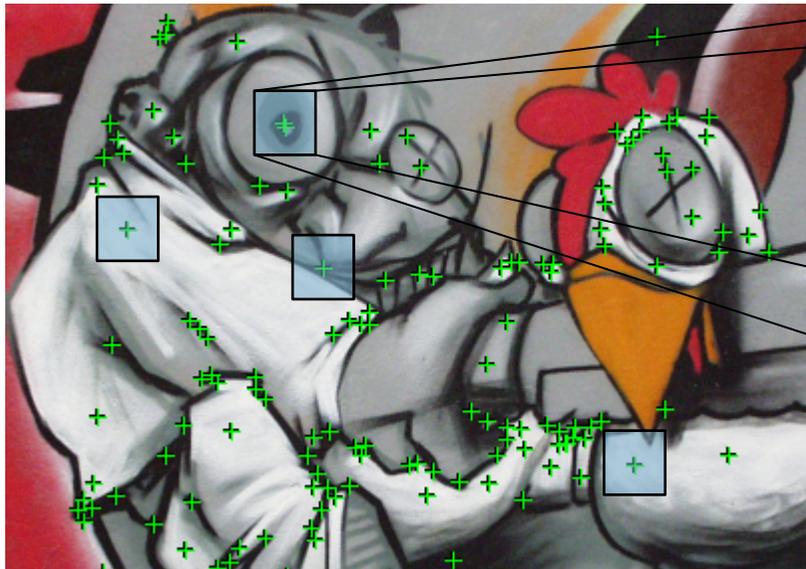
Common assumptions for detected local features:

- 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**.

# Image Feature Detection

## Concept of Detector and Descriptor

	Detector	Descriptor
<b>Purpose</b>	<b>Localization</b> of feature points	<b>Matching</b> of feature points
<b>Approach</b>	Spatial detection done via a mathematical operator e.g.: DoG, $\det(H)$ , $\text{trace}(H)$	Description by the structure of the local neighborhood e.g.: weighted sampling of gradient field
<b>Desired properties</b>	<ul style="list-style-type: none"><li>Stability – detection independent of changes in the image conditions Measured in terms of repeatability, i.e. 3D point detectable in two images capturing the same scene</li><li>Localization precision</li></ul>	<ul style="list-style-type: none"><li>Distinctiveness</li><li>Robustness to occlusions</li><li>Invariance to image transformations.</li></ul> Reached by appropriate size and sampling of the gradient field

# Image Feature Detection

Corners and blobs in an image are complementary features

## Corner Detector

## Blob Detector

### Examples



Harris corner detector



MSER detector

### Feature definition

Location where two dominant directions intersect

Areas brighter or darker than the surrounding.

### Detector

is sensitive to local regions which have a high degree of variation in all directions.

localizes blobs by a well-defined point - the blob center.

## Related Work in Feature Detection

In recent work blob detectors have found increasingly popular use

### Corner Detectors

- Harris (1988): **Harris** corner detector: Based on second moment matrix; measures the local change of the image.
- Smith, Brady (1995): **SUSAN**: Based on the definition of corners using no image derivatives.
- Mikolajczyk, Schmid (2002): **Harris Laplace** & **Harris-Affine**: Scale space adaption of the second moment matrix (representation as pyramid); the latter is also invariant to affine transformations.
- Rosten, Drummond (2005): **FAST**, efficient implementation of SUSAN using decision trees; extensively used in SLAM.

### Blob Detectors

- Lowe (1999): **SIFT** (Scale invariant feature transform): DoG filter to search for features in scale space; Efficient implementation for descriptor matching.
- Matas (2002): **MSER** (Maximally stable extremal regions): no scale space representation, but still scale invariant; detects blobs by subsequently thresholding the image.
- Mikolajczyk, Schmid (2002): **Hessian-Affine**: Similar to the Harris-Affine, but Hessian based detector finds blobs in an image.
- Bay et. al. (2006): **SURF** (Speed up robust features): Hessian based detector; similar to SIFT, but fast implementation by usage of integral images and block filters.

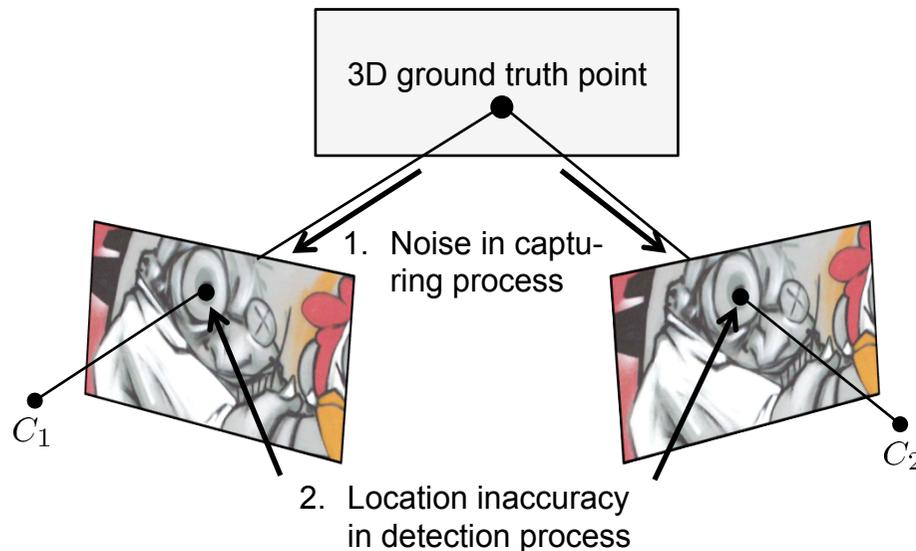
## Uncertainty Estimation Framework

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**.

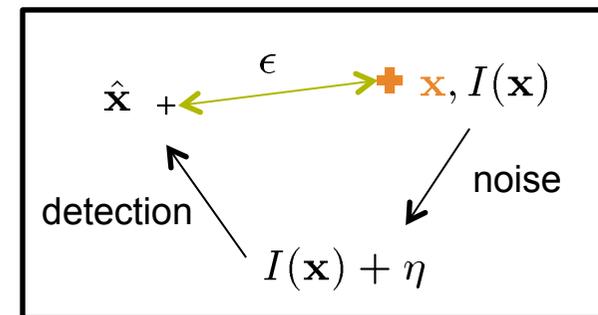
True feature point  $\mathbf{x}$  is mapped to different points according to the noise distribution for each image taken from the scene.



### Detection Algorithm

Feature point detection algorithms are based on the exploration of the local scene structure (1<sup>st</sup> or 2<sup>nd</sup> derivatives).

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



$$\epsilon = \hat{\mathbf{x}} - \mathbf{x}$$

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

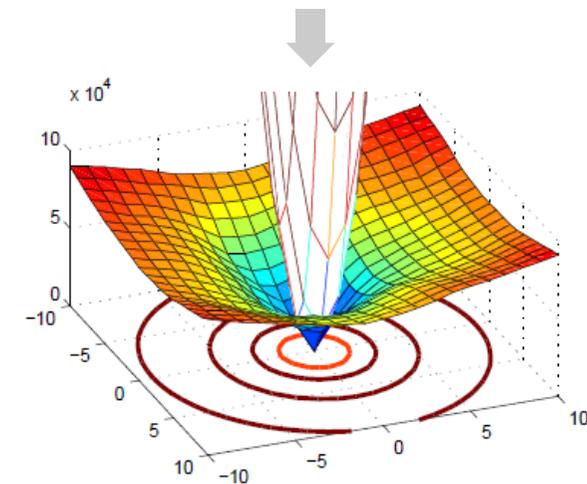
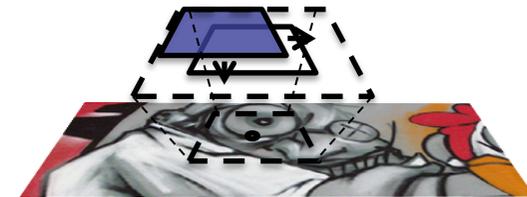
$\mathbf{x} \dots$  ground truth feature point

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

## Related Work in Uncertainty Estimation

### Localization error addressed without scale-space consideration

- Kanatani & Kanazawa (2001):  
Argue that meaningful covariance matrices can be computed from the **self-matching residual**, but **no improvement** for homography or fundamental matrix computation.
- Brooks et. al. (2001):  
Curvature of **self-matching residual** for covariance estimation. Show performance **gain for fundamental matrix** computation.
- Steel & Jaynes (2005):  
Argue that gradient based methods overestimate covariances. Derive covariances from **different noise models** for pixel intensities.
- Haja et. al. (2008):  
Comparison of detectors with respect to localization accuracy. Localization accuracy evaluated in terms of **matching precision**.
- Wu et. al. (2008):  
Observe **relation between location precision and detection scale**. Introduce less weighting for points at higher scales.



Residual from template matching

# Related Work in Uncertainty Estimation

## Localization error addressed without scale-space consideration

- Kanatani & Kanazawa (2001):  
Argue that meaningful covariance matrices can be computed from the **self-matching residual**, but **no improvement** for homography or fundamental matrix computation.
- Brooks et. al. (2001):  
Curvature of **self-matching residual** for covariance estimation. Show performance **gain for fundamental matrix** computation.
- Steel & Jaynes (2005):  
Argue that gradient based methods overestimate covariances. Derive covariances from **different noise models** for pixel intensities.
- Haja et. al. (2008):  
Comparison of detectors with respect to localization accuracy. Localization accuracy evaluated in terms of **matching precision**.
- Wu et. al. (2008):  
Observe **relation between localization precision and detection scale**. Introduce less weighting for points at higher scales.

Due to focus on scale invariant interest regions, we claim:

1. **Shape of covariance will be in general anisotropic.**

2. **Magnitude will vary according to detection scale.**

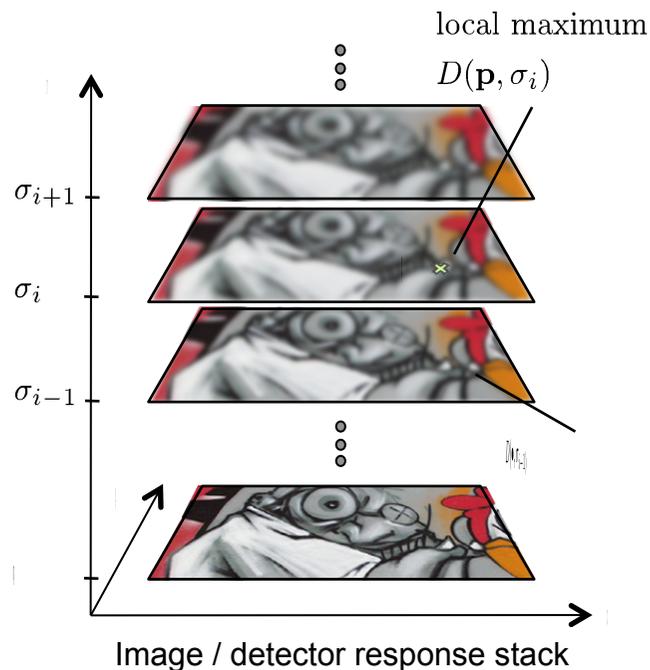
No scale space consideration

Observation of relation between scale and position

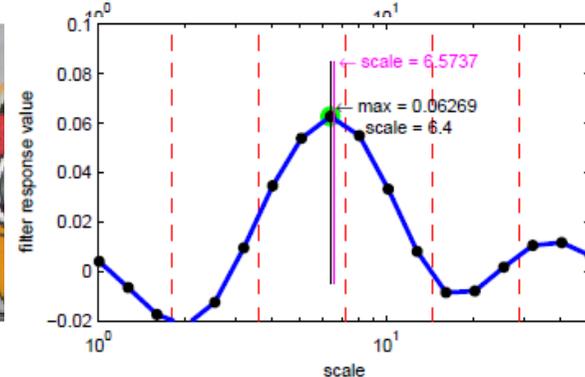
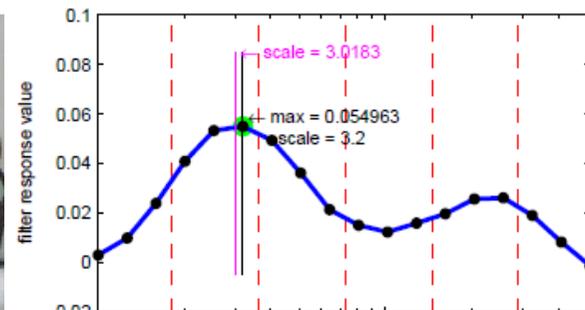
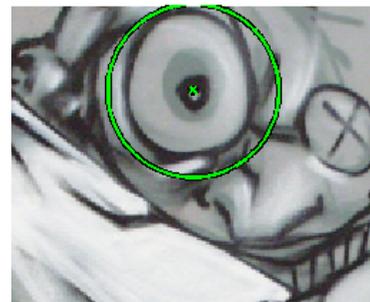
# Scale Invariant Feature Detection

The same feature can be detected at different scales

## Scale Space Representation



## Characteristic Scale Selection



# Scale Invariant Feature Detection

A novel general formulation for feature detection in scale space

## Scale Space Representation

- Scale space representation of the image with detection operator  $\mathbf{f}_{\text{dec}}$ .

$$D(\mathbf{x}, \sigma_i) = \mathbf{f}_{\text{dec}}(I(\mathcal{N}_{\mathbf{x}}), \sigma_i)$$

- Candidate set  $\mathbb{P}_1$  of feature points  $\langle \mathbf{p}, \sigma_i \rangle$  by local maxima detection in  $D(\bullet, \sigma_i)$ .

$$\mathbb{P}_1 := \bigcup_{i=1}^N \left\{ \langle \mathbf{p}, \sigma_i \rangle \mid \mathbf{p} = \arg \max_{\mathbf{x} \in \mathcal{N}_{\mathbf{p}}} D(\mathbf{x}, \sigma_i) \right\}$$

## Scale Selection

- Scale space representation of the image with scale selection operator  $\mathbf{f}_{\text{sel}}$ .

$$S(\mathbf{p}, \sigma_i) = \mathbf{f}_{\text{sel}}(I(\mathcal{N}_{\mathbf{p}}), \sigma_i)$$

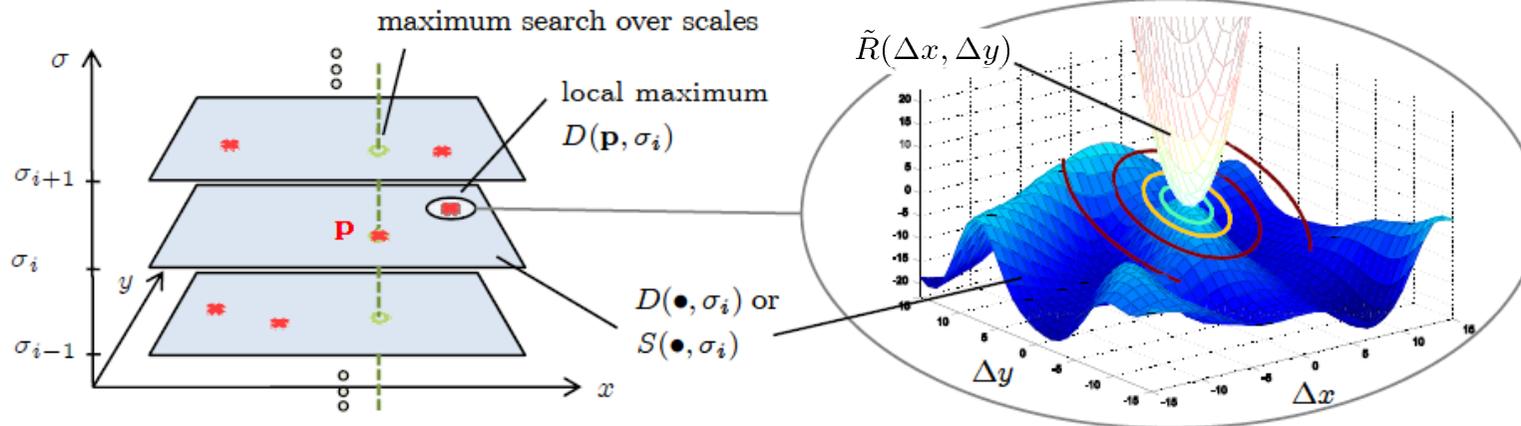
- Set  $\mathbb{P}_2$  of feature points  $\langle \mathbf{p}, \sigma \rangle$  by local maxima detection at positions  $\mathbf{p}$  in  $S(\mathbf{p}, \sigma_i)$

$$\mathbb{P}_2 := \left\{ \langle \mathbf{p}, \sigma \rangle \mid \langle \mathbf{p}, \sigma \rangle \in \mathbb{P}_1, \sigma = \arg \max_{\sigma_i} S(\mathbf{p}, \sigma_i), S(\mathbf{p}, \sigma) > \tau \right\}$$

Applicable for all detectors building upon a representation in scale space.

# Uncertainty Evaluation Framework

Covariance are estimated from the detector response curvature



**Residual** at feature point:

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

$$\mathbf{p} = \arg \max_{\mathbf{x} \in \mathcal{N}_{\mathbf{p}}} D(\mathbf{x}, \sigma) = \arg \min_{\Delta \mathbf{p} \in \mathcal{N}_0} R(\Delta \mathbf{p})$$

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

**Covariance** based on Hessian:

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

$\Sigma^{(0)} =$

low curvature  $\rightarrow$  error due to the missing discriminative behavior of  $D(\bullet, \sigma_0)$  in  $\mathcal{N}_{\mathbf{p}}$ .  
 high curvature  $\rightarrow$  detection process more accurate

# Framework Application

Application is identically for SIFT and SURF

	SIFT	SURF
<b>Detector function</b>	$D(\mathbf{x}, \sigma_i) = \underbrace{(G(\mathbf{x}, \sigma_{i+1}) - G(\mathbf{x}, \sigma_i))}_{:=S(\mathbf{x}, \sigma_i)} * I(\mathbf{x}) \approx \nabla^2 G(\mathbf{x}, \sigma_i)$	$D(\mathbf{x}, \sigma_i) = \det \begin{bmatrix} L_{xx}(\mathbf{x}, \sigma_i) & L_{xy}(\mathbf{x}, \sigma_i) \\ L_{xy}(\mathbf{x}, \sigma_i) & L_{yy}(\mathbf{x}, \sigma_i) \end{bmatrix} := S(\mathbf{x}, \sigma_i)$
<b>Covariance calculation</b>	$\Sigma = \left( \mp \sum_{i,j \in \mathcal{N}_{\mathbf{p}}} w(i,j) \cdot \begin{bmatrix} D_{xx}(i,j,\sigma) & D_{xy}(i,j,\sigma) \\ D_{xy}(i,j,\sigma) & D_{yy}(i,j,\sigma) \end{bmatrix} \right)^{-1}$ $D_{**} = d_{**} \cdot D(\mathcal{N}_{\mathbf{p}}, \sigma_i)$	
<b>Back projection</b>	$\Sigma^{(0)} = \Sigma \cdot (2^{\text{octave}})^2$	



## Optimal Images for SIFT and SURF

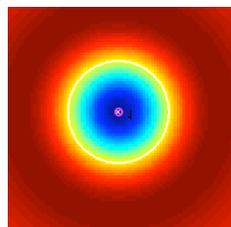
Image feature for maximal detector response are computable

### SIFT

DoG is a linear filter

→ matched filter approach

$$I(\mathbf{x}) = G(\mathbf{x} - \mathbf{p}, \sigma_{i+1}) - G(\mathbf{x} - \mathbf{p}, \sigma_i)$$



### SURF

det(H) is non-linear detector:

→ Optimization via quadratically constrained quadratic program

$$\det \mathbf{H} = \det \begin{bmatrix} h_{xx} & h_{xy} \\ h_{yx} & h_{yy} \end{bmatrix}$$

$$h_{xx} = \mathbf{f}_{xx}^\top \mathbf{I}$$

$$h_{xy} = h_{yx} = \mathbf{f}_{xy}^\top \mathbf{I}$$

$$h_{yy} = \mathbf{f}_{yy}^\top \mathbf{I}$$

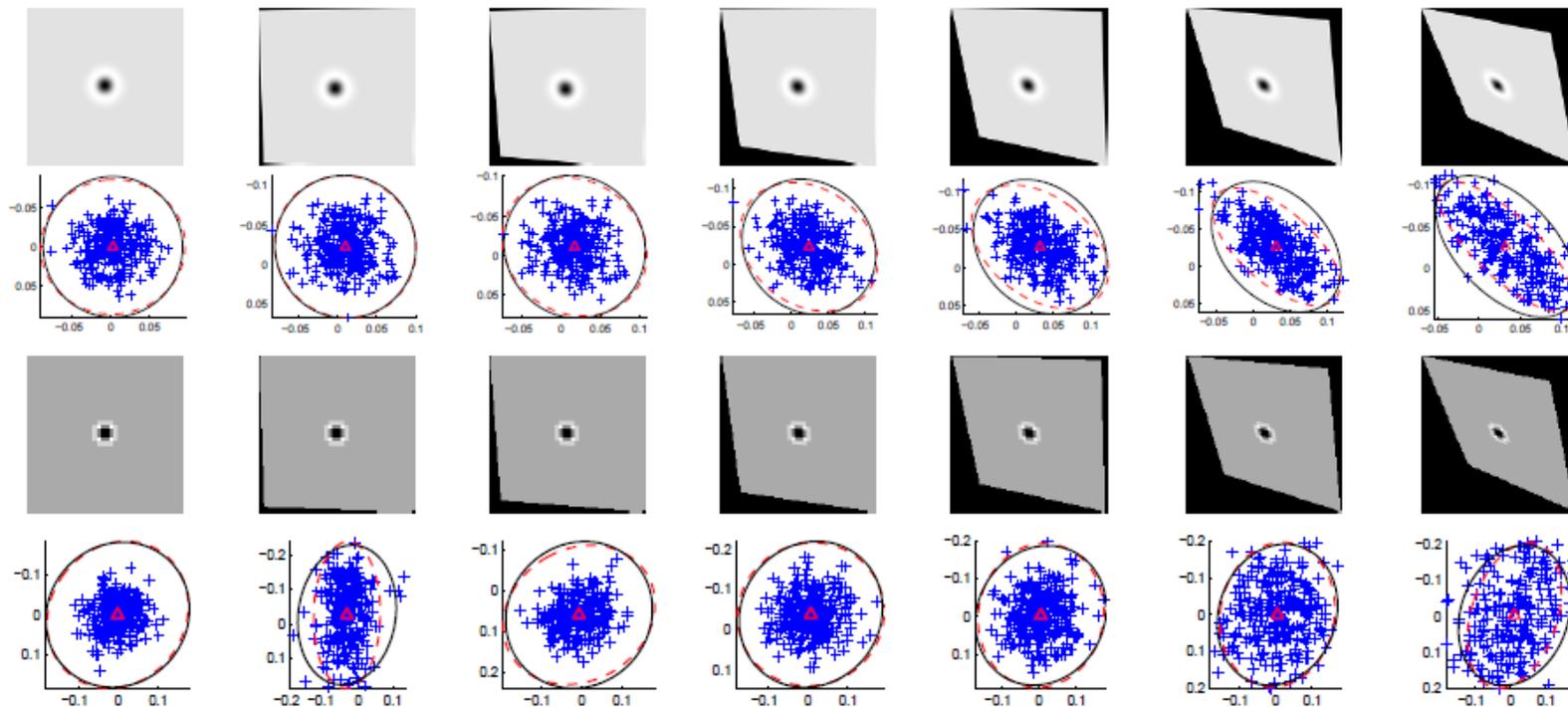
$$\hat{\mathbf{I}} = \arg \max_{\mathbf{I}} \mathbf{I}^\top \underbrace{(\mathbf{f}_{xx}\mathbf{f}_{yy}^\top - \mathbf{f}_{xy}\mathbf{f}_{xy}^\top)}_{\mathbf{A}} \mathbf{I}$$

*s.t.*  $\|\hat{\mathbf{I}}\| = 1$

Location and scale where feature will be detected is **controllable**.

# Statistical Error Modeling

Maximum likelihood estimate and our covariance coincide

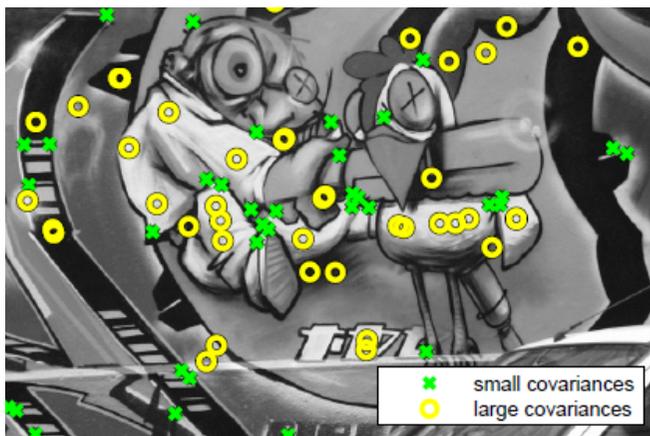
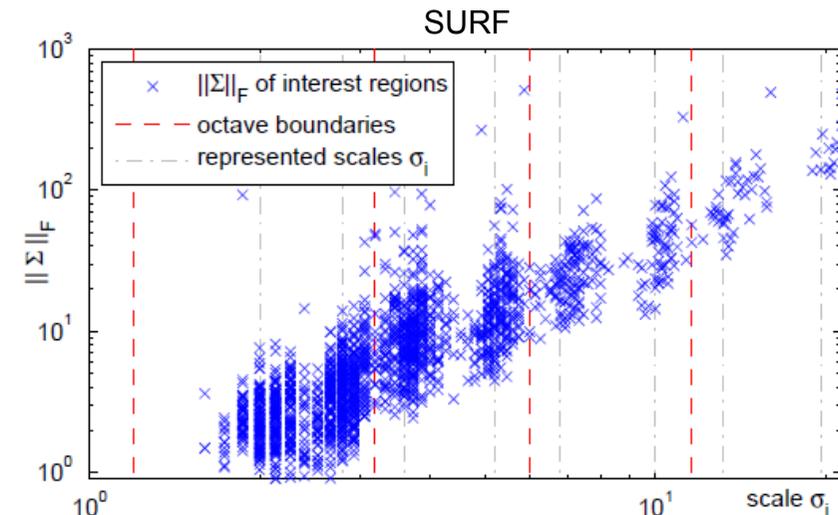
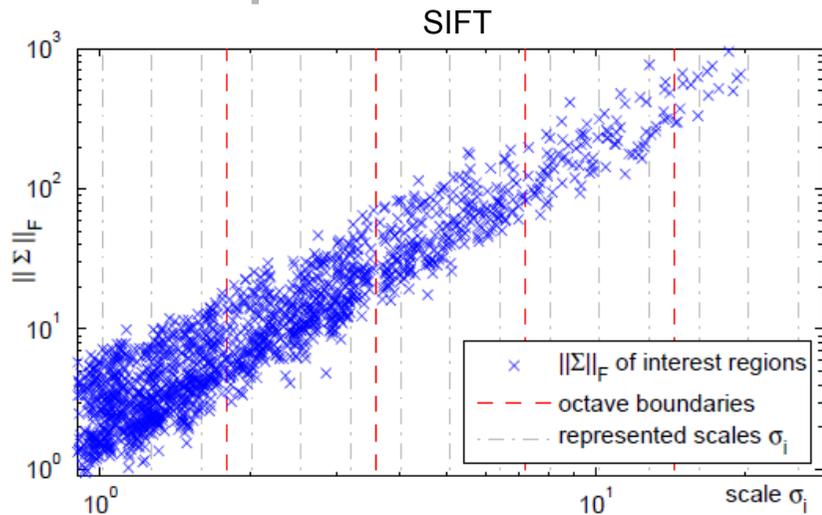


(+) distribution of location error      (--) covariance estimate      (- -) maximum likelihood estimate

The **covariance estimates** fit the modeled error distribution

# 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.

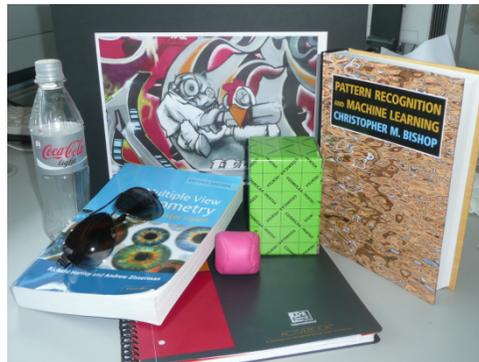
Feature points with small ( $\sigma < 2.1$ ) or large ( $\sigma > 8$ ) covariances.

Blobs are worse localized than distinctive image points.

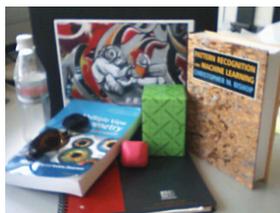
# Covariance Dependence on Scale

Quality of image has no influence on the localization accuracy

High and low resolution images:

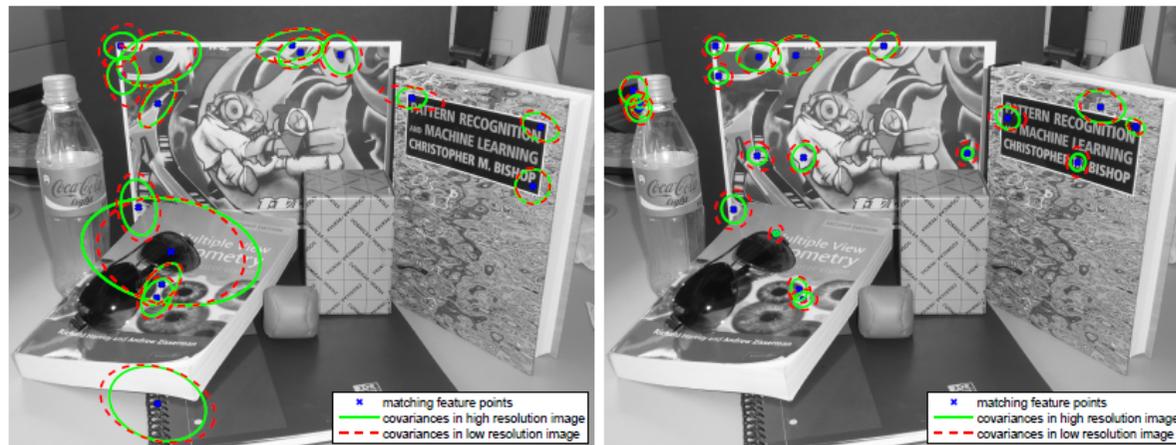


3072x2304 pixel



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;  
however particular feature shapes are equally sized  
→ Localization error is similar in both images in relation to their size

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

# Covariance Dependence on Projective Transformations

## Projective distortions modify covariance shape and scale

Tracking of a feature points in an image sequence and estimation of covariance matrix in each frame

Test criteria: stability

→ expected outcome: smooth change of covariance according to viewpoint change between frames.

Zoom



Covariances are bigger the larger the feature is in the image

Perspective transform.



Covariances are horizontally compressed due to the perspective distortion

Rotation

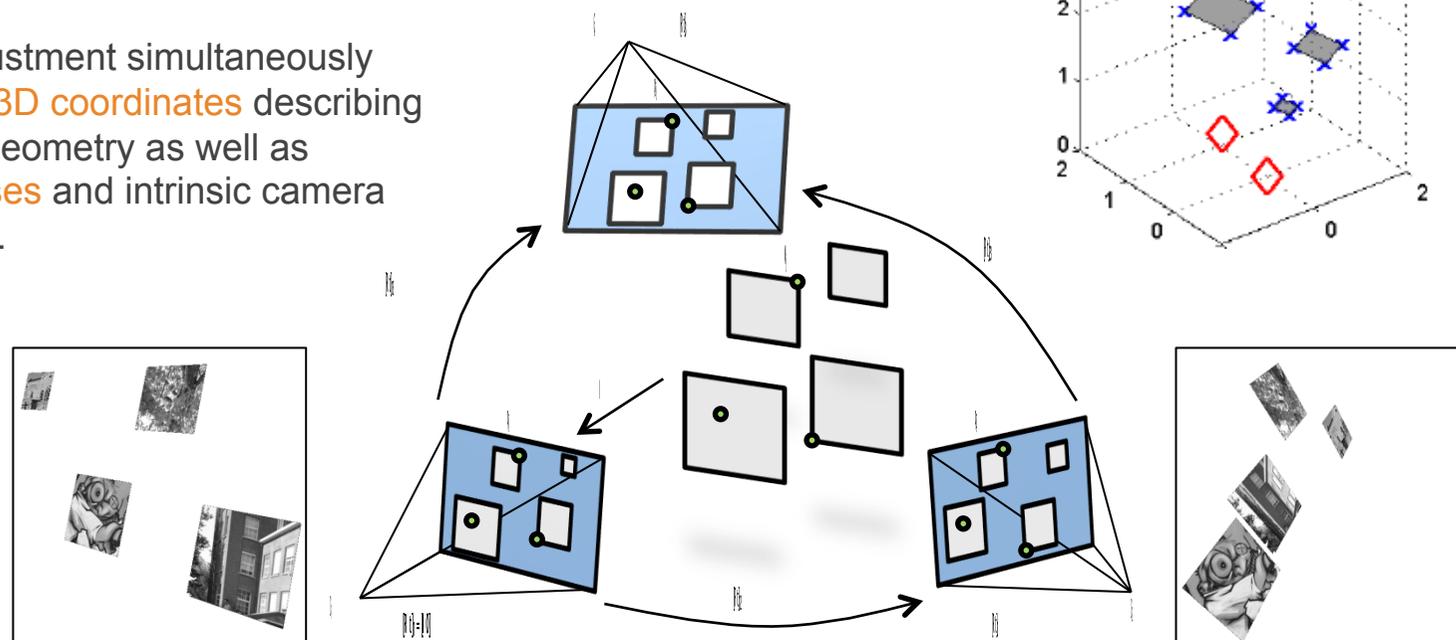


Covariances follow the rotation of the pattern

# 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.

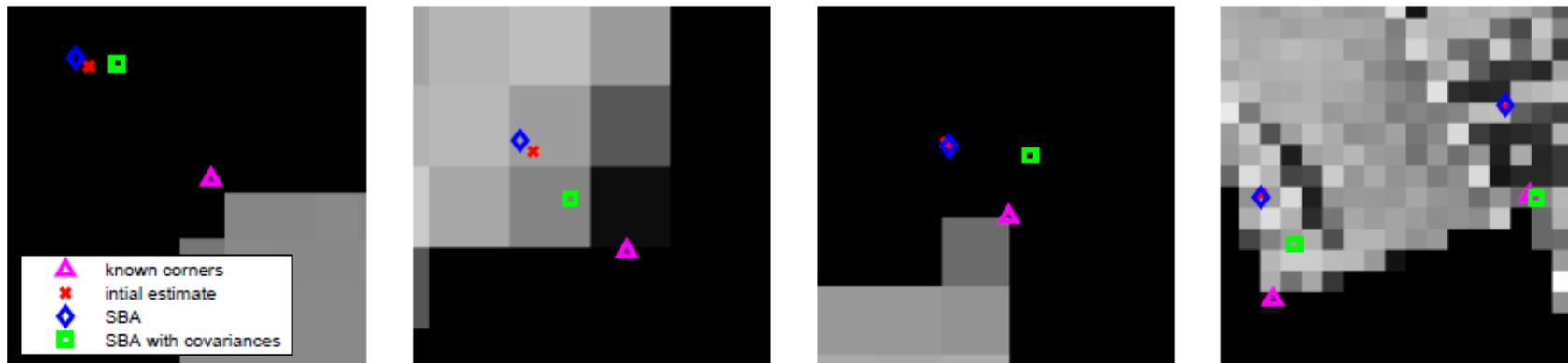


Euclidian distance: 
$$\arg \min_{[\mathbf{R} \ \mathbf{t}]_i, \mathbf{M}_j} \sum_{i=1}^N \sum_{j=1}^m v_{ij} \cdot (\mathbf{p}_{ij} - w(\mathbf{K}[\mathbf{R} \ \mathbf{t}]_i \mathbf{M}_j))^T (\mathbf{p}_{ij} - w(\mathbf{K}[\mathbf{R} \ \mathbf{t}]_i \mathbf{M}_j))$$

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

# Bundle Adjustment

Performance is evaluated with the reprojection error of corner points



Reprojection error of 3D corner points:

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

Mean performance as pixel offset for about 100 different image pairs

	mean all patches		smallest patch		largest patch	
	no	yes	no	yes	no	yes
SIFT	2.031	1.759	1.941	1.672	2.088	1.828
SURF	2.554	2.363	2.518	2.292	2.631	2.464

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

## I did not do this work alone



*Pierre Schroeder*



*Selim Benhimane*



*Juergen Sotke*



*Stefan Hinterstoiser*



*Stefan Holzer*

But also

Adrien Bartoli (universite d'Auvergne)

Bernard Ziesl

Mirko Appel (siemens CT)

Sumit Paranjape, Xinxing Feng, Xavier Fernandez, Hagen Kaiser

## Conclusion and Perspectives

### About this method

- General formulation for feature detection in scale space
- Automatic computation of covariance
- Justification of correctness for the covariance estimates
- Performance improvement for bundle adjustment

### About the application

- Landmark based registration
- Automatic 6 DoF stereo registration
- Bundle Adjustment with fix 3D points
- 3D Measurements
- Inspection / Documentation

## Future Work

- Further registration evaluation
- Comparison to methods for Harris
- Improvement for Matching?
- Scale stable corner?
- Mobile AR using keyframes
- Guided 3D reconstruction
- 3D Annotations



**Thank for the attention**

(Any) **Questions?**

# References

## Key papers.

- Bay, H. and Tuytelaars, T. and Van Gool, L., *SURF: Speeded Up Robust Features*, Computer Vision and Image Understanding (CVIU), 2008
- Brooks, MJ, et. al., *What value covariance information in estimating vision parameters?*, ICCV 2001
- Kanazawa, Y., Kanatani, K., *Do we really have to consider covariance matrices for image features?*, ICCV 2001
- Lindeberg, T, *Scale-Space: A framework for handling image structures at multiple scales*, In Proc. CERN School of Computing, 1996
- Lowe, D., *Distinctive image features from scale-invariant keypoints*, International Journal of Computer Vision, 2004
- Matas, J. et. al., *Robust wide-baseline stereo from maximally stable extremal regions*, Image & Vision Computing, 2004
- Mikolajczyk, K., Schmid, C., *An affine invariant interest point detector*, ICCV 2002
- Mikolajczyk, K., Schmid, C., *Scale & Affine Invariant Interest Point Detectors*, Intern. Journal of Computer Vision, 2004
- Ochoa, B. and Belongie, S., *Covariance Propagation for Guided Matching*, 2006
- Park, Y., Lepetit, V., Woo, W., *Multiple 3D Object Tracking for Augmented Reality*, ISMAR 2008
- Schmid, C., Mohr, R., Bauckhage, C., *Evaluation of Interest Point Detectors*, Intern. Journal of Computer Vision, 2000
- Steele, R.M., Jaynes, C., *Feature Uncertainty Arising from Covariant Image Noise*, Proc. of IEEE CVPR 2005
- Sur, F., Noury, N., Berger, M., *Computing the uncertainty of the 8 point algorithm for fundamental matrix estimation*, Proc. of the British Machine Vision Conference (BMVC), 2008
- Triggs, B. et.al., *Bundle Adjustment – A Modern Synthesis*, Proc. of the Intern. Workshop on Vision Algorithms, ICCV 1999