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 shipped last month
Photo-based Augmented Reality

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
Common assumptions for detected local features:

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

<table>
<thead>
<tr>
<th>Purpose</th>
<th>Detector</th>
<th>Descriptor</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Localization</strong> of feature points</td>
<td>Spatial detection done via a mathematical operator e.g.: DoG, det(H), trace(H)</td>
<td>Description by the structure of the local neighborhood e.g.: weighted sampling of gradient field</td>
</tr>
<tr>
<td><strong>Desired properties</strong></td>
<td>- 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 - Localization precision</td>
<td>- Distinctiveness - Robustness to occlusions - Invariance to image transformations. Reached by appropriate size and sampling of the gradient field</td>
</tr>
</tbody>
</table>

- **Stability** – detection independent of changes in the image conditions
- **Localization precision**
Image Feature Detection
Corners and blobs in an image are complementary features

<table>
<thead>
<tr>
<th>Feature definition</th>
<th>Corner Detector</th>
<th>Blob Detector</th>
</tr>
</thead>
<tbody>
<tr>
<td>Detector</td>
<td>Harris corner detector</td>
<td>MSER detector</td>
</tr>
<tr>
<td>Location where two dominant directions intersect</td>
<td>Areas brighter or darker than the surrounding.</td>
<td>localizes blobs by a well-defined point - the blob center.</td>
</tr>
</tbody>
</table>
## 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.
- 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 $X$ 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 (1st or 2nd derivatives).
$\rightarrow$ Additional error introduced for the feature point $\hat{X}$ depending on the algorithm.

$\hat{X} = \hat{x} - x$
$\hat{x} = \text{dect}(I(x) + \eta)$
$x \ldots \text{ground truth feature point}$
$I(x) \ldots \text{intensity value at } x$
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.
- 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.
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.
- **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.

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.**
Scale Invariant Feature Detection
The same feature can be detected at different scales

Scale Space Representation

Characteristic Scale Selection

Mikolajczyk, K., Schmid, C., Scale & Affine Invariant Interest Point Detectors, 2004
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 $f_{\text{dec}}$.
  \[ D(x, \sigma_i) = f_{\text{dec}}(I(N_x), \sigma_i) \]
- Candidate set $\mathbb{P}_1$ of feature points $\langle p, \sigma_i \rangle$ by local maxima detection in $D(\bullet, \sigma_i)$.
  \[ \mathbb{P}_1 := \bigcup_{i=1}^{N} \left\{ \langle p, \sigma_i \rangle \mid p = \arg \max_{x \in N_p} D(x, \sigma_i) \right\} \]

**Scale Selection**
- Scale space representation of the image with scale selection operator $f_{\text{sel}}$.
  \[ S(p, \sigma_i) = f_{\text{sel}}(I(N_p), \sigma_i) \]
- Set $\mathbb{P}_2$ of feature points $\langle p, \sigma \rangle$ by local maxima detection at positions $p$ in $S(p, \sigma_i)$.
  \[ \mathbb{P}_2 := \left\{ \langle p, \sigma \rangle \mid \langle p, \sigma \rangle \in \mathbb{P}_1, \sigma = \arg \max_{\sigma_i} S(p, \sigma_i), S(p, \sigma) > \tau \right\} \]

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:
\[ R(\Delta p) = |D(p, \sigma) - D(p + \Delta p, \sigma)| \]
\[ p = \arg \max_{x \in N_p} D(x, \sigma) = \arg \min_{\Delta p \in N_0} R(\Delta p) \]
\[ R(\Delta p) \approx \tilde{R}(\Delta p) = \frac{1}{2} \Delta p^T H \Delta p \]

Covariance based on Hessian:
\[ \Sigma = H^{-1} = \left[ \begin{array}{cc} D_{xx}(p, \sigma) & D_{xy}(p, \sigma) \\ D_{xy}(p, \sigma) & D_{yy}(p, \sigma) \end{array} \right]^{-1} \]

low curvature \(\rightarrow\) error due to the missing discriminative behavior of \(D(\bullet, \sigma_0)\) in \(N_P\).

high curvature \(\rightarrow\) detection process more accurate
Framework Application
Application is identically for SIFT and SURF

Detector function

\[
D(x, \sigma_i) = \frac{G(x, \sigma_{i+1}) - G(x, \sigma_i)}{\nabla^2 G(x, \sigma_i)} * I(x)
\]

\[
D(x, \sigma_i) = \text{det} \begin{bmatrix}
L_{xx}(x, \sigma_i) & L_{xy}(x, \sigma_i) \\
L_{xy}(x, \sigma_i) & L_{yy}(x, \sigma_i)
\end{bmatrix}
\]

Covariance calculation

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

\[
D_{xx} = d_{xx} \cdot D(N_p, \sigma_i)
\]

\[
\Sigma^{(0)} = \Sigma \cdot (2^{\text{octave}})^2
\]

Back projection

Scale calculation

\[
S = b \cdot 2^{\text{octave + interval}/N_{\text{intervals}}}
\]

\[
S = \sum_s^b
\]
Optimal Images for SIFT and SURF

Image feature for maximal detector response are computable

**SIFT**

DoG is a linear filter

→ matched filter approach

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

**SURF**

det(H) is non-linear detector:

→ Optimization via quadratically constrained quadratic program

\[ \det H = \det \begin{bmatrix} h_{xx} & h_{xy} \\ h_{yx} & h_{yy} \end{bmatrix} \]

\[ h_{xx} = f_{xx}^\top I \]

\[ h_{xy} = h_{yx} = f_{xy}^\top I \]

\[ h_{yy} = f_{yy}^\top I. \]

\[ \hat{I} = \arg\max_I I^\top \left( f_{xx}f_{yy}^\top - f_{xy}f_{xy}^\top \right) I \]

\[ s.t. \ |\hat{I}| = 1 \]

Location and scale where feature will be detected is controllable.
Statistical Error Modeling

Maximum likelihood estimate and our covariance coincide

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

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 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: $\arg\min_{[R, t], M} \sum_{i=1}^{N} \sum_{j=1}^{m} v_{ij} \cdot (p_{ij} - w(K[R, t]_i M_j)) \sum_{ij}^{\text{T}} (p_{ij} - w(K[R, t]_i M_j))$

Mahalanobis distance: $\arg\min_{[R, t], M} \sum_{i=1}^{N} \sum_{j=1}^{m} v_{ij} \cdot (p_{ij} - w(K[R, t]_i M_j)) \Sigma_{ij}^{-1}(p_{ij} - w(K[R, t]_i M_j))$
Bundle Adjustment
Performance is evaluated with the reprojection error of corner points

Reprojection error of 3D corner points:

\[ e = \frac{1}{4} \sum_{i=1}^{4} \left\| \overline{c}_i - w(\mathbf{Tc}_i) \right\| \]

Mean performance as pixel offset for about 100 different image pairs

<table>
<thead>
<tr>
<th>covariance usage</th>
<th>mean all patches</th>
<th>smallest patch</th>
<th>largest patch</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>no</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>SIFT</td>
<td>2.031</td>
<td>1.759</td>
<td>1.941</td>
</tr>
<tr>
<td>SURF</td>
<td>2.554</td>
<td>2.363</td>
<td>2.518</td>
</tr>
</tbody>
</table>

We get a performance improvement for the reconstruction with bundle adjustment using our feature point covariances.
I did not do this work alone

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

Code and binaries for SIFT and SURF local feature detection and covariance estimation:
http://campar.in.tum.de/Main/CovarianceEstimator
Thank for the attention

(Any) Questions?

http://wwwnavab.in.tum.de/Main/PierreGeorgel - Pierre.Georgel@gmail.com
References
Key papers.

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
Matas, J. et. al., *Robust wide-baseline stereo from maximally stable extremal regions*, Image & Vision Computing, 2004