Optimization methods to calibrate a stereo rig with increased accuracy for vehicular applications

András Bódis-Szomorú, Tamás Dabóczi

bodis@mit.bme.hu
http://www.mit.bme.hu/~bodis

Budapest University of Technology and Economics
Dept. of Measurement and Information Systems
Budapest, Hungary

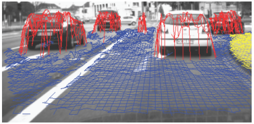
I2MTC Graz, May 14, 2012

Outline

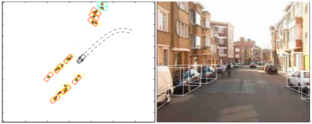
Outline

1. Introduction
2. Max. likelihood solution of pose estimation
3. Improve inter-camera pose
4. Rig pose estimation
5. Summary

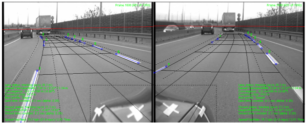
Computer Vision in Advanced Driver Assistance Systems



Nedevschi et al. 2008, TU Cluj-Napoca



Cornelis et al. IJCV 2008, ETH Zurich/KU Leuven



Bódis et al. 2009, TU Budapest

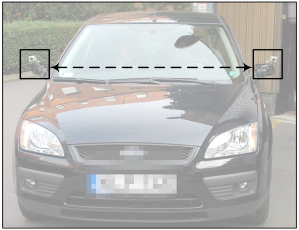
Vision tasks in ADAS

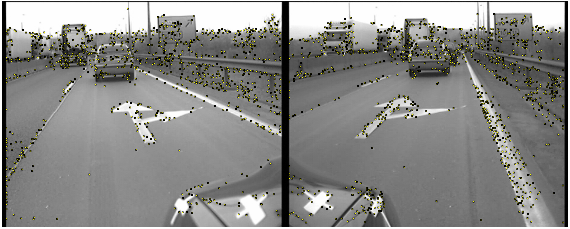
- Low-level 3D reconstruction (SfM, stereo)
- Drivable region/obstacle detection from geometry (obstacle, ground, facade...)
- Lane detection
- Object detection and recognition (pedestrian, vehicle, traffic sign...)
- Tracking (obstacle, lane, vehicle...)

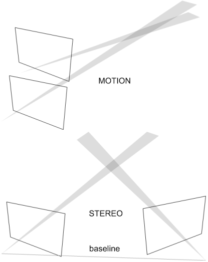
References (Surveys)

- *M. Bertozzi et al.*, **Artificial Vision in Road Vehicles**, *Proceedings of the IEEE, 90(7):1258–1271, 2002*
- *V. Kastrinaki, M. Zervakis, K. Kalaitzakis,* **A survey of video processing techniques for traffic applications**, *Image and Vision Computing, 21(4):359–381, 2003*
- *Z. Sun, G. Bebis, and R. Miller,* **On-road vehicle detection: A review**,*IEEE Trans. on Pattern Analysis and Machine Intelligenece, 28(5):694–711, 2006*

Focus of interest: wide-baseline stereo



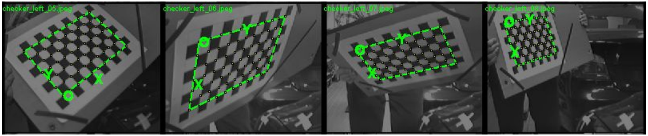




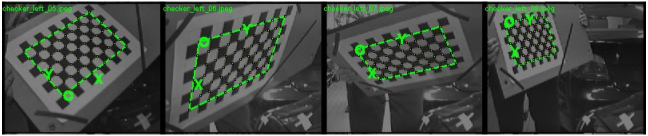
Properties of wide-baseline stereo

- stereo ⇒ instantaneous depth capture
- more suitable for far-range than single-view SfM
- difficult matching problem
- accurate calibration is crucial at any time

1. Offline intrinsic calibration (lens, sensor) $\Leftarrow$ planar pattern

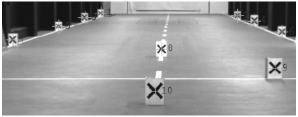


1. Offline intrinsic calibration (lens, sensor) $\Leftarrow$ planar pattern

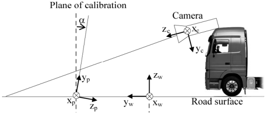


2. Offline pose estimation (position, orientation) $\Leftarrow$ planar arrangement/pattern

Broggi et al. ICRA 2001, Univ.Parma



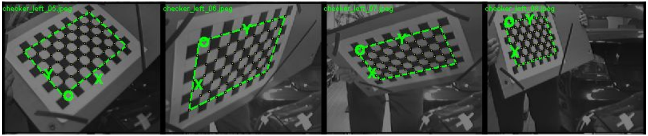
Marita et al. IVS 2006, TU Cluj-Napoca



Bellino et al. ITSC 2005, EPFL

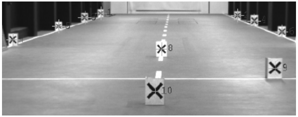
Typical calibration steps today

1. Offline intrinsic calibration (lens, sensor) $\Leftarrow$ planar pattern

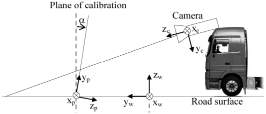


2. Offline pose estimation (position, orientation) $\Leftarrow$ planar arrangement/pattern

Broggi et al. ICRA 2001, Univ.Parma



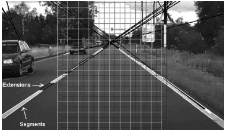
Marita et al. IVS 2006, TU Cluj-Napoca



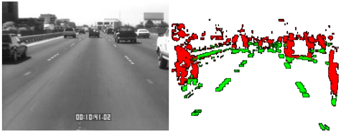
Bellino et al. ITSC 2005, EPFL

3. Online self-checking / re-estimation $\Leftarrow$ markers, disparities, flatness, rigidity, tracking

Broggi et al. ICRA 2001, Univ.Parma



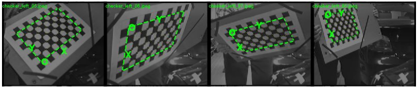
Nedevschi et al. ITSC 2006, TU Cluj-Napoca



Weber et al. IVS 1995, UC Berkeley

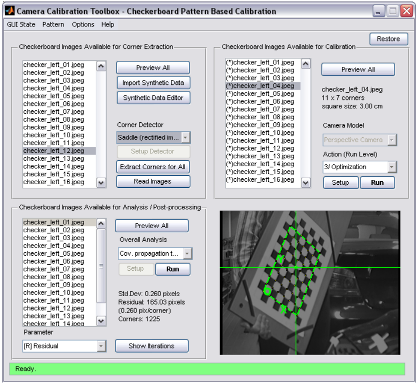
Outline

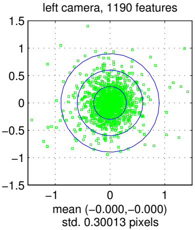
Camera model: pinhole projection + radial lens distortion model
Intrinsic calibration $\Leftarrow$ checkerboard dataset (16 poses per camera)

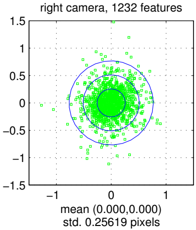


Min. reprojection error w.r.t
9 intrinsic params & 6-DoF board poses

$$\mathcal{C}_{int}(\mathbf{p}) = \sum_{j=1}^{b} \sum_{i=1}^{n_j} d^2(\mathbf{m}_i^j, \varphi(\mathbf{p}^j, \mathbf{M}_i)) \longrightarrow \min_{\mathbf{p}}$$

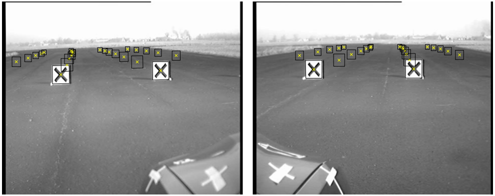


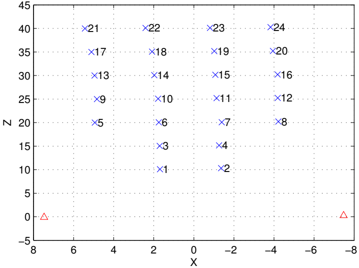




http://www.mit.bme.hu/~bodis/ccalgui.html

Camera poses w.r.t. vehicle $\Leftarrow$ far range arrangement: X-markers dataset

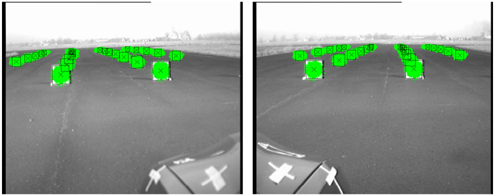


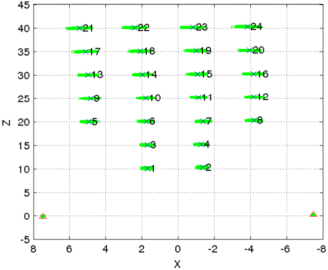


$$\mathcal{C}_{ML}(\mathbf{a}_l, \mathbf{t}_l, \mathbf{a}_r, \mathbf{t}_r, \hat{\mathbf{M}}) = \overbrace{\frac{1}{\sigma^2}||\bar{\mathbf{m}} - \hat{\mathbf{m}}||_2^2}^{\textit{error in the images}} + \overbrace{||\bar{\mathbf{M}} - \hat{\mathbf{M}}||_{\mathbf{\Sigma}}^2}^{\textit{error in 3D space}} \rightarrow \min_{\mathbf{a}_l, \mathbf{t}_l, \mathbf{a}_r, \mathbf{t}_r, \hat{\mathbf{M}}}$$

Requirement: error covariances σ, $\mathbf{\Sigma}$

Camera poses w.r.t. vehicle $\Leftarrow$ far range arrangement: X-markers dataset

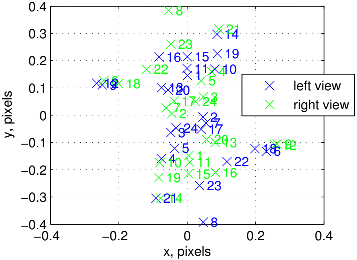


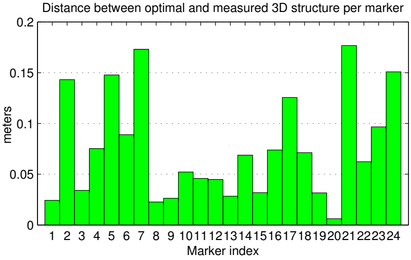


$$\mathcal{C}_{ML}(\mathbf{a}_l, \mathbf{t}_l, \mathbf{a}_r, \mathbf{t}_r, \hat{\mathbf{M}}) = \overbrace{\frac{1}{\sigma^2}||\bar{\mathbf{m}} - \hat{\mathbf{m}}||_2^2}^{\textit{error in the images}} + \overbrace{||\bar{\mathbf{M}} - \hat{\mathbf{M}}||_{\boldsymbol{\Sigma}}^2}^{\textit{error in 3D space}} \rightarrow \min_{\mathbf{a}_l, \mathbf{t}_l, \mathbf{a}_r, \mathbf{t}_r, \hat{\mathbf{M}}}$$

Requirement: error covariances σ, $\boldsymbol{\Sigma}$

Residual errors in the images and in 3D





$$\mathcal{C}_{ML}(\mathbf{a}_l, \mathbf{t}_l, \mathbf{a}_r, \mathbf{t}_r, \hat{\mathbf{M}}) = \overbrace{\frac{1}{\sigma^2}||\bar{\mathbf{m}} - \hat{\mathbf{m}}||_2^2}^{error\ \ in\ \ the\ \ images} + \overbrace{||\bar{\mathbf{M}} - \hat{\mathbf{M}}||_{\mathbf{\Sigma}}^2}^{error\ \ in\ \ 3D\ \ space} \rightarrow \min_{\mathbf{a}_l, \mathbf{t}_l, \mathbf{a}_r, \mathbf{t}_r, \hat{\mathbf{M}}}$$

Requirement: error covariances σ, $\mathbf{\Sigma}$

Outline

Problem: small number of features $\Rightarrow$ likely to overfit

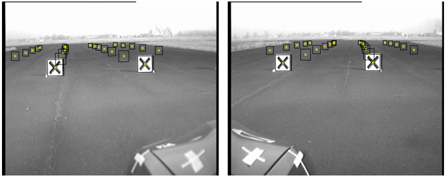
Key ideas

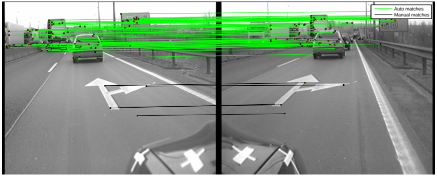
Extension: Decoupled pose estimation

Problem: small number of features ⇒ likely to overfit

Key ideas

- use many stereo matches from on-line videos ⇒ new dataset



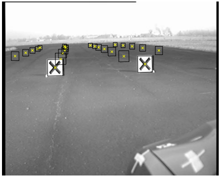


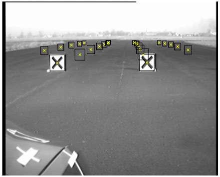
Extension: Decoupled pose estimation

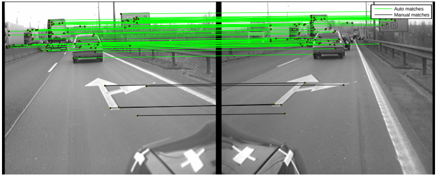
Problem: small number of features $\Rightarrow$ likely to overfit

Key ideas

- use many stereo matches from on-line videos $\Rightarrow$ new dataset
- validation: use the new dataset to evaluate earlier results



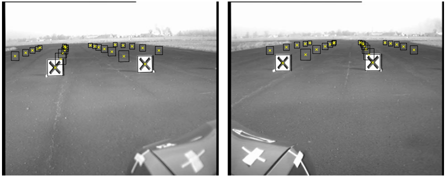


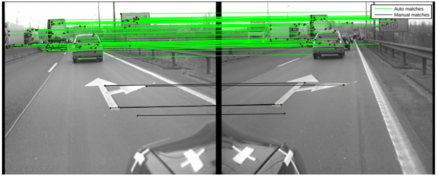


Problem: small number of features $\Rightarrow$ likely to overfit

Key ideas

- use many stereo matches from on-line videos $\Rightarrow$ new dataset
- validation: use the new dataset to evaluate earlier results
- decoupling: inter-camera pose $\{\mathbf{R}, \mathbf{t}\}$ vs. rig pose $\{\mathbf{R}_r, \mathbf{t}_r\}$



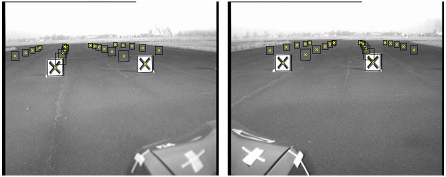


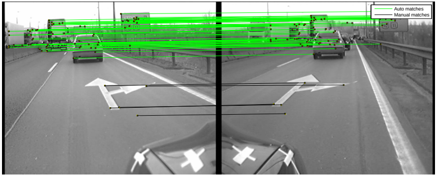
Extension: Decoupled pose estimation

Problem: small number of features $\Rightarrow$ likely to overfit

Key ideas

- use many stereo matches from on-line videos $\Rightarrow$ new dataset
- validation: use the new dataset to evaluate earlier results
- decoupling: inter-camera pose $\{\mathbf{R}, \mathbf{t}\}$ vs. rig pose $\{\mathbf{R}_r, \mathbf{t}_r\}$
- new dataset $\Rightarrow$ $\{\mathbf{R}, \mathbf{t}\}$ up to scale



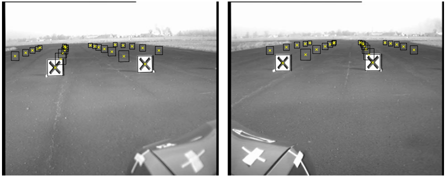


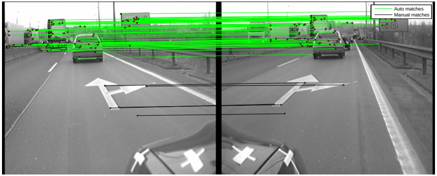
Extension: Decoupled pose estimation

Problem: small number of features ⇒ likely to overfit

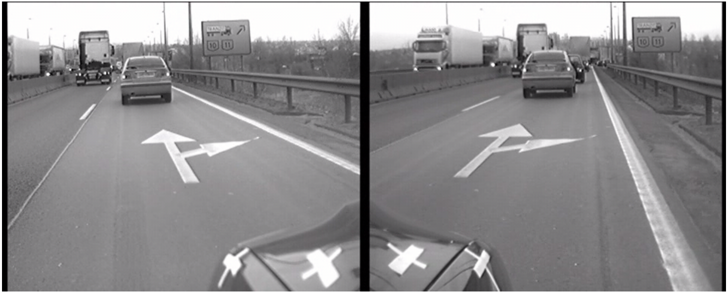
Key ideas

- use many stereo matches from on-line videos ⇒ new dataset
- validation: use the new dataset to evaluate earlier results
- decoupling: inter-camera pose $\{\mathbf{R}, \mathbf{t}\}$ vs. rig pose $\{\mathbf{R}_r, \mathbf{t}_r\}$
- new dataset ⇒ $\{\mathbf{R}, \mathbf{t}\}$ up to scale
- X-markers ⇒ global scale λ and rig-to-world pose $\{\mathbf{R}_r, \mathbf{t}_r\}$



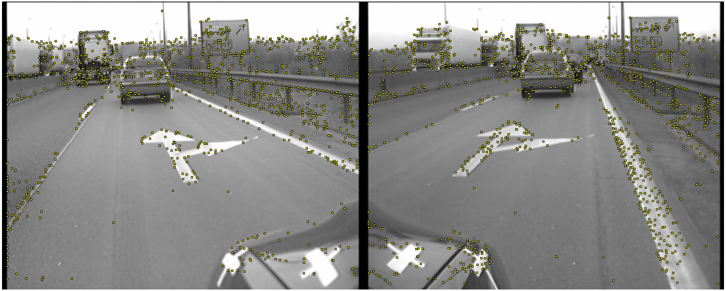


Inter-camera pose from new on-line dataset



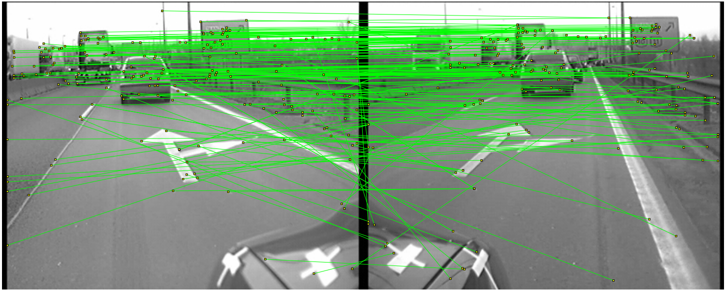
Inter-camera pose from new on-line dataset

1. Feature detection (SIFT of A. Vedaldi)



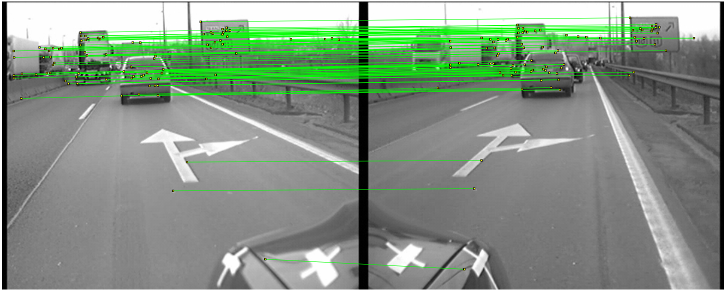
Inter-camera pose from new on-line dataset

1. Feature detection (SIFT of A. Vedaldi)
2. Automatic stereo matching



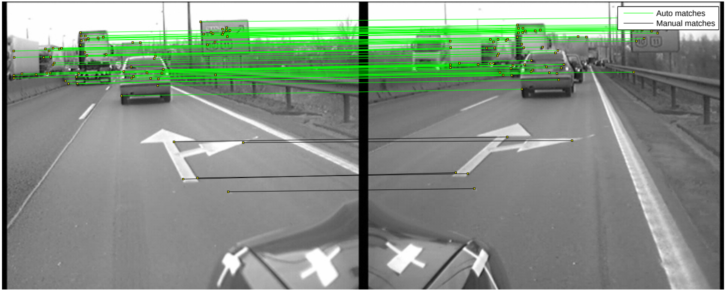
Inter-camera pose from new on-line dataset

1. Feature detection (SIFT of A. Vedaldi)
2. Automatic stereo matching
3. Outlier removal (RANSAC)



Inter-camera pose from new on-line dataset

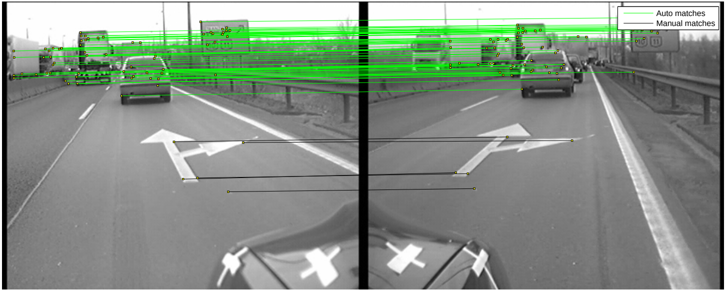
1. Feature detection (SIFT of A. Vedaldi)
2. Automatic stereo matching
3. Outlier removal (RANSAC)
4. Remove all residual outliers manually ⇒ 2000 good matches in total



Computing inter-camera pose

Inter-camera pose from new on-line dataset

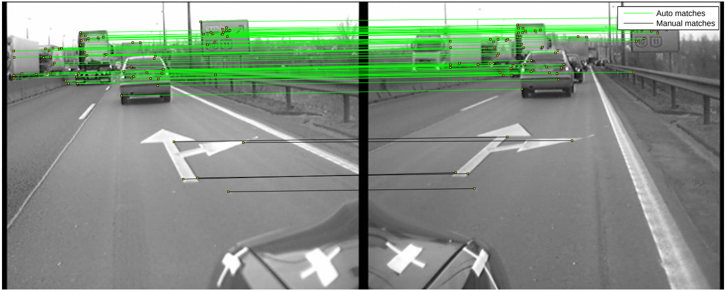
1. Feature detection (SIFT of A. Vedaldi)
2. Automatic stereo matching
3. Outlier removal (RANSAC)
4. Remove all residual outliers manually $\Rightarrow$ 2000 good matches in total
5. Compute: $\mathbf{F} \Rightarrow \mathbf{E} \Rightarrow$ inter-camera pose $\{\mathbf{R}, \mathbf{t}\}$



Computing inter-camera pose

Inter-camera pose from new on-line dataset

1. Feature detection (SIFT of A. Vedaldi)
2. Automatic stereo matching
3. Outlier removal (RANSAC)
4. Remove all residual outliers manually $\Rightarrow$ 2000 good matches in total
5. Compute: $\mathbf{F} \Rightarrow \mathbf{E} \Rightarrow$ inter-camera pose $\{\mathbf{R}, \mathbf{t}\}$
6. Triangulate $\Rightarrow$ min. reprojection errors w.r.t. $\{\mathbf{R}, \mathbf{t}\}$

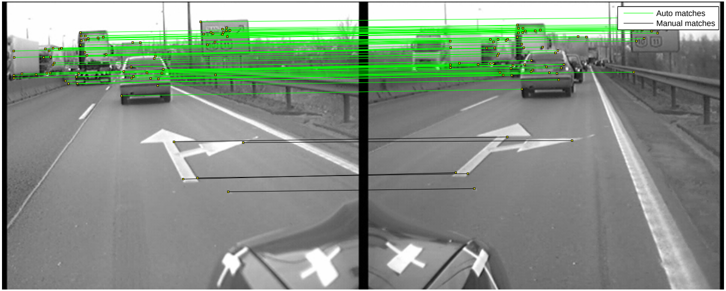


Computing inter-camera pose

Inter-camera pose from new on-line dataset

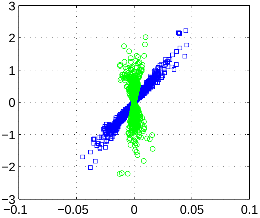
1. Feature detection (SIFT of A. Vedaldi)
2. Automatic stereo matching
3. Outlier removal (RANSAC)
4. Remove all residual outliers manually $\Rightarrow$ 2000 good matches in total
5. Compute: $\mathbf{F} \Rightarrow \mathbf{E} \Rightarrow$ inter-camera pose $\{\mathbf{R}, \mathbf{t}\}$
6. Triangulate $\Rightarrow$ min. reprojection errors w.r.t. $\{\mathbf{R}, \mathbf{t}\}$

Result: $\{\mathbf{R}, \mathbf{t}\}$ up to scale, $||\mathbf{t}|| = \lambda$ unknown

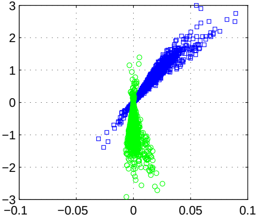


Validation: Epipolar errors in the SIFT-dataset (2000 matches)

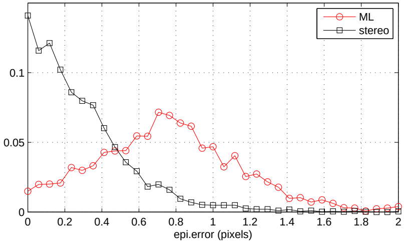
F from SIFT/stereo dataset



F from X dataset (ML)



Comparison (histogram)



Rig pose from far range arrangement

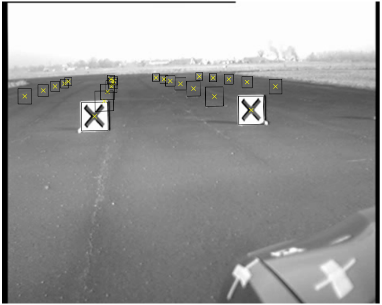
Fixed: inter-camera pose $\{\mathbf{R}, \mathbf{t}\}$ up to scale, ($||\mathbf{t}|| = \lambda$ unknown)
Estimate: scale λ and rig pose $\{\mathbf{R}_r, \mathbf{t}_r\}$

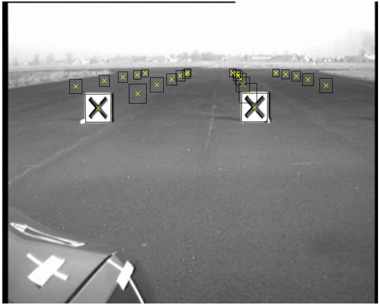
Rig pose from far range arrangement

Fixed: inter-camera pose $\{\mathbf{R}, \mathbf{t}\}$ up to scale, ($||\mathbf{t}|| = \lambda$ unknown)
Estimate: scale λ and rig pose $\{\mathbf{R}_r, \mathbf{t}_r\}$

Method 1: reprojection error $\sum_i ||\mathbf{m}_i - \hat{\mathbf{m}}_i||^2 \rightarrow \min_{\{\mathbf{R}_r, \mathbf{t}_r, \lambda\}}$

Method 2: 3D registration error $\sum_i ||\mathbf{M}_i - \hat{\mathbf{M}}_i||^2 \rightarrow \min_{\{\mathbf{R}_r, \mathbf{t}_r, \lambda\}}$



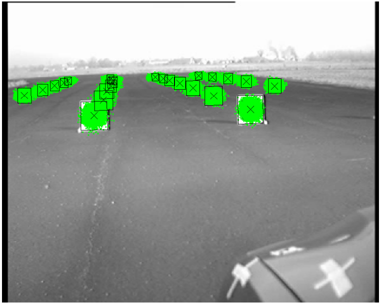


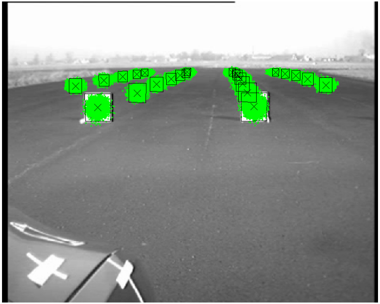
Rig pose from far range arrangement

Fixed: inter-camera pose $\{\mathbf{R}, \mathbf{t}\}$ up to scale, $(||\mathbf{t}|| = \lambda$ unknown$)$
Estimate: scale λ and rig pose $\{\mathbf{R}_r, \mathbf{t}_r\}$

Method 1: reprojection error $\sum_i ||\mathbf{m}_i - \hat{\mathbf{m}}_i||^2 \rightarrow \min_{\{\mathbf{R}_r, \mathbf{t}_r, \lambda\}}$

Method 2: 3D registration error $\sum_i ||\mathbf{M}_i - \hat{\mathbf{M}}_i||^2 \rightarrow \min_{\{\mathbf{R}_r, \mathbf{t}_r, \lambda\}}$

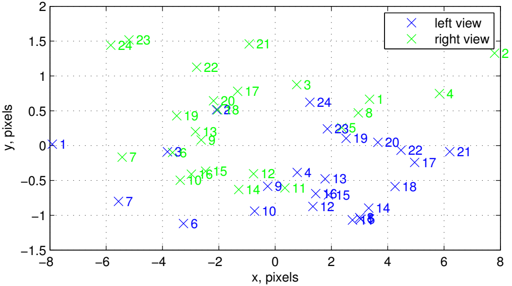




Fixed: inter-camera pose $\{\mathbf{R}, \mathbf{t}\}$ up to scale, ($||\mathbf{t}|| = \lambda$ unknown)
Estimate: scale λ and rig pose $\{\mathbf{R}_r, \mathbf{t}_r\}$

Method 1: reprojection error $\sum_i ||\mathbf{m}_i - \hat{\mathbf{m}}_i||^2 \rightarrow \min_{\{\mathbf{R}_r, \mathbf{t}_r, \lambda\}}$

Method 2: 3D registration error $\sum_i ||\mathbf{M}_i - \hat{\mathbf{M}}_i||^2 \rightarrow \min_{\{\mathbf{R}_r, \mathbf{t}_r, \lambda\}}$



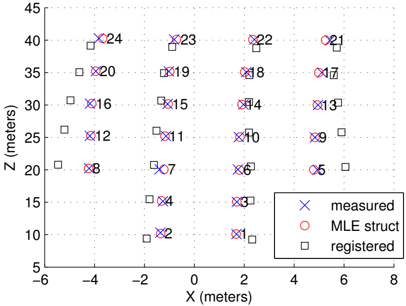
Result (Method 1): Residuals are ± 8 pixels $\Rightarrow$ too high!

Rig pose from far range arrangement

Fixed: inter-camera pose $\{\mathbf{R}, \mathbf{t}\}$ up to scale, ($||\mathbf{t}|| = \lambda$ unknown)
Estimate: scale λ and rig pose $\{\mathbf{R}_r, \mathbf{t}_r\}$

Method 1: reprojection error $\sum_i ||\mathbf{m}_i - \hat{\mathbf{m}}_i||^2 \rightarrow \min_{\{\mathbf{R}_r, \mathbf{t}_r, \lambda\}}$

Method 2: 3D registration error $\sum_i ||\mathbf{M}_i - \hat{\mathbf{M}}_i||^2 \rightarrow \min_{\{\mathbf{R}_r, \mathbf{t}_r, \lambda\}}$



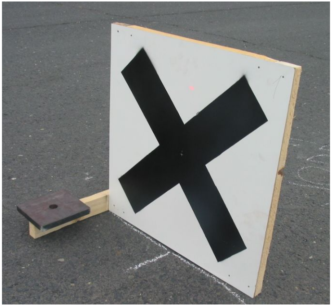
Result (Method 2): residual errors 1.0 meters RMS (1.4 meters max)

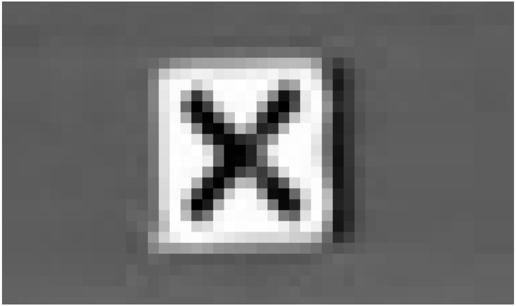
Rig pose from far range arrangement

Fixed: inter-camera pose $\{\mathbf{R}, \mathbf{t}\}$ up to scale, ($||\mathbf{t}|| = \lambda$ unknown)
Estimate: scale λ and rig pose $\{\mathbf{R}_r, \mathbf{t}_r\}$

Method 1: reprojection error $\sum_i ||\mathbf{m}_i - \hat{\mathbf{m}}_i||^2 \rightarrow \min_{\{\mathbf{R}_r, \mathbf{t}_r, \lambda\}}$

Method 2: 3D registration error $\sum_i ||\mathbf{M}_i - \hat{\mathbf{M}}_i||^2 \rightarrow \min_{\{\mathbf{R}_r, \mathbf{t}_r, \lambda\}}$





Result (Method 2): residual errors 1.0 meters RMS (1.4 meters max)

Rig pose from far range arrangement

Fixed: inter-camera pose $\{\mathbf{R}, \mathbf{t}\}$ up to scale, ($||\mathbf{t}|| = \lambda$ unknown)
Estimate: scale λ and rig pose $\{\mathbf{R}_r, \mathbf{t}_r\}$

Method 1: reprojection error $\sum_i ||\mathbf{m}_i - \hat{\mathbf{m}}_i||^2 \rightarrow \min_{\{\mathbf{R}_r, \mathbf{t}_r, \lambda\}}$

Method 2: 3D registration error $\sum_i ||\mathbf{M}_i - \hat{\mathbf{M}}_i||^2 \rightarrow \min_{\{\mathbf{R}_r, \mathbf{t}_r, \lambda\}}$

Potential causes for structural inconsistency

- inaccurate measurements in the X-dataset
- inaccurate inter-camera pose
- inaccurate intrinsic parameters

Rig pose from far range arrangement

Fixed: inter-camera pose $\{\mathbf{R}, \mathbf{t}\}$ up to scale, ($||\mathbf{t}|| = \lambda$ unknown)
Estimate: scale λ and rig pose $\{\mathbf{R}_r, \mathbf{t}_r\}$

Method 1: reprojection error $\sum_i ||\mathbf{m}_i - \hat{\mathbf{m}}_i||^2 \rightarrow \min_{\{\mathbf{R}_r, \mathbf{t}_r, \lambda\}}$

Method 2: 3D registration error $\sum_i ||\mathbf{M}_i - \hat{\mathbf{M}}_i||^2 \rightarrow \min_{\{\mathbf{R}_r, \mathbf{t}_r, \lambda\}}$

Potential causes for structural inconsistency

- inaccurate measurements in the X-dataset $\Rightarrow$NO
- inaccurate inter-camera pose $\Rightarrow$NO
- inaccurate intrinsic parameters $\Rightarrow$YES

Proposed far-range pose estimation method

Result: far-range 3D registration very sensitive to intrinsics

Proposed far-range pose estimation method

Result: far-range 3D registration very sensitive to intrinsics
Idea: fine-tune intrinsics to far range for better 3D registration

Proposed far-range pose estimation method

Result: far-range 3D registration very sensitive to intrinsics
Idea: fine-tune intrinsics to far range for better 3D registration
Joint optimization: rig pose $\{\mathbf{R}_r, \mathbf{t}_r\}$, scale λ and intrinsics $\{f, x_0, y_0\}$

Proposed far-range pose estimation method

Result: far-range 3D registration very sensitive to intrinsics
Idea: fine-tune intrinsics to far range for better 3D registration
Joint optimization: rig pose $\{\mathbf{R}_r, \mathbf{t}_r\}$, scale λ and intrinsics $\{f, x_0, y_0\}$

Method 3 (Modified, iterative 3D registration)

1. Radial correction using $\{f, x_0, y_0\}$
2. SIFT matches $\Rightarrow$ inter-camera pose
3. 3D registration $\Rightarrow$ rig pose $\{\mathbf{R}_r, \mathbf{t}_r\}$ and λ
4. Change $\{f, x_0, y_0\}$ and go to Step 1 until convergence of

$$\mathcal{C}_{3D}(f, x_0, y_0 \mid \mathbf{R}_r, \mathbf{t}_r, \lambda) = \sum_{i=1}^{n} ||\mathbf{M}_i - \hat{\mathbf{M}}_i||^2 \rightarrow \min_{\{f, x_0, y_0\}}$$

Proposed far-range pose estimation method

Result: far-range 3D registration very sensitive to intrinsics
Idea: fine-tune intrinsics to far range for better 3D registration
Joint optimization: rig pose $\{\mathbf{R}_r, \mathbf{t}_r\}$, scale λ and intrinsics $\{f, x_0, y_0\}$

Method 3 (Modified, iterative 3D registration)

1. Radial correction using $\{f, x_0, y_0\}$
2. SIFT matches $\Rightarrow$ inter-camera pose
3. 3D registration $\Rightarrow$ rig pose $\{\mathbf{R}_r, \mathbf{t}_r\}$ and λ
4. Change $\{f, x_0, y_0\}$ and go to Step 1 until convergence of

$$\mathcal{C}_{3D}(f, x_0, y_0 \mid \mathbf{R}_r, \mathbf{t}_r, \lambda) = \sum_{i=1}^{n} ||\mathbf{M}_i - \hat{\mathbf{M}}_i||^2 \rightarrow \min_{\{f, x_0, y_0\}}$$

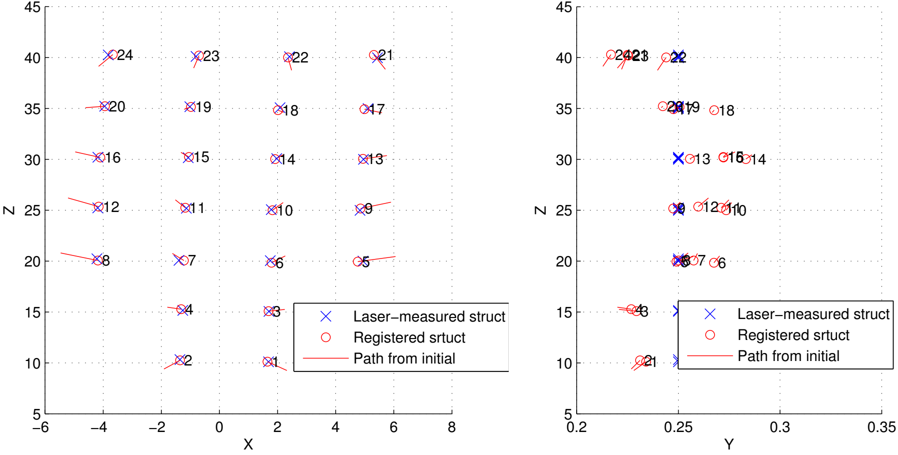
Method 4 (Max. Likelihood for fixed-inter-camera pose)

$$\mathcal{C}_{ML}(\mathbf{R}_r, \mathbf{t}_r, \lambda, \hat{\mathbf{M}}) = \underbrace{\tfrac{1}{\sigma^2}||\bar{\mathbf{m}} - \hat{\mathbf{m}}||_2^2}_{\textit{error in the images}} + \underbrace{||\bar{\mathbf{M}} - \hat{\mathbf{M}}||_{\mathbf{\Sigma}}^2}_{\textit{error in 3D space}} \rightarrow \min_{\{\mathbf{R}_r, \mathbf{t}_r, \lambda, \hat{\mathbf{M}}\}}$$

Optimal changes (left,right):

$$f: -5.8, -5.2 \qquad x_0: +0.6, -1.9 \qquad y_0: +1.6, +2.8$$

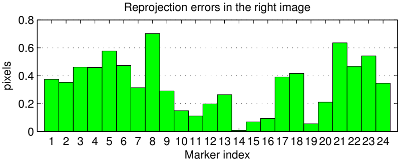
Effect of these changes to 3D reconstruction

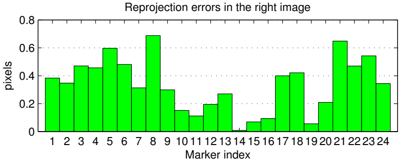


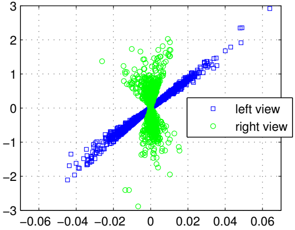
Optimal changes (left,right):

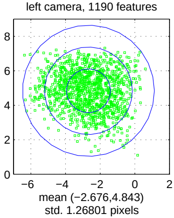
$$f : -5.8, -5.2 \qquad x_0 : +0.6, -1.9 \qquad y_0 : +1.6, +2.8$$

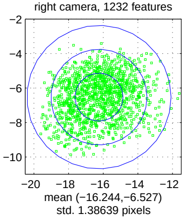
Residual reprojection errors





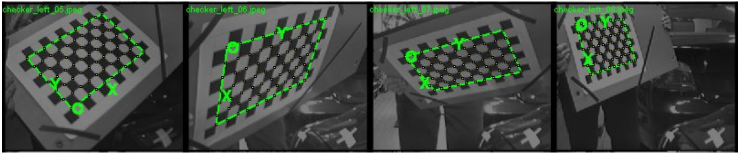


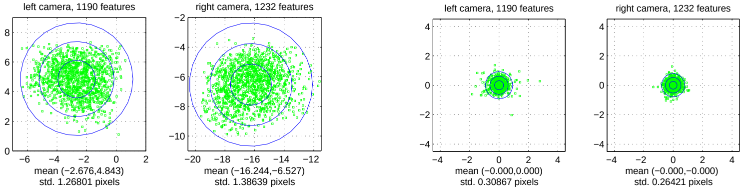




Correction for increased checkerboard reprojection errors

Problem: we changed intrinsics $\Rightarrow$ not optimal for checkerboard dataset
Correction: optimize checkerboard poses (intrinsics fixed)





Result: successful compensation via repositioning, while poses are of no interest!

Comparison

RMS of residual error norms in the different datasets for datasets $\times$ methods

		$\mid\longleftarrow$inter-camera pose from matches$\longrightarrow\mid$			
	ML	reproj.	3D reg.	intr-to-X	ML rig
checker,left (pixels)	0.300	0.300	0.300	0.309*	0.309*
checker,right (pixels)	0.256	0.256	0.256	0.264**	0.264**
X 3D (meters)	0.091	0	**0.952**	0.144	0.089
X images (pixels)	0.21	**3.60**	0.44	0.38	0.39
SIFT,epipolar (pixels)	**0.90**	0.42	0.42	0.49	0.49

***5.7** and ****17.3** before optimized repositioning of the checkerboards

Outline

Summary

Far range stereo calibration

1. Far-range setup for full pose estimation (ML method)
2. Problem: few points for pose $\Leftrightarrow$ many for intrinsic
3. stereo matches from on-line videos (new dataset)
4. Decoupling: inter-camera pose + rig pose
5. Rig pose estimation: (1) reprojection, (2) 3D registration
6. Inconsistency $\Leftarrow$ inaccurate intrinsics
7. (3) Iterative 3D registration: fine-tune intrinsics based on far-range arrangement
8. (4) ML rig pose (fixed inter-camera pose)
9. Good consistency over all datasets
10. Useful to evaluate on-line pose/autocalibration methods...

Optimization methods to calibrate a stereo rig with increased accuracy for vehicular applications

András Bódis-Szomorú, Tamás Dabóczi

bodis@mit.bme.hu
http://www.mit.bme.hu/~bodis

Budapest University of Technology and Economics
Dept. of Measurement and Information Systems
Budapest, Hungary

I2MTC Graz, May 14, 2012