Berlin University of Technology
Computer Vision & Remote Sensing

Evaluation of Relative Pose Estimation Methods for Multi-Camera Setups

Volker Rodehorst, Matthias Heinrichs and Olaf Hellwich

1. Introduction

Problem:

- Fully **automatic** and **reliable** calculation of **relative camera pose** from image correspondences.
- Especially in case of **critical camera motions** or when observing **special point configurations**.
- Some methods provide **multiple solutions** for the relative orientation.

Main Contribution:

- Comparison of various techniques and analysis of their difficulties.
- Recognition that pose estimation of a single calibrated camera is still difficult.
- New constraints for **multiple cameras** that are **fixed on a rig**, which significantly stabilize the pose estimation process.

2. Relative Pose Recovery

Motion of **calibrated** cameras is analyzed using the **essential matrix** $\mathbf{E}$ for corresponding image points $\mathbf{x} \quad \mathbf{x}'$ with known calibration matrices $\mathbf{K}$ and $\mathbf{K}'$.

Properties:

1. Coplanarity or **epipolar-constraint** in terms of normalized coordinates $\mathbf{u} \quad \mathbf{u}'$:
   $\mathbf{u}_i'^{T}\mathbf{E}\mathbf{u}_i \quad 0$ $\quad$ $\mathbf{u}_i \quad \mathbf{K}^{-1}\mathbf{x}_i \quad \mathbf{u}_i' \quad \mathbf{K}'^{-1}\mathbf{x}_i'$
2. Cubic singularity or **rank-constraint**:
   $\det(\mathbf{E}) \quad 0$
3. Cubic **trace-constraint**:
   $2\mathbf{E}\mathbf{E}^{T}\mathbf{E} \quad \mathrm{trace}(\mathbf{E}\mathbf{E}^{T})\mathbf{E} \quad \mathbf{0}$

3. Evaluated Methods

Analysis of 8 state-of-the-art algorithms:

- **Linear 8-point** (Hartley, 1997)
  Estimation from equation (1) and later insertion of constraints (2) and (3) with:
  $\tilde{\mathbf{E}} \quad \mathbf{U}\,\mathrm{diag}(\;,\; ,0)\,\mathbf{V}^{T} \qquad (_1 \quad _2)/2$
- **7-point** (Hartley & Zisserman, 2004)
  Third-order polynomial using (2).
- **Linear 6-point** (Philip, 1996/98)
  Nine third-order polynomials from (3).
- **6-point** (Pizarro et al., 2003)
  Sixth-degree polynomial from (3).
- **Minimal 5-point** (Nistér, 2004)
  Tenth-order polynomial from (3) and (2). Sturm sequences to bracket the roots.
- **Minimal 5-point** (Stewénius et al, 2006)
  Solution with eigen decomposition.
- **Minimal 5-point** (Li & Hartley, 2006)
  Simultaneous parameter estimation.
- **Non-linear 5-point** (Batra et al., 2007)
  Levenberg-Marquardt with random initialization.

Method	Points	Deg.	Solutions
Hartley, 1997	≥ 8	a)	1
Hartley & Zisserman, 2004	≥ 7	a)	≤ 3
Philip, 1996/98	≥ 6	a)	1
Pizarro et al., 2003	≥ 6	b)	≤ 6
Nister, 2004 Stewénius et al., 2006 Li & Hartley, 2006	≥ 5	b)	≤ 10

Table 1: Direct solvers for relative orientation

Degenerate configurations:

a) Coplanar object points and ruled quadric containing the projection centers.
b) Orthogonal ruled quadric, especially cylinder containing projection centers.

4. Experimental Results Using Synthetic Data

Ground Truth Generation

- 2 random cameras, 100 random object points and projected image points.
- Prevent degenerated pointsets, outliers and critical camera setups.

Figure 1: Example out of 100 datasets.

Test Szenario

- Varying **noise** for image coordinates.
- **Selection strategy** for multiple solutions
  - **Sampson**-distance of additional points:
    $$d_{error} \quad \sum_i \frac{\left(\mathbf{u}_i'\mathbf{E}\mathbf{u}_i\right)^2}{\left(\mathbf{E}\mathbf{u}_i\right)_x^2 \quad \left(\mathbf{E}\mathbf{u}_i\right)_y^2 \quad \left(\mathbf{E}^{T}\mathbf{u}_i'\right)_x^2 \quad \left(\mathbf{E}^{T}\mathbf{u}_i'\right)_y^2}$$
  - **Cheirality**-test to maximize the number of points in front of both cameras in case of enough camera translation:
    $$\mathbf{P} = \left[\mathbf{I}|0\right], \quad \mathbf{P}' = \left[\mathbf{R}|\mathbf{t}\right], \quad \frac{\|\mathbf{u}' \quad \mathbf{R}\mathbf{u}\|}{\|\mathbf{u}\|\;\|\mathbf{u}'\|} \quad t_{mov}$$
- Minimal vs. **over-determined** solutions.
- **Data conditioning** for $\mathbf{E}$ is not necessary
- Deconditioning with similarity transformations $\mathbf{T}$ and $\mathbf{T}'$ of the null-vectors $\mathbf{E}_i$ must be applied **before** root searching:
  $$\tilde{\mathbf{E}}_i \quad \mathbf{T}'^{T}\mathbf{E}_i\mathbf{T} \qquad \mathbf{E} \quad \sum_{i \; 1}^{4} \;_i\mathbf{E}_i$$

Visual Verification

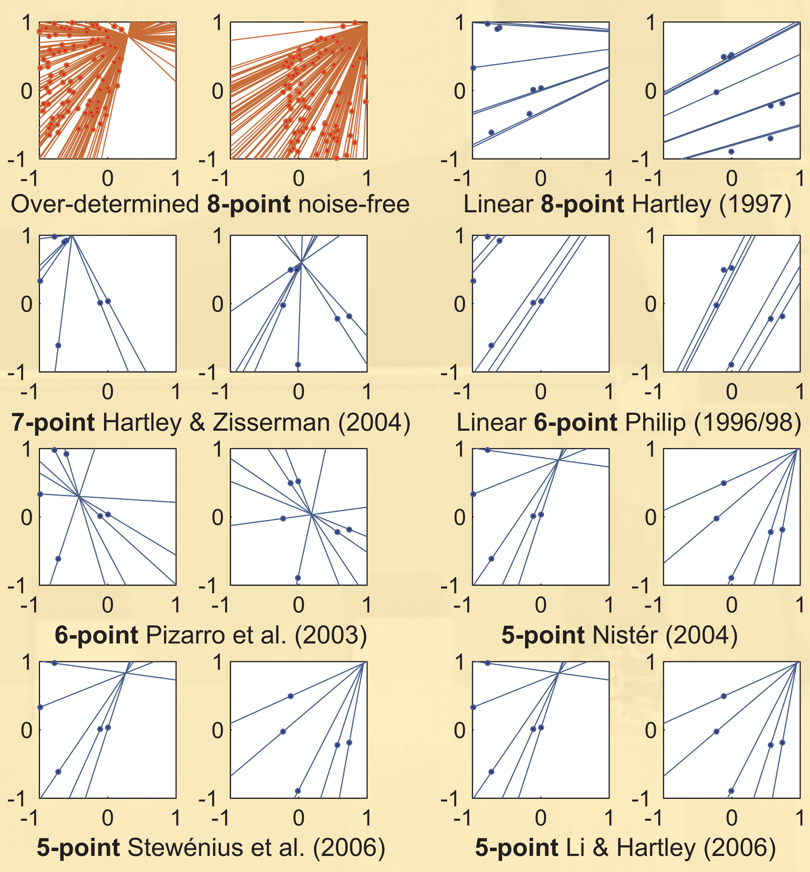

Figure 2: Estimated relative orientation using normalized image pairs with overlaid epipolar rays (ground truth in red).

Error Analysis

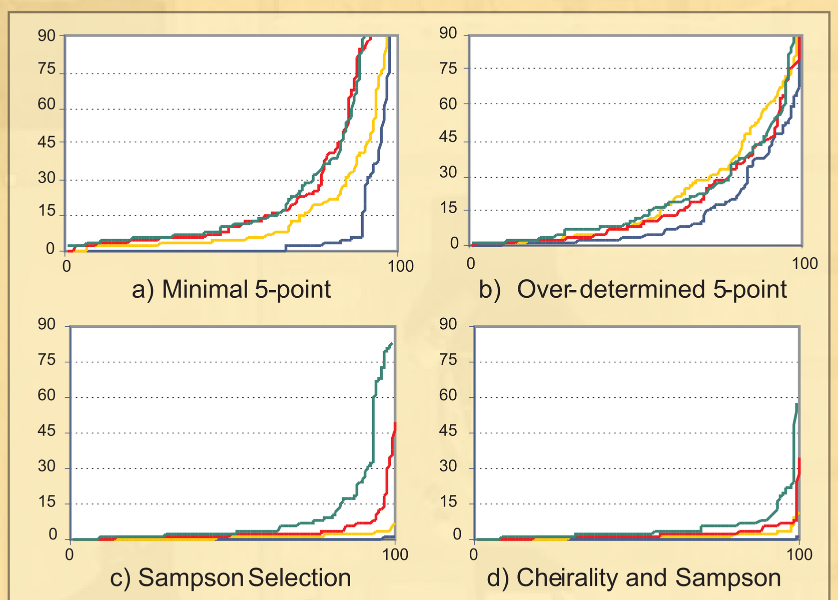

Figure 3: Translation error of the direct 5-point solver (Nistér, 2004) in degrees for 100 runs.
(=0.07, =0.5, =0.9 and =1.3)

Quantitative Evaluation

Rotation Errors

Method	σ = 0.07		σ = 0.5		σ = 0.9		σ = 1.3	
	cnt	mean	cnt	mean	cnt	mean	cnt	mean
Ground Truth	100	0.0860	100	0.7475	100	1.0872	100	2.0434
Evaluation of Different Algorithms								
8-Point	90	2.2589	52	5.0803	32	5.1768	18	6.3161
7-Point	82	1.7805	42	4.0002	33	5.2407	24	6.3662
6-Linear	73	2.4829	38	4.4165	29	4.6610	16	5.7286
6-Pizzaro	78	1.0380	49	4.1056	39	4.5695	24	4.7601
5-Nistér	89	0.9935	67	3.1534	52	4.5321	50	5.4091
5-Stewénius	89	0.9935	67	3.1534	52	4.5321	50	5.4091
5-Li	89	0.9935	67	3.1534	52	4.5321	36	5.6712
5-Non-linear	30	0.7699	24	1.0879	32	1.5902	32	3.8587
Selection Strategy for Multiple Solutions								
Sampson S	100	0.2826	100	1.4059	94	2.3680	81	3.1766
Cheirality C	38	2.3562	36	4.1240	19	3.9483	15	4.9022
C-5 + S	100	0.2610	99	1.5168	99	2.2784	91	3.1963
C-all + S	100	0.2420	100	1.2317	96	2.2333	88	2.7443
Over-determined Solution								
8-Point	99	0.3966	96	1.8661	81	3.1579	62	4.5419
5-Nistér	70	2.9331	51	3.5674	52	3.4531	48	4.3646
Data Conditioning								
8-Uncond	99	0.3197	94	2.0592	83	3.3274	61	4.1966
8-Cond	98	0.3433	93	2.0068	84	3.0240	64	3.6310
5-Uncond	100	0.2692	98	1.2245	95	2.2597	90	2.9573
5-Cond	100	0.2314	100	1.3404	97	2.1464	92	2.8859

Translation Errors

Method	σ = 0.07		σ = 0.5		σ = 0.9		σ = 1.3	
	cnt	mean	cnt	mean	cnt	mean	cnt	mean
Ground Truth	100	0.0003	100	0.0030	100	0.0048	100	0.0086
Evaluation of Different Algorithms								
8-Point	95	0.3361	67	0.7335	57	0.9643	32	1.0265
7-Point	85	0.2191	69	0.6678	55	0.8858	45	0.9150
6-Linear	81	0.3507	51	0.7330	44	0.8677	33	1.0573
6-Pizzaro	79	0.1871	57	0.7736	44	0.7868	32	1.0351
5-Nistér *	88	0.1346	77	0.6184	64	0.7910	54	0.9737
5-Stewénius	88	0.1346	77	0.6184	64	0.7910	54	0.9737
5-Li	88	0.1346	77	0.6184	64	0.7910	38	1.0040
5-Non-linear	40	0.4367	31	0.4205	37	0.3403	37	0.4715
Selection Strategy for Multiple Solutions								
Sampson S *	100	0.0364	100	0.2069	95	0.3381	87	0.5475
Cheirality C	38	0.6003	40	0.7801	25	0.8582	21	1.1246
C-5 + S *	100	0.0350	100	0.2140	99	0.3184	96	0.4879
C-all + S	100	0.0280	100	0.1619	96	0.2800	90	0.4413
Over-determined Solution								
8-Point	97	0.0433	96	0.2529	89	0.4730	69	0.6653
5-Nistér *	84	0.5310	70	0.6350	73	0.5765	76	0.7148
Data Conditioning								
8-Uncond	94	0.0429	95	0.2776	92	0.4581	73	0.6194
8-Cond	98	0.0460	95	0.2817	86	0.3988	73	0.6678
5-Uncond	100	0.0354	99	0.1724	98	0.3231	95	0.5198
5-Cond	100	0.0292	100	0.1621	100	0.3037	94	0.4303

Table 2: Percentage of correct solutions (rotation error < 2°, translation error < 10°) and mean errors (* method shown in Figure 3).

5. Pose Estimation with Multiple Cameras

Problem:

- The main drawback of the 5-point algorithms are the 10 possible solutions.

Idea:

- Constrained motion of $n \geq 2$ cameras.
- Fixed relationship by mounting the cameras on a common frame.
- Select a unique pair of essential matrices from the two solution sets.
- We denote a camera i at time j with $\mathbf{P}_i^j$.
- The reference camera $\mathbf{P}_1^1$ is at origin.
- The cameras $\mathbf{P}_1^1$ and $\mathbf{P}_2^1$ have fixed relative orientation $?\mathbf{T}_2$.

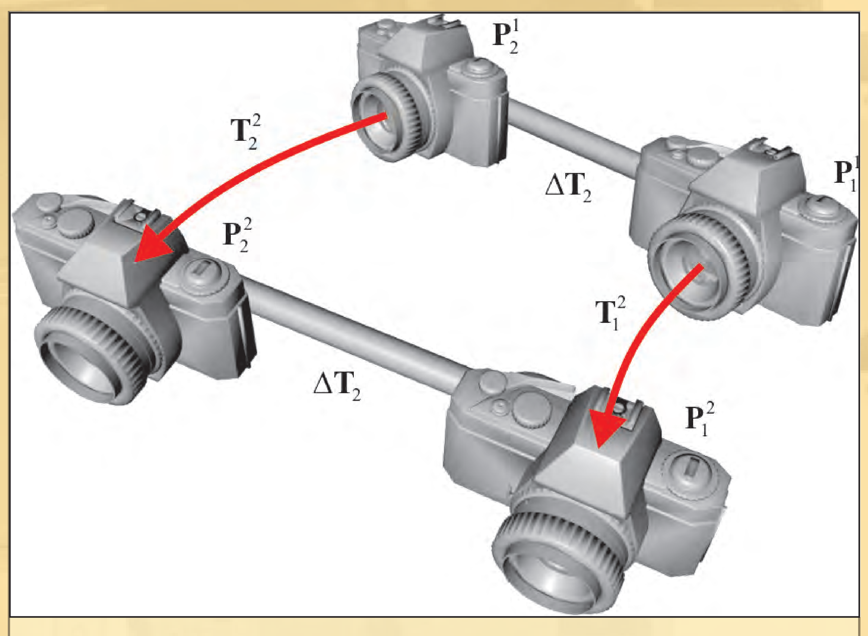

Figure 4: Constrained multiple-camera motion.

6. Multi-Camera Constraints

- $\mathbf{E}$ contains relative orientation up to scale
- Estimate the **relative scale** factors μ and λ using translation $\mathbf{C}$ and rotation $\mathbf{R}$:
  $$\mathbf{C}_1^2 \qquad \mathbf{R}_2\mathbf{C}_2^2 \qquad\qquad \mathbf{C}_2 \quad \mathbf{R}_1^2 \quad \mathbf{C}_2$$
- For critical motions use 5 spatial points $\mathbf{X}$
  $$0.2 \sum_{i\;1}^{5} \frac{\|\mathbf{X}_i\|}{\|\mathbf{X}_i^1\|}, \qquad 0.2 \sum_{i\;1}^{5} \frac{\|\mathbf{X}_i\|}{\|\mathbf{X}_i^2\|}$$
- Our **cost function** to select a pair of solutions is defined as
  $$e_i^j \quad w_{trans} \quad \|\mathbf{r}\| \quad w_{rot} \quad r_{error}$$
  with weighting factors $w_{trans} \quad 5w_{rot}$
  residual errors $\mathbf{r} \quad \mathbf{A}\mathbf{x} \quad \mathbf{b}$
  and $r_{error} \quad \frac{1}{3}\sum_{i \; x,y,z} \mathrm{acos} \quad \mathbf{R}_1\mathbf{e}_i \;^{T} \mathbf{R}_2\mathbf{e}_i$

7. Multi-Camera Experiments

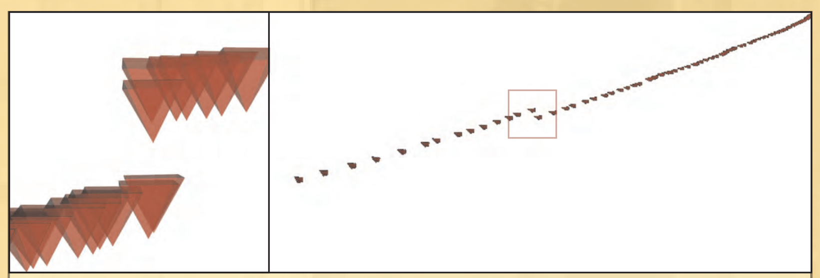

Figure 5: Reconstructed single-camera path over 110 frames.

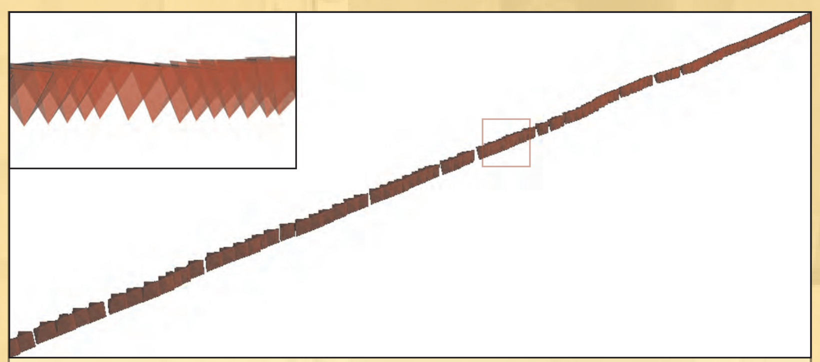

Figure 6: Reconstructed multi-camera path over 110 frames.

Figure 7: Reconstructed multi-camera path over 900 frames.

8. Conclusions

- Data conditioning for $\mathbf{E}$ is not necessary.
- The **minimal 5-point** solvers produce better results than all other methods, especially in presence of noise.
- The estimation of camera **rotation** is more reliable than the **translation**.
- In case of multiple solutions, the best selection criterion is a combination of a preceding **cheirality**-test with **minimal** points followed by the computation of the **Sampson**-distance over **all** available points.
- Using **over-determined** variants of the minimal solver not necessarily increase the accuracy of the essential matrix.
- The proposed multi-camera constraints allow extensive path reconstructions.
- The result may be improved with bundle adjustment.

Acknowledgements

This work was founded by the german research foundation

www.cv.tu-berlin.de Volker Rodehorst +49 30 314 23163 +49 30 314 21104 vr@cs.tu-berlin.de