Single-Scanline Relative Pose Estimation for Rolling Shutter Cameras

Supplementary Material

A. Decomposition of the Trifocal Tensor into Canonical Cameras

Given a trifocal tensor $\mathbf{T} \in \mathbb{R}^{2,2,2}$, we describe that we can decompose it into two possible sets of 2×3 matrices $(\mathbf{A}_{c,1}, \mathbf{A}_{c,2}, \mathbf{A}_{c,3})$ in the canonical form. described in Eq. (20) of the main paper.

First, we note that calculating the tensor elements according to Eq. (23) of the main paper from cameras $\mathbf{A}_{c,1}, \mathbf{A}_{c,2}, \mathbf{A}_{c,3}$ yields

- $T_{000} = 1$
- $T_{001} = \alpha_7$
- $T_{010} = \alpha_3$
- $T_{011} = \alpha_3\alpha_7 - \alpha_4\alpha_6$
- $T_{100} = \alpha_1$
- $T_{010} = \alpha_1\alpha_7 - \alpha_1\alpha_5$
- $T_{110} = \alpha_1\alpha_3 - \alpha_1\alpha_2$
- $T_{111} = \alpha_1\alpha_3\alpha_7 - \alpha_1\alpha_4\alpha_6 - \alpha_1\alpha_2\alpha_7 + \alpha_1\alpha_4\alpha_5 + \alpha_1\alpha_2\alpha_6 - \alpha_1\alpha_3\alpha_5$

We normalize the tensor such that its first entry is 1:

$$\mathbf{T} \leftarrow \frac{\mathbf{T}}{T_{000}}. \tag{1}$$

From the normalized tensor, we extract the parameters:

$$\alpha_1 = T_{100}, \quad \alpha_3 = T_{010}, \quad \alpha_7 = T_{001}, \tag{2}$$

$$\alpha_5 = \frac{-T_{101} + \alpha_1\alpha_7}{\alpha_1}, \quad \alpha_2 = \frac{-T_{110} + \alpha_1\alpha_3}{\alpha_1}. \tag{3}$$

Now, T_{100} and T_{111} give two constraints for α_4, α_6. In the equation for T_{111}, we substitute $\alpha_3\alpha_7 - T_{100}$ to get a linear constraint in α_4, α_6. Expressing α_6 from this constraint and substituting into the constraint for T_{100} yields a quadratic equation

$$c_A\alpha_4^2 + c_B\alpha_4 + c_C = 0 \tag{4}$$

with coefficients

$$\begin{aligned} c_A &= -\frac{\alpha_5}{\alpha_2}, \\ c_B &= \frac{T_{111} - \alpha_1\alpha_3\alpha_7 + \alpha_1(\alpha_3\alpha_7 - T_{011}) + \alpha_1\alpha_2\alpha_7 + \alpha_1\alpha_3\alpha_5}{\alpha_1\alpha_2}, \\ c_C &= T_{011} - \alpha_3\alpha_7. \end{aligned} \tag{5}$$

Solving (4), we obtain two solutions:

$$\begin{aligned} \alpha_{4a} &= \frac{-c_B + \sqrt{c_B^2 - 4c_Ac_C}}{2c_A}, \\ \alpha_{4b} &= \frac{-c_B - \sqrt{c_B^2 - 4c_Ac_C}}{2c_A}. \end{aligned} \tag{6}$$

Finally, we find the corresponding values α_{6_a}, α_{6_b} of α_6 from the equation for T_{111}.

Using these values, we construct two sets of projection matrices:

$$\mathbf{A}_1 = \begin{bmatrix} 1 & 0 & 0 \\ \alpha_1 & \alpha_1 & \alpha_1 \end{bmatrix}, \tag{7}$$

$$\mathbf{A}_2 = \begin{bmatrix} 0 & 1 & 0 \\ \alpha_2 & \alpha_3 & \alpha_4 \end{bmatrix}, \tag{8}$$

$$\mathbf{A}_3 = \begin{bmatrix} 0 & 0 & 1 \\ \alpha_5 & \alpha_6 & \alpha_7 \end{bmatrix}, \tag{9}$$

where (α_4, α_6) takes either $(\alpha_{4a}, \alpha_{6a})$ or $(\alpha_{4b}, \alpha_{6b})$, yielding two valid decompositions.

This method provides a direct way to extract the projection matrices from a given trifocal tensor while preserving the canonical structure. This also shows, that for a generic tensor $\mathbf{T} \in \mathbb{R}^{2,2,2}$ there are two ways to decompose it into matrices in the canonical form.

B. Triangulation of Lines from a Single Scanline

Given the projections of a 3D line onto a single scanline, we describe how we estimate a point $\mathbf{L}_0$ on the line and measure the reprojection error. Since the line direction $\mathbf{L}_d$ is known and corresponds to the y-axis, the problem reduces to estimating the 3D position of one point $\mathbf{L}_0$ on the line. Let $\mathbf{A}_i \in \mathbb{R}^{2\times3}$ be the projection matrix encoding the pose of the i-th camera, and let $\mathbf{u}_i \in \mathbb{R}^2$ be the measured projection of the line on the scanline. We construct the constraint matrix $\mathbf{M} \in \mathbb{R}^{m\times3}$ as:

$$\mathbf{M} = \begin{bmatrix} \frac{\mathbf{u}_1^T\mathbf{A}_1}{\|\mathbf{u}_1^T\mathbf{A}_1\|} \\ \dots \\ \frac{\mathbf{u}_m^T\mathbf{A}_m}{\|\mathbf{u}_m^T\mathbf{A}_m\|} \end{bmatrix}. \tag{10}$$

We solve for the homogeneous line representation $\mathbf{L}_h$ as the right singular vector corresponding to the smallest singular value of $\mathbf{M}$:

$$\mathbf{L}_h = \arg\min_{\|\mathbf{L}\|=1} \|\mathbf{M}\mathbf{L}\|. \tag{11}$$

Normalizing $\mathbf{L}_h$ such that its third component is 1, we obtain:

$$\mathbf{L}_h = \frac{\mathbf{L}_h}{\mathbf{L}_{h(3)}}. \tag{12}$$

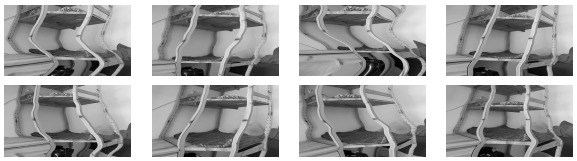

Figure 1. Custom images used in the experiment in Section C.2.

The estimated 3D point on the line is then given by:

$$\mathbf{L}_0 = \begin{bmatrix} \mathbf{L}_{h(1)} \\ 0 \\ \mathbf{L}_{h(2)} \end{bmatrix}. \tag{13}$$

To evaluate the reprojection error, we compute the projected coordinates of $\mathbf{L}_h$:

$$\mathbf{p}_i = \mathbf{A}_i \mathbf{L}_h, \tag{14}$$

and measure the angular deviation from the observed scanline projection:

$$e_i = \left| \frac{p_{i,2}}{p_{i,1}} + \frac{u_{i,0}}{u_{i,1}} \right|. \tag{15}$$

We get the total reprojection error by:

$$e_{\text{total}} = \sum_{i=1}^{m} e_i^2. \tag{16}$$

C. Additional experiments

C.1. Detailed evaluation of the experiments from the main paper

Here, we present detailed results of the experiments conducted on the real-world dataset Fastec [2]. In the case of solvers $(E, 3, 5)$, $(E, 4, 4)$, and $(D, 3, 7)$, we present median and minimum rotation and translation errors for every sequence from the Fastec dataset, as well as the percentages of scenes, whose pose error (obtained as the maximum over the rotation and translation error) does not exceed $10°$, $20°$, and $30°$. In the case of solver $(B, 3, 7)$, we cannot measure the pose error, so we present the median and minimum tensor error for every sequence from the Fastec dataset.

C.2. Evaluation on images with very large rolling-shutter distortions

Here, we present an evaluation of our method on a custom dataset simulated very large rolling-shutter distortions. This scenario, while not very realistic, demonstrates that the proposed solvers are not limited to a particular type of motion. We have taken multiple high speed videos of the same scene with an iPhone, and simulated rolling shutter images (Fig. 1) by selecting one scanline per frame. We have generated ground truth poses for every frame and obtained the camera intrinsics using COLMAP [4].

Seq	R,t errors in ° Median		Minimum		Percentage of errors below 10° / 20° / 30°
4	1.0	71.8	6.0	12.0	0.0 / 18.5 / 33.3
9	4.0	45.3	2.6	16.9	0.0 / 3.2 / 9.7
10	3.8	40.5	0.9	16.2	0.0 / 9.4 / 34.4
11	2.7	32.1	0.2	8.9	3.1 / 21.9 / 40.6
13	3.2	38.1	0.9	7.5	3.1 / 12.5 / 12.5
16	4.1	42.4	3.1	32.7	0.0 / 0.0 / 0.0
17	2.0	38.6	1.0	13.8	0.0 / 12.5 / 31.3
18	6.0	44.7	2.2	20.3	0.0 / 0.0 / 25.0
20	2.4	42.8	1.0	20.0	0.0 / 0.0 / 31.3
21	2.8	63.8	6.3	32.6	0.0 / 0.0 / 0.0
32	3.8	43.6	11.1	26.3	0.0 / 0.0 / 8.0
50	0.6	34.3	0.1	20.1	0.0 / 0.0 / 37.5
64	1.9	71.2	0.5	10.8	0.0 / 4.5 / 4.5
74	1.9	118.8	1.9	115.8	0.0 / 0.0 / 0.0
76	0.8	111.4	0.3	51.3	0.0 / 0.0 / 0.0
79	1.4	25.2	0.4	10.7	0.0 / 15.6 / 62.5
88	3.4	40.0	1.8	27.1	0.0 / 0.0 / 10.0
92	2.2	25.9	0.6	3.3	12.5 / 12.5 / 59.4
100	4.0	44.3	7.6	19.0	0.0 / 6.3 / 15.6

Table 1. Detailed evaluation of the errors for solver $(E, 3, 5)$ in the multi-view setting. For every sequence of Fastec [2], we give the median and the minimum rotation and translation error, as well as the percentage of pose errors below $10°, 20°, 30°$.

While it is very difficult to detect the whole lines, our method only needs to detect the segment of the line in the vicinity of a selected scanline. To detect these, we have found Canny edges in the images, and fitted conics into the edge points using [1]. To match the curves, we obtained point matches using LoFTR [5], associated every detected curve with keypoints within 2px from it, and used the number of shared associated keypoints as the score for matching the curves.

We estimated the pose for every camera triplet from Fig. 1 using solver $(E, 3, 5)$. The median rotation and translation error are $4.8°$ and $14.8°$. 30.6% triplets have pose error below $10°$, and 56.5% have pose error below $20°$. Every image is in at least one triplet with pose error below $6°$.

We have further evaluated the five point solver [3] on all images pairs from Fig. 1. The median rotation and translation error are $6.6°$ and $44.4°$. 10.7% triplets have pose error below $10°$, and 21.4% have pose error below $20°$. This demonstrates that our method can handle large rolling-shutter distortions better than the five point solver.

C.3. Pose estimation with local optimization

We have added local optimization (LO) after the solvers $(E, 3, 5)$ and $(D, 3, 7)$, and evaluated them on synthetic and real data. The result on synthetic data with 10 lines per scene (Fig. 2) shows that the LO improves the pose. How-

Seq	R,t errors in °				Percentage of errors
	Median		Minimum		below 10° / 20° / 30°
4	1.8	112.4	0.4	4.6	4.3 / 17.4 / 26.1
9	5.3	43.8	1.1	10.6	0.0 / 10.0 / 16.7
10	3.1	27.0	0.7	2.4	3.2 / 22.6 / 58.1
11	3.8	36.1	3.8	7.8	3.2 / 22.6 / 35.5
13	3.5	36.8	1.7	2.6	6.5 / 19.4 / 35.5
16	6.9	50.5	3.7	32.3	0.0 / 0.0 / 0.0
17	2.8	40.4	1.2	3.4	3.2 / 6.5 / 25.8
18	7.0	71.3	4.5	36.5	0.0 / 0.0 / 0.0
20	2.8	35.7	0.6	12.4	0.0 / 16.1 / 32.3
21	5.2	70.4	3.0	9.4	3.2 / 3.2 / 3.2
32	5.5	54.3	19.5	27.8	0.0 / 0.0 / 5.3
50	2.2	62.4	1.1	17.9	0.0 / 6.5 / 22.6
64	1.4	82.2	2.0	30.6	0.0 / 0.0 / 0.0
74	5.1	115.5	5.1	115.5	0.0 / 0.0 / 0.0
76	3.1	128.4	3.5	110.6	0.0 / 0.0 / 0.0
79	2.3	43.0	0.7	21.7	0.0 / 0.0 / 20.7
88	3.5	28.2	4.3	21.7	0.0 / 0.0 / 66.7
92	4.6	26.4	2.3	4.0	9.7 / 22.6 / 54.8
100	3.8	42.8	0.9	6.5	3.2 / 3.2 / 22.6

Table 2. Detailed evaluation of the errors for solver $(E, 4, 4)$ in the multi-view setting. For every sequence of Fastec [2], we give the median and the minimum rotation and translation error, as well as the percentage of pose errors below $10°, 20°, 30°$.

Seq	R,t errors in °				Percentage of errors
	Median		Minimum		below 10° / 20° / 30°
4	1.6	83.3	0.9	14.4	0.0 / 7.4 / 14.8
9	8.6	55.0	16.8	24.6	0.0 / 0.0 / 3.2
10	6.1	49.3	2.5	1.2	6.3 / 18.8 / 21.9
11	10.4	59.7	2.1	7.3	6.5 / 22.6 / 25.8
13	5.1	41.0	0.4	8.6	3.1 / 18.8 / 40.6
16	13.0	57.7	6.2	35.4	0.0 / 0.0 / 0.0
17	3.8	37.2	2.4	3.2	9.4 / 21.9 / 46.9
18	8.1	93.5	3.0	24.2	0.0 / 0.0 / 16.7
20	4.2	57.0	1.9	10.4	0.0 / 9.4 / 18.8
21	6.3	85.8	4.2	4.8	3.1 / 9.4 / 12.5
32	6.4	85.6	2.4	1.1	16.0 / 24.0 / 36.0
50	3.2	94.7	1.5	12.5	0.0 / 3.1 / 9.4
64	31.7	153.4	2.0	17.9	0.0 / 4.3 / 4.3
74	180.0	180.0	180.0	180.0	0.0 / 0.0 / 0.0
76	5.7	141.3	4.4	10.4	0.0 / 5.0 / 10.0
79	6.1	44.8	2.1	2.8	3.1 / 15.6 / 28.1
88	3.1	10.0	3.0	2.7	50.0 / 70.0 / 80.0
92	6.8	53.8	6.5	9.7	6.3 / 25.0 / 28.1
100	8.0	54.0	1.0	8.5	3.1 / 6.3 / 15.6

Table 3. Detailed evaluation of the errors for solver $(D, 3, 7)$ in the multi-view setting. For every sequence of Fastec [2], we give the median and the minimum rotation and translation error, as well as the percentage of pose errors below $10°, 20°, 30°$.

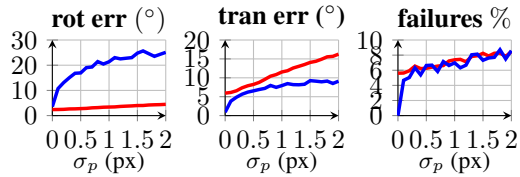

Figure 2. Synthetic tests for solvers $(E, 3, 5)$ and $(D, 3, 7)$ with local optimization using 10 lines. $\sigma_v = 1°$.

ever, the effect of the local optimization on real data from the Fastec dataset and from Fig. 1 is marginal, mostly due to a low number of lines in the images.

References

[1] Radim Halır and Jan Flusser. Numerically stable direct least squares fitting of ellipses. In *Proc. 6th International Conference in Central Europe on Computer Graphics and Visualization. WSCG*, pages 125–132. Citeseer, 1998. 2

[2] Peidong Liu, Zhaopeng Cui, Viktor Larsson, and Marc Pollefeys. Deep shutter unrolling network. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2020. 2, 3, 4, 5

[3] David Nistér. An efficient solution to the five-point relative pose problem. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2004. 2

[4] Johannes Lutz Schönberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. 2

[5] Jiaming Sun, Zehong Shen, Yuang Wang, Hujun Bao, and Xiaowei Zhou. Loftr: Detector-free local feature matching with transformers. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8922–8931, 2021. 2

Seq	R,t errors in °				Percentage of errors below 10° / 20° / 30°
	Median		Minimum		
4	4.7	55.8	0.3	29.9	0.00 / 0.00 / 2.94
9	0.5	117.8	0.5	2.6	3.03 / 3.03 / 6.06
10	0.4	49.7	0.4	20.0	0.00 / 2.94 / 2.94
11	1.3	52.2	3.0	12.0	0.00 / 5.88 / 5.88
13	0.5	93.8	0.1	36.3	0.00 / 0.00 / 0.00
16	7.1	129.6	0.2	28.1	0.00 / 0.00 / 4.00
17	1.7	51.7	0.3	43.2	0.00 / 0.00 / 0.00
18	89.9	128.7	5.2	41.8	0.00 / 0.00 / 0.00
20	0.3	51.0	0.6	1.2	2.94 / 2.94 / 11.76
21	0.3	29.2	0.2	4.3	2.94 / 32.35 / 52.94
32	3.2	130.1	3.9	46.9	0.00 / 0.00 / 0.00
50	4.4	125.9	1.2	47.7	0.00 / 0.00 / 0.00
64	90.1	136.2	8.4	48.1	0.00 / 0.00 / 0.00
74	179.9	179.9	179.9	179.9	0.00 / 0.00 / 0.00
76	47.6	133.9	0.2	24.8	0.00 / 0.00 / 5.00
79	0.7	115.5	8.1	23.9	0.00 / 0.00 / 2.94
88	10.8	55.3	0.1	46.7	0.00 / 0.00 / 0.00
92	0.7	133.3	0.3	37.6	0.00 / 0.00 / 0.00
100	0.4	96.2	2.7	21.3	0.00 / 0.00 / 5.88

Table 4. Detailed evaluation of the errors for solver $(E, 3, 5)$ in the single-view setting. For every sequence of Fastec [2], we give the median and the minimum rotation and translation error, as well as the percentage of pose errors below $10^\circ, 20^\circ, 30^\circ$.

Seq	R,t errors in °				Percentage of errors below 10° / 20° / 30°
	Median		Minimum		
4	1.4	81.2	1.7	1.5	11.11 / 18.52 / 22.22
9	0.3	86.7	0.2	22.7	0.00 / 0.00 / 6.06
10	0.3	49.8	1.2	6.0	3.03 / 9.09 / 15.15
11	0.9	77.1	0.3	10.2	0.00 / 8.82 / 11.76
13	0.6	111.6	0.0	51.5	0.00 / 0.00 / 0.00
16	0.1	57.5	0.1	23.0	0.00 / 0.00 / 7.69
17	0.7	71.0	0.5	3.0	5.88 / 14.71 / 17.65
18	0.5	127.4	8.5	20.1	0.00 / 0.00 / 12.50
20	0.3	59.5	0.7	11.6	0.00 / 11.76 / 17.65
21	0.2	35.4	0.3	15.4	0.00 / 11.76 / 47.06
32	0.3	49.8	1.7	22.2	0.00 / 0.00 / 10.53
50	0.0	125.2	11.3	68.9	0.00 / 0.00 / 0.00
64	1.3	134.6	3.7	27.0	0.00 / 0.00 / 7.69
74	179.9	179.9	179.9	179.9	0.00 / 0.00 / 0.00
76	0.3	74.0	1.0	26.8	0.00 / 0.00 / 11.11
79	0.3	121.5	0.1	38.8	0.00 / 0.00 / 0.00
88	0.0	47.5	0.0	46.6	0.00 / 0.00 / 0.00
92	0.1	47.2	0.1	42.0	0.00 / 0.00 / 0.00
100	0.2	87.3	0.4	5.2	2.94 / 2.94 / 2.94

Table 5. Detailed evaluation of the errors for solver $(E, 4, 4)$ in the single-view setting. For every sequence of Fastec [2], we give the median and the minimum rotation and translation error, as well as the percentage of pose errors below $10^\circ, 20^\circ, 30^\circ$.

Seq	R,t errors in °				Percentage of errors below 10° / 20° / 30°
	Median		Minimum		
4	1.8	48.5	32.4	43.2	0.00 / 0.00 / 0.00
9	0.2	52.9	1.5	3.9	3.03 / 6.06 / 15.15
10	2.1	48.0	25.3	36.0	0.00 / 0.00 / 0.00
11	7.6	48.4	1.1	17.6	0.00 / 2.94 / 5.88
13	3.2	55.9	0.1	24.8	0.00 / 0.00 / 14.71
16	2.7	50.1	14.6	43.4	0.00 / 0.00 / 0.00
17	2.2	49.7	22.8	36.5	0.00 / 0.00 / 0.00
18	6.9	56.8	6.6	26.2	0.00 / 0.00 / 4.17
20	3.2	48.5	18.0	3.1	0.00 / 2.94 / 8.82
21	1.5	28.4	2.1	3.0	2.94 / 20.59 / 61.76
32	1.5	48.8	26.2	33.3	0.00 / 0.00 / 0.00
50	10.3	60.5	4.4	49.9	0.00 / 0.00 / 0.00
64	3.2	44.2	5.4	39.8	0.00 / 0.00 / 0.00
74	180.0	180.0	180.0	180.0	0.00 / 0.00 / 0.00
76	2.2	45.4	0.1	41.4	0.00 / 0.00 / 0.00
79	0.4	63.5	1.2	19.5	0.00 / 2.94 / 8.82
88	5.0	48.6	17.7	39.2	0.00 / 0.00 / 0.00
92	1.3	46.1	22.8	34.0	0.00 / 0.00 / 0.00
100	0.5	50.0	0.8	3.3	2.94 / 5.88 / 5.88

Table 6. Detailed evaluation of the errors for solver $(D, 3, 7)$ in the single-view setting. For every sequence of Fastec [2], we give the median and the minimum rotation and translation error, as well as the percentage of pose errors below $10^\circ, 20^\circ, 30^\circ$.

Seq	Median tensor error	Minimum tensor error
4	0.79	0.33
9	0.74	0.23
10	0.74	0.24
11	0.63	0.20
13	0.61	0.07
16	0.77	0.63
17	0.72	0.30
18	0.93	0.46
20	0.73	0.32
21	1.12	0.72
32	0.93	0.49
50	0.67	0.40
64	1.37	0.52
74	1.41	1.41
76	1.15	0.87
79	0.52	0.10
88	0.79	0.53
92	0.57	0.25
100	0.86	0.26

Table 7. Detailed evaluation of the errors for solver $(B, 3, 7)$ in the multi-view setting. For every sequence of Fastec [2], we give the median and the minimum tensor error.

Seq	Median error	Minimum error
4	0.84	0.75
9	0.87	0.06
10	0.84	0.81
11	0.81	0.32
13	0.86	0.44
16	0.86	0.83
17	0.84	0.15
18	0.90	0.53
20	0.84	0.82
21	0.47	0.08
32	0.84	0.77
50	0.91	0.88
64	0.77	0.66
74	2.00	2.00
76	0.80	0.58
79	1.05	0.40
88	0.83	0.80
92	0.80	0.78
100	0.82	0.04

Table 8. Detailed evaluation of the errors for solver $(B, 3, 7)$ in the single-view setting. For every sequence of Fastec [2], we give the median and the minimum tensor error.