On the Two-View Geometry of Unsynchronized Cameras Appendix

Cenek Albl[1] Zuzana Kukelova[1] Andrew Fitzgibbon[2] Jan Heller [3] Matej Smid[1] Tomas Pajdla[1]

[1]Czech Technical University in Prague
Prague
Czechia
{alblcene,kukelova}@cmp.felk.cvut.cz
smidm@cmp.felk.cvut.cz,pajdla@cvut.cz

[2] HoloLens, Microsoft
Cambridge
UK
awf@microsoft.com

[3]Magik Eye Inc.
New York
US
jan@magik-eye.com

In this appendix we present additional experimental results that provide details that are out of the scope of the main paper.

1. Subframe synchronization

One issue that was not directly elaborated upon in the main paper is the ability of the solvers to synchronize sub-frame time shifts, *i.e.*, shifts where β_{gt} is not an integer. In the real datasets, images were either hardware synchronized, *i.e.*, $\beta_{gt} = 0$, or we did not have precise enough ground truth information about the subframe time shift. Therefore, we tested the subframe synchronization on the synthetic data only. The results in Figure 1 show that the subframe synchronization is very precise for various levels of noise. Figure 2 shows an example of a randomly generated scene for the synthetic experiments.

2. Iterative algorithm visualization

In Figure 3, we provide a visualization of one run of the iterative algorithm with $p_{max} = 5$. Each iteration is marked by a black square and denoted by the number of the iteration k and the distance d used for interpolation in the given iteration. The algorithm greedily searches for a larger number of inliers (the top figure) and uses the estimated β_k to change the correspondences, which results in change of the current ground truth shift (bottom figure). This particular run converged in 6 iterations, even though the initial time shift (50) was larger than the maximum interpolation distance $d = 32$. Moreover, the algorithm only used interpolation distances $d = 1, 2$. These distances were enough to provide a good enough estimate of the time shift that lead to an increased number of inliers.

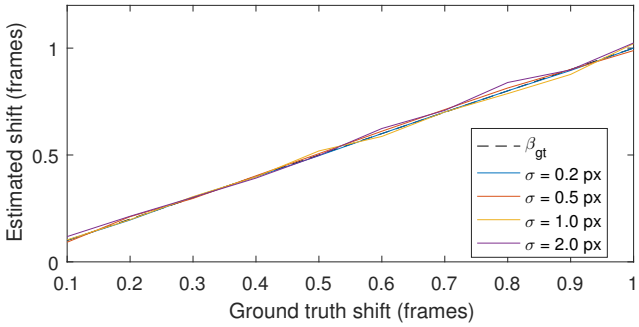

Figure 1. Subframe time shift estimation using the fundamental matrix solver. The solver was tested with different levels of image noise.

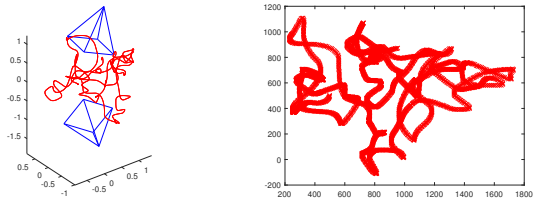

Figure 2. An example of the randomly generated scene for the synthetic experiments. On the left is the 3D trajectory with cameras and on the right is an image projected into one of the cameras.

3. Accuracy of the estimated geometry

3.1. Synthetic data

On the same data as used in Section 5.1 of the main paper, we evaluated the estimated relative rotations $\mathtt{R}$ and translations $\mathbf{t}$. The results in Figure 4 show that we are able to estimate $\mathtt{R}$ and $\mathbf{t}$ significantly better than the classical 7 point algorithm. The utility of our solver is especially apparent from the zoomed in figures with smaller time shifts. The error in $\mathtt{R}$ and $\mathbf{t}$ is almost zero up to 5 frames shift, for shorter interpolation distances $d = 1, 2, 4, 8$. In contrast, such shift

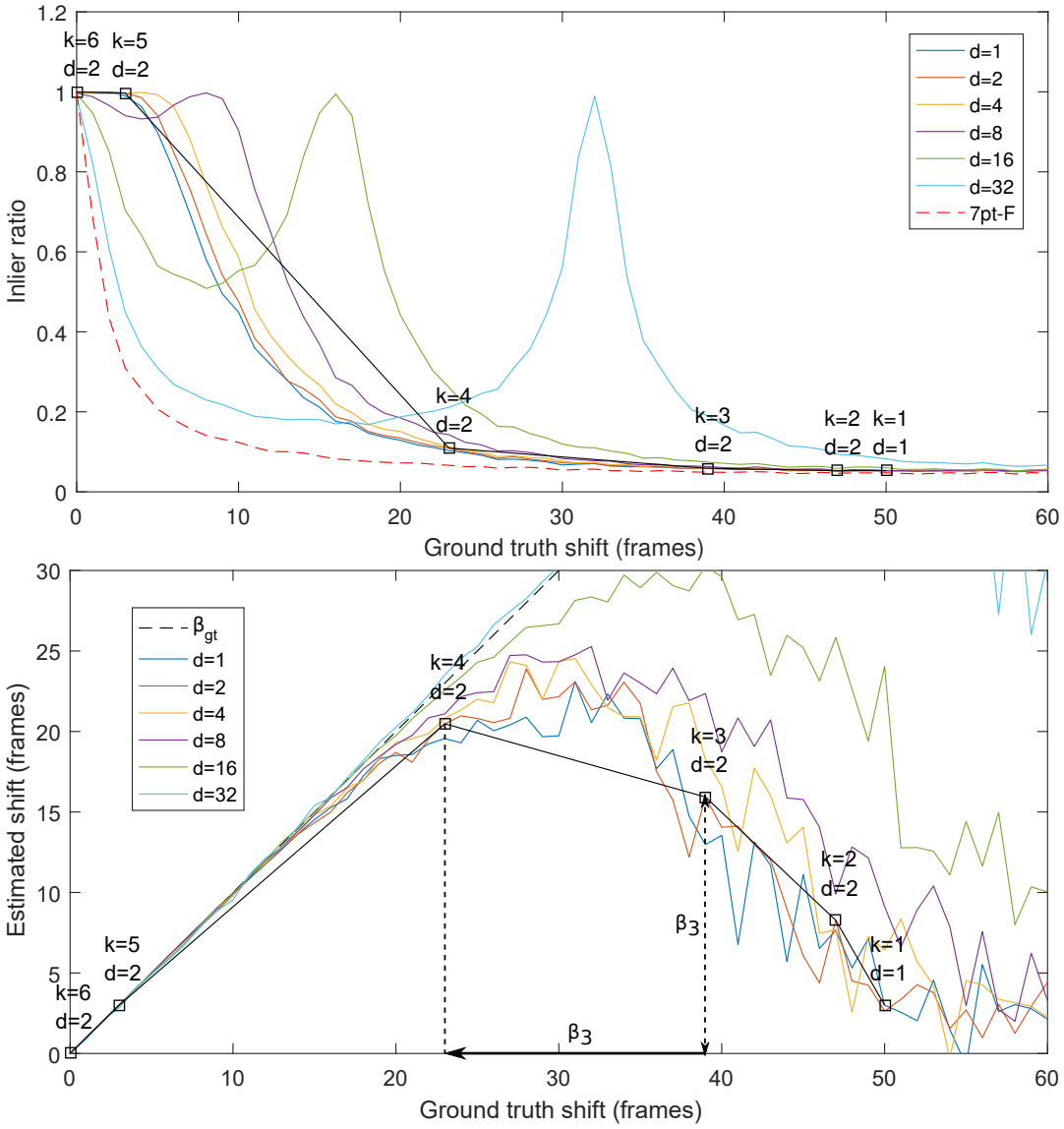

Figure 3. An example of one run of the iterative algorithm. k is the iteration number and d is the interpolation distance used. Beginning with time shift of 50 frames, the algorithm would converge in 6 iterations.

causes a significant drop in performance of the classical 7-point algorithm, resulting in errors up to 5 degrees in orientation and relative error of 5% in the translation vector.

Even for the long interpolation distances $d = 16, 32$—although not as good as for $d = 1, 2, 4, 8$—the performance is still better than that of the classical 7-point algorithm. The performance of $d = 16$ and $d = 32$ improves with increasing ground truth time shift and peaks, as expected, on time shifts 16 and 32, respectively. Note that in our iterative algorithm, we only use the right hand side of the results in the above graphs, because both d and $-d$ are used at each iteration.

3.2. Real data

The only real world dataset used for the main paper experiments for which the ground truth spatial calibration is provided is the UvA dataset. We extracted the ground truth relative $\mathtt{R}_{gt}$ and $\mathbf{t}_{gt}$ from the dataset camera matrices and compared them to the values estimated by all algorithms. Figure 5 shows the angular error of $\mathtt{R}$, measured as the rotation angle of $\mathtt{R}_{err} = \mathtt{R}^\top \mathtt{R}_{gt}$, and the relative translation error measured as $||\mathbf{t}_{gt} - \mathbf{t}||$, where both $\mathbf{t}_{gt}$ and $\mathbf{t}$ are normalized to unit lengths. Errors are averaged over 100 runs for each datapoint.

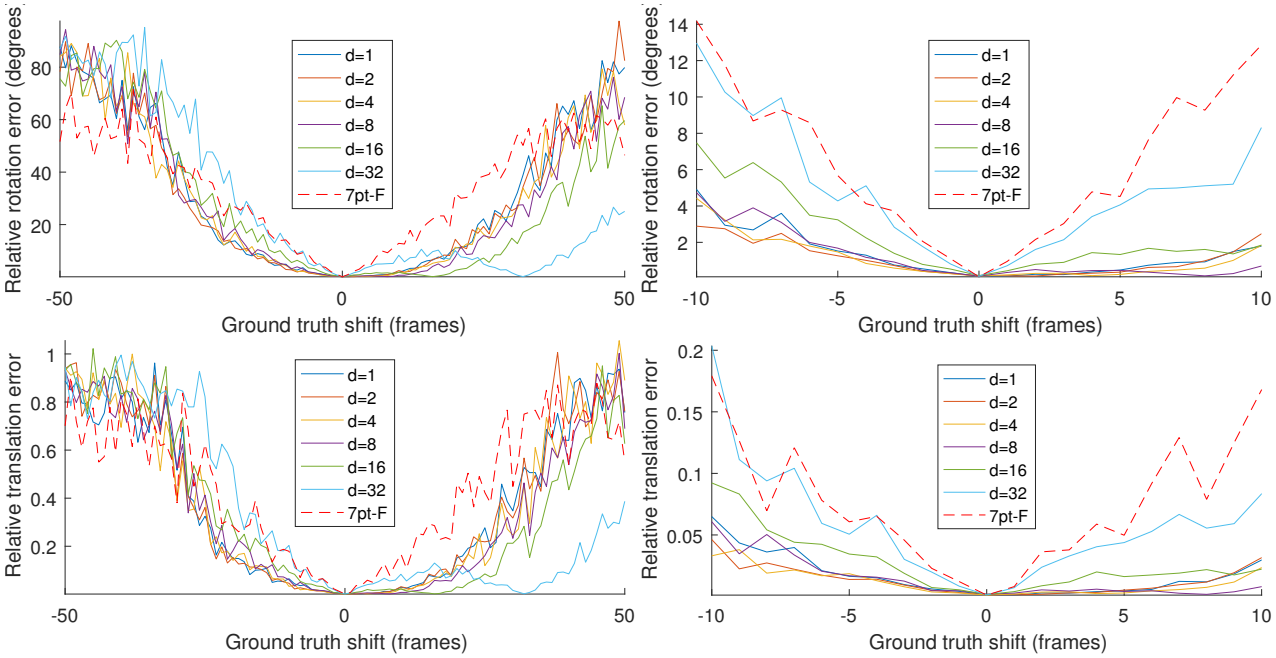

Figure 4. Error in the relative rotation and translation between the two cameras from synthetic data extracted from the computed fundamental matrix. Our solvers provide significantly better rotation and translation estimates than the classic 7-point algorithm. Note that in our iterative algorithm, we only use the right hand side of the results in above graphs, because both d and $-d$ are used at each iteration.

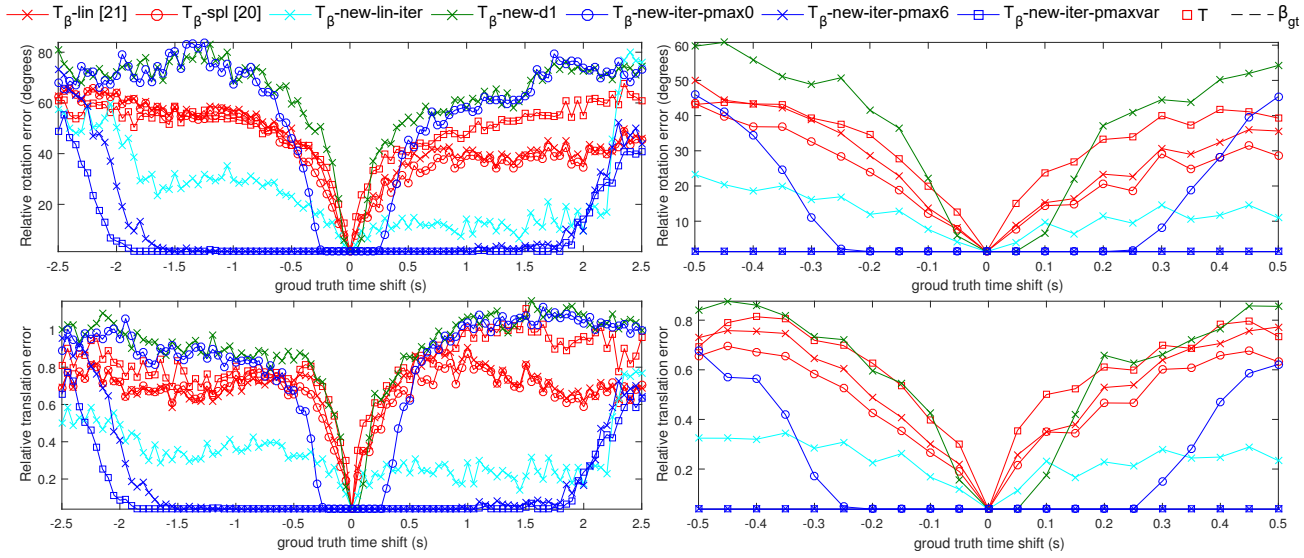

Figure 5. Error in relative rotation and translation between the two cameras from UvA dataset. All algorithms were tested, taking the resulting fundamental matrix and decomposing it into $\mathtt{R}$ and $\mathtt{t}$.

The results follow the pattern of the results in Figure 4 of the paper, where when an algorithm successfully estimated the time shift, it also provided a good geometry estimate.

Both iterative algorithms T_β-new-iter-pmax6 and T_β-new-iter-pmaxvar, which have $pmax$ large enough to cover the required time shifts, perform well over almost the entire range of time shifts. The efficient T_β-new-iter-pmax0 which iteratively uses $d = 1$ performed well up till the time shifts of 0.25 s (5 frames). T_β-new-d1, which is the solver using $d = 1$ in RANSAC, was able to estimate the geometry reliably only for a time shift of 1 frame. All the algorithms based on the 7-point algorithm, including the 7-point algorithm itself in RANSAC, performed poorly on this dataset.