Supplementary Document for "What Correspondences Reveal about Unknown Camera and Motion Models?"

Thomas Probst[1], Ajad Chhatkuli[1], Danda Pani Paudel[1], Luc Van Gool[1,2]

[1]Computer Vision Laboratory, ETH Zurich, Switzerland
[2]VISICS, ESAT/PSI, KU Leuven, Belgium

1. Additional Experimental Results

We provide additional experimental results in the document and give more details on the experimental setups for the synthetic and real data experiments.

1.1. Synthetic Data

In the synthetic experiment, we also show the model estimation error. Given an estimated model $M \in \mathbb{R}^{3 \times 3}$ and the ground-truth model $M_g \in \mathbb{R}^{3 \times 3}$, such that $\|M\|_{\mathcal{F}} = \|M_g\|_{\mathcal{F}} = 1$, we estimate the model computation error as,

$$e = \|M_g - M\|_{\mathcal{F}}$$

where $\|.\|_{\mathcal{F}}$ denotes the operator for the Frobenius norm.

Additional details. We use $m = 100$ points to evaluate on the synthetic data with noise and outliers for all methods. For *ransac* and *ransac-M*, we use 1000 iterations and use $\mathcal{L}_1$ norm to detect outliers. We use the same threshold in our method and the compared methods. Figure I shows the plots of the comparison for correspondences with noise and outliers.

The results show that the model error of *sparse-basis* stays lower than *ransac-M* particularly for low noise and outlier rate. At high noise, the *sparse-basis* model estimation deviates stronger than *ransac-M*. However, in all cases, *ransac* performs poorly due to the fact that some of the ground-truth models are the 2d rotation-induced homography.

1.2. Real Data

We show the ground-truth trajectory of the Robot car dataset sequence [2] used in the main paper in figure III. We additionally run our method *sparse-basis* and the compared method *ransac* on a driving sequence of the KITTI dataset [1]. The KITTI dataset consists of lower quality images with lower frame rate compared to the Robot car
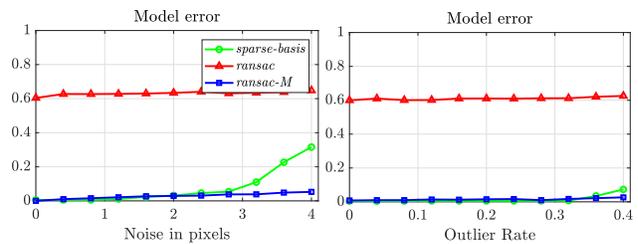
Figure I: Model error on the synthetic data.

dataset [2]. We plot the results for the KITTI dataset sequence the same way as for the TUM dataset [3] in the main paper.

Additional information on the Oxford Robotcar dataset [2] experiment.
Average number of inliers:
Number of inliers of *sparse-basis*: **95.6**
Number of inliers of *ransac*: 82.7

The plots in figure II show that our method obtains poses with very few outliers and corresponds better to the probable physical motion of the car compared to *ransac*. We further plot the execution time of our method compared to the *global* method for varying outlier rate in figure IV.

2. Proofs

In this section, we provide the proofs of the propositions (3.2, 3.3 and 4.2) presented in the main text. For the convenience of the reader we rewrite each proposition below, followed by its proof.

Proposition 3.2 *If* $z \in \{0, 1\}^n$ *represents the sparsity of the basis vector* y, *Problem (2) is then equivalent to solving the*
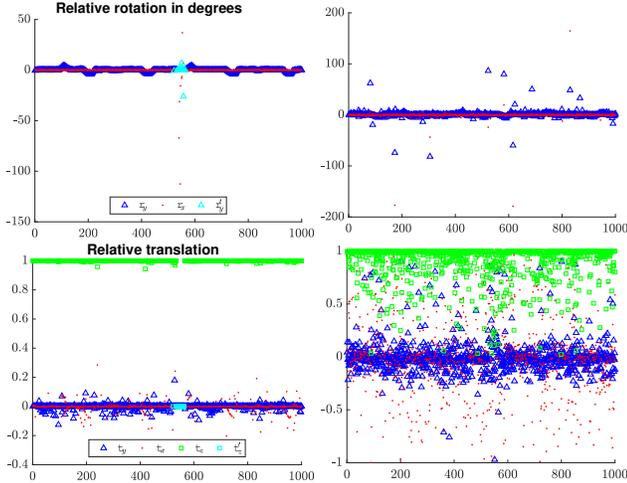
Figure II: **Two-view poses for KITTI.** We are able to capture the turns and the degenerate motions much better despite the short baseline due to the sparse basis search.
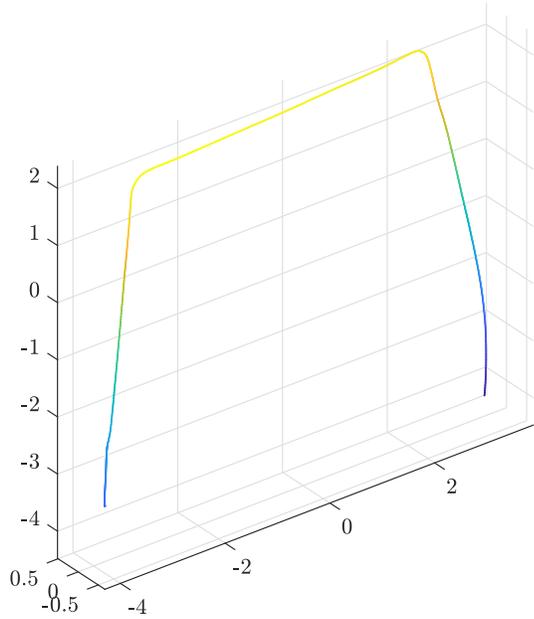
Figure III: Trajectory path of the car.

following MIP.

$$\min_{y\in\mathbb{R}^n,z\in\{0,1\}^n} \quad \sum_i z_i,$$

$$\textit{subject to} \quad \|M_d(\Omega)y\|_\infty \le \epsilon,$$

$$|y_i| \le z_i, y^\top w = 0, \quad \forall i, \forall w \in \mathcal{W}, \quad (3)$$

$$\|y\|_\infty = 1,$$

$$\sum_i z_i \ge 1.$$

Proof For each component y_i of the basis vector y, $z_i \in$
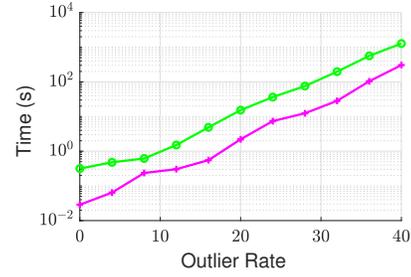
Figure IV: Mean time plot for $n = 100$ points. Our method is slower than the *global* method which is expected due to the higher number of binary variables.

$\{0, 1\}$. We now consider the two possibilities for the value of z_i:

Case I: $z_i = 0$

If $z_i = 0$, the constraint $|y_i| \le z_i$ implies that y_i must be 0.

Case II: $z_i = 1$

If $z_i = 1$, $|y_i| \le 1$. However, at the global minimum of the objective function $\sum z_i$, z_i must be 0 for $y_i = 0$. Therefore, the binary vector z represents the sparsity of the vector y and consequently problems (2) and (3) are equivalent.

Proposition 3.3 *For* $M = M_d(\Omega)$, $z \in \{0, 1\}^n$, $s \in \{0, 1\}^m$, *and* $y \in \mathbb{R}^n$, *the following MIP ensures that at least half of the correspondences respect the sparse basis obtained by solving,*

$$\min_{y,z,s} \quad \sum_{i=1}^n z_i,$$

$$\textit{subject to} \quad m_j^\top y \le \epsilon + s_j m, \qquad \forall j = 1, \ldots, m,$$

$$|y_i| \le z_i, y^\top w = 0, \qquad \forall i, \forall w \in \mathcal{W}, \quad (4)$$

$$\|y\|_\infty = 1,$$

$$\sum_i z_i \ge 1, \sum_j s_j \le m/2.$$

Proof The proof for the equivalent objective is already provided in Proposition 3.2. For the rest, we use a similar line of argument as in (3). If (4) has a solution, then the constraint $\sum s_j \le m/2$ must be satisfied. Consequently, at least half of s_j must be 0. We now consider two possibilities for (u_j, v_j).

Case I: (u_j, v_j) is an inlier:

If (u_j, v_j) is an inlier: then the constraint $m_j^\top y \le \epsilon$ must be satisfied for the sparse basis y, allowing s_j to be 0.

Case II: (u_j, v_j) is an outlier:

If (u_j, v_j) is an outlier: then the constraint $m_j^\top y \le \epsilon$ no longer holds, however, $m_j^\top y \le \epsilon + s_j m$ still holds for $s_j = 1$, where m here becomes the large scalar constant of the big-M formulation.

Therefore, solving problem (4) gives a model basis y such that at least half of the points agree. The full inlier-

outliers can be obtained by simply checking the constraint $m_j^\top y \leq \epsilon$ after recovering y.

Proposition 4.1 *For the rotation-induced 2d homography* H, *with intrinsics* $K \neq I$ *and translation* $t = 0$, *we have,*

$$HH^\top \neq I, \qquad (5)$$

where, $r \neq \pm[0\ 0\ \pi/2]^\top$.

Proof We consider $R \in SO3$ to be the camera rotation and the translation $t = 0$. We can express the homography H for the uncalibrated correspondences as:
$H = KRK^{-1}$.
By right multiplying both sides with K, we have,
$HK = KR$.
Multiplying both sides by their transpose, we obtain:
$(HK)^\top(HK) = (KR)^\top(KR)$.
We proceed to the proof by contradiction and assume $H \in SO3$. This implies, $K^\top K = R^\top K^\top KR^\top$.
Let $K^\top K = Q\Lambda Q^\top$ be the Eigen decomposition of the symmetric matrix $K^\top K$, where $Q \in SO3$ and Λ is a diagonal matrix. We therefore obtain: $Q\Lambda Q^\top = R^\top Q\Lambda Q^\top R$.
As $R^\top Q$ is an orthogonal matrix, it must represent the eigenvectors of $K^\top K$. In other words, vectors in Q and RQ must be identical up to sign and permutation. It is straightforward to see that this holds if and only if, $R = I$ or in special cases with two equal eigenvalues of $K^\top K$ when R is a permutation matrix for the first two eigen vectors. Therefore, in general $H \notin SO3$ and $HH^\top \neq I$.

References

[1] A. Geiger, P. Lenz, and R. Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *CVPR*, 2012. 1

[2] W. Maddern, G. Pascoe, C. Linegar, and P. Newman. 1 Year, 1000km: The Oxford RobotCar Dataset. *The International Journal of Robotics Research (IJRR)*, 36(1):3–15, 2017. 1

[3] J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers. A benchmark for the evaluation of rgb-d slam systems. In *IROS*, 2012. 1