Match Selection and Refinement for Highly Accurate Two-View Structure from Motion — Supplementary Material —

Zhe Liu, Pascal Monasse, Renaud Marlet

ECCV 2014

This supplementary material further supports and illustrates some of the points mentioned in paper "Match Selection and Refinement for Highly Accurate Two-View Structure from Motion", published at ECCV 2014. It is organized as follows.

- Section 1 discusses possible biases for match selection. It first explains the reason for putting the RANSAC stage as the last step of match selection and not earlier. Second, it shows the impact of using different ranking functions in match selection. In particular, it shows that the distance to the epipolar line is not suitable.
- Section 2 displays statistics on the proportion of matches that are selected by our algorithm. On our test dataset, it selects on average 61% of the matches (MS), or 78% if match refinement is applied first (MR+MS).
- Section 3 provides some visual illustrations that the improvement of calibration accuracy with our algorithm leads to a reduction of the reconstruction error of 3D points.

Please also note the following missing definition at line 182 of the paper: $e_{3D}(M) = e_{3D}(M, R_M, t_M)$.

1 Bad alternative choices for match selection due to bias

1.1 Cleaning up matches with RANSAC before selection is biased

A preliminary step, before actual match selection, consists in eliminating likely outliers (cf. paper, Section 3, "Cleaning up input matches"). It is crucial *not* to introduce any bias at this stage.

As mentioned in the paper, there would be a bias if we were to filter the matches using RANSAC and an estimated epipolar geometry. This is illustrated on Figure 1 ("ORSA before MS"), on the 6 scenes of Strecha et al.'s dataset [1]: an increase in both rotation and translation errors can be observed if match selection (MS) is preceded by ORSA [2] to first clean up input matches.

1.2 Distance to the epipolar line is biased for ranking matches

Match selection relies on a ranking function ϕ to order the matches (cf. paper, Section 3, "Ranking matches"). However, using geometrical information in function ϕ introduces a bias. In particular, it is not appropriate to use the distance to

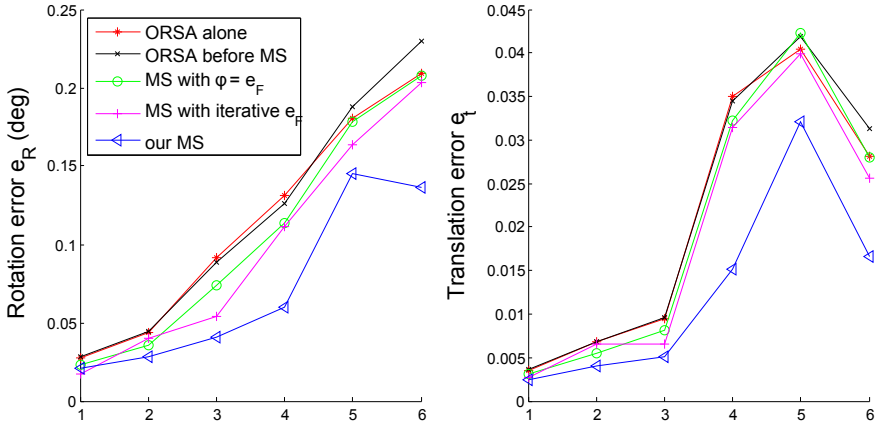

Fig. 1. Possible bias with inappropriate match selection. Left: rotation error e_R on Strecha et al.'s dataset. Right: translation error e_t. Lines are defined as follows:
-✳-: ordinary ORSA alone (an a-contrario variant of RANSAC),
-×-: MS preceded by ORSA to first clean up input matches,
-○-: MS using distance to epipolar line as ranking function ϕ,
-+-: MS using iterated distance to epipolar line and $r_{\min} = 0.4$,
-◁-: our MS method.
Scenes are ordered by increasing rotation error for ORSA alone.

the estimated epipolar line to rank the matches, e.g., to define $\phi(m) = e_F(M, m)$. This is illustrated on Figure 1 ("MS with $\phi = e_F$"), also on the 6 scenes of Strecha et al.'s dataset: results are not as good as with our unbiased ranking function.

This estimate can be slightly improved, although still with a bias. After estimating a fundamental matrix $F_{M'}$ for a given subset of matches $M' \subset M$, and considering another subset of matches $M_{sub} \subset M$, we can compute $e_F(M', M_{sub})$, the root mean square error of the distance of matches in M_{sub} to the $F_{M'}$-epipolar lines. The matches $m \in M$ can then be ordered by increasing distance $e_F(M', m)$ as a sequence $(m_i)_{1 \le i \le |M|}$ such that $i < j \Rightarrow e_F(M', m_i) \le e_F(M', m_j)$. Noting $M'_{|n} = \{m_i \mid 1 \le n\}$ the first n matches in M' and setting a minimum number of matches $N_{\min}$ to retain, we can easily find the exact optimal subset $M'^* \subset M$ with respect to $F_{M'}$:

$$
\begin{aligned}
M'^* &= \underset{\substack{M_{sub} \subset M \\ N_{\min} \le |M_{sub}|}}{\arg\min} \frac{e_F(M', M_{sub})^2}{|M_{sub}|} \\
&= \underset{\substack{M_{sub} = M'_{|n} \\ N_{\min} \le n \le |M|}}{\arg\min} \frac{e_F(M', M_{sub})^2}{|M_{sub}|} \\
&= M'_{|n^*}, \text{ with } n^* = \underset{N_{\min} \le n \le |M|}{\arg\min} \frac{e_F(M', M'_{|n})^2}{n}
\end{aligned}
$$

A linear exploration of n in $\{N_{\min}, \ldots, |M|\}$ is enough to compute n^*, and then $M'^* = M'_{|n^*}$. Starting with $M'_0 = M$, defining $M'_{k+1} = M'^*_k$, and stopping when $M'^*_k = M'_k$, we can iteratively get a good estimate for $M^*_{sub} \subset M$ defined as:

$$M^*_{sub} = \underset{M_{sub} \subset M}{\arg\min} \frac{e_F(M_{sub}, M_{sub})^2}{|M_{sub}|} \quad (1)$$

As shown of Figure 1 ("MS with iterative e_F"), results with this estimate for minimum ratio of kept points $r_{\min} = N_{\min}/|M'| = 40\%$ are slightly better on average than with $\phi(m) = e_F(M, m)$. However, experiments show that this algorithm tends to lead to values of $|M'^*_k|$ that are close to $N_{\min}$, which means it is not well behaved.

2 Number of matches kept by match selection

Match selection (cf. paper, Section 3) removes matches when they are likely to degrade accuracy. Experiments (cf. paper, Section 5) shows that the remaining matches reduce the rotation and translation error with respect to actual ground truth. It is interesting to look at the number or proportion of matches that are discarded.

This is illustrated in Figure 2. Match selection alone (MS) keeps 61% of the matches on average. But preceded by match refinement (MR), match selection (MR+MS) keeps on average 78% of the matches, as they are more reliable. Note that the number of used matches may slightly increase after match refinement because some matches that were previously discarded by the final RANSAC stage (to compute motion) are now considered as inliers. Note also that the ratio of used matched N rarely goes down to 40%, which justifies our heuristic for exploring only discrete fractions of $M_{sub}(N)$ starting from ratio $r = 0.4$ up (cf. paper, Section 3, "Exploring subsets of matches").

3 Accuracy of 3D reconstruction

We now illustrate the accuracy of our method regarding 3D reconstruction, i.e., structure. The problem is that a 3D ground truth is not available for the considered datasets. It is why we could not provide figures for the 3D error e_{3D} in the paper; we could only measure the rotation error e_R and the translation error e_t with respect to the ground truth (cf. paper, Tables 1 and 2).

To get round this problem, we construct a *pseudo ground truth* based on exact rotation and translation, but approximate point matches: for each match $m = (\mathbf{x}, \mathbf{x}')$, in images I, I' with ground-truth camera centers C, C', we construct a 3D point $X_{\mathbf{x}}(\mathbf{x}')$ as the point on line $\overline{C\mathbf{x}}$ that is the closest to line $\overline{C'\mathbf{x}'}$.

Note that we do not resort to ordinary triangulation here, e.g., mid-point of lines $\overline{C\mathbf{x}}$ and $\overline{C'\mathbf{x}'}$, gold-standard algorithm, etc. [3]. The reason is that a 3D point $X_{(\mathbf{x},\mathbf{x}')}$ originating from ordinary triangulation provides a kind of middle ground between views $\mathbf{x}$ and $\mathbf{x}'$, where $(\mathbf{x}, \mathbf{x}')$ does not try to aim at a *specific* 3D

point. As a result, it does not make sense with respect to match refinement. The fact is, as described in the paper (cf. Section 4), match refinement is asymmetric; it only moves points in image I'. It yields a new putative match $(\mathbf{x}, \mathbf{x}'')$ that tries to better locate $\mathbf{x}$ in 3D, which is different from $X_{(\mathbf{x},\mathbf{x}')}$. On the contrary, if we consider 3D points $X_{\mathbf{x}}(\mathbf{x}')$ as indicated above, match refinement make sense: we then try to get closer to the 3D ground truth location of $\mathbf{x}$ both before or after refinement.

A drawback, though, is that the error of the pseudo ground truth with respect to the unknown actual ground truth might be doubled compared to the ordinary triangulation case. We accept that and consider the measure as relative but fair in the sense that we evaluate all SfM methods with the exactly same 3D reconstruction principle.

Figures 3 and 4 show how our approach compares to RANSAC-only: reconstructed 3D points are much closer to the pseudo ground truth with our method. Note that points on the top left and top right parts of the views are not outliers; they correspond to points on the roof. Figures 5 and 6 provide a similar example.

References

1. Strecha, C., von Hansen, W., Van Gool, L., Fua, P., Thoennessen, U.: On benchmarking camera calibration and multi-view stereo for high resolution imagery. In: CVPR. (2008)
2. Moisan, L., Stival, B.: A probabilistic criterion to detect rigid point matches between two images and estimate the fundamental matrix. IJCV **57**(3) (2004) 201–218
3. Hartley, R.I., Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge University Press (2004)

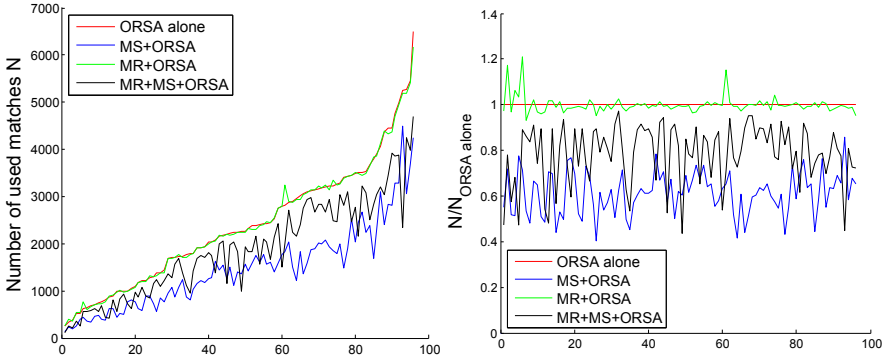

Fig. 2. Left: number of matches selected to compute motion for image pairs in Strecha et al.'s dataset. Right: proportion of selected matches. (The ratio can be greater than 1 with MR-based methods as match refinement can turn outliers that are near inliers into actual inliers.) Image pairs are ordered by increasing number of matches for ORSA alone.

Fig. 3. An image pair in Strecha et al.'s dataset.

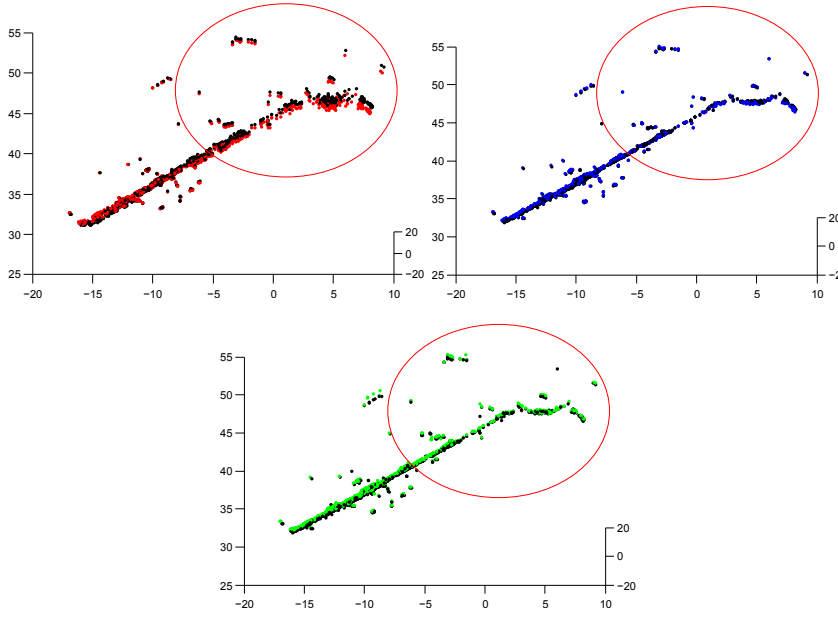

Fig. 4. View from above of the 3D points reconstructed from the image pair in Figure 3. The colors are as follows:

- **black**: pseudo ground truth,
- **red**: using ORSA alone,
- **blue**: using match selection (MS) before ORSA,
- **green**: our method, i.e., match refinement followed by match selection (MR+MS).

Fig. 5. Another image pair in Strecha et al.'s dataset.

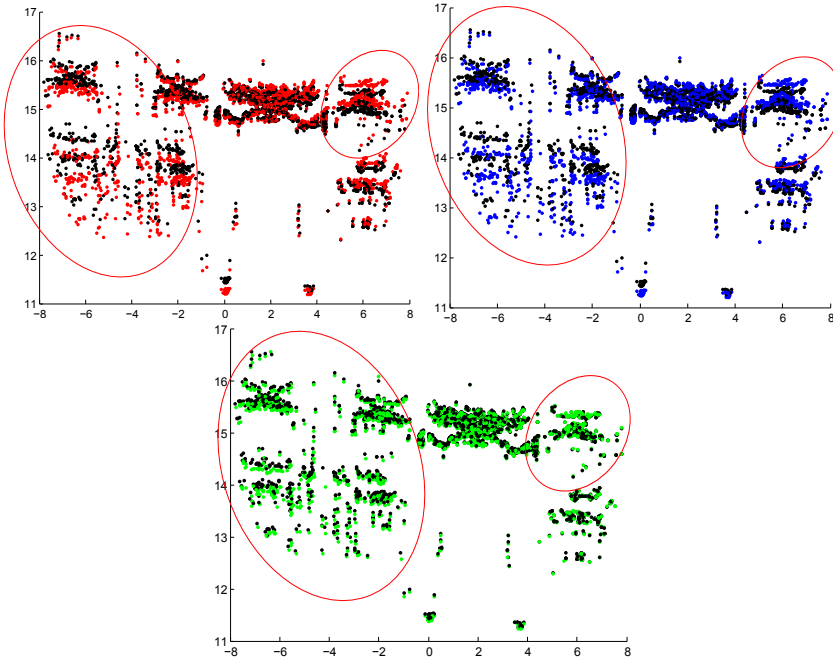

Fig. 6. Front view of the 3D point cloud reconstructed from the image pair in Figure 5. The colors are as follows:

- **black**: pseudo ground truth,
- **red**: using ORSA alone,
- **blue**: using match selection (MS) before ORSA,
- **green**: our method, i.e., match refinement followed by match selection (MR+MS).