Feature Match Selection and Refinement for Highly Accurate Two-View Structure from Motion

Anonymous ECCV submission

Paper ID 534

This supplementary material further illustrates or supports some of the points mentioned in the paper as the following organization: Section 1 discuses first the reason of putting robust methods only in the last step of match selection but not earlier; second, the impact of using different score functions in match selection, notably the distance to epipolar is not suitable. Section 2 adds some detail about how many matches of matches are kept after our algorithm, we observe the proportion of conversed match is correlated to match accuracy. Section 3 provides some visual prove that the improvement of calibration precision by our algorithm leads to decrease 3D point reconstruction error.

1 Some facts about match selection

1.1 Cleaning up matches with RANSAC before match 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, 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 epipolar line is a biased value for scoring matches

Match selection relies on a score function ϕ to order the matches (cf. paper, Section 3, "Scoring matches"). However, using geometrical information in function ϕ introduces a bias. In particular, it is not appropriate to use the distance to the estimated epipolar line as function to score the matches, i.e., to define $\phi(m) = e_F(M, m)$. This is illustrated on Figure 1, again on the 6 scenes of Strecha et al.'s dataset: results are not as good as with our unbiased score function.

This estimate can be slightly improved, although still with a bias. For this, after estimating a fundamental matrix $F_{M'}$ for a given subset of matches M', and for any other 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 be also ordered by increasing distance $e_F(M', m)$

as a sequence $(m_i)_{1\leq i\leq |M|}$ such that $i<j \Rightarrow e_F(M',m_i)\leq e_F(M',m_j)$. We also define $M'_n = \{m_i \mid 1 \leq n\}$ as the first n matches in M'. Considering now a minimum number of matches $N_{\min}$ to retain, we can easily find the exact optimal subset $M'^* \subset M$ with respect to $F_{M'}$:

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

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

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

As shown of Figure 1, 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 section (cf. paper, Section 3) removes matches because 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 Figures 2. Match selection alone (MS) keeps 61% of the matches on average. Preceded by match refinement (MR), match selection (MR+MS) now 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 illustrate here the acuracy 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, 2, 3).

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 camera centers C, C', we construct a 3D point $X_\perp$ 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 $X_\triangleright$ originating from ordinary triangulation provides a kind of middle ground between views $\mathbf{x}$ and $\mathbf{x}'$, where $(\mathbf{x}, \mathbf{x}')$ do not try to aim at a *specific* 3D point. As a result, it does not make sense with respect to point 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 $X_\triangleright$. On the contrary, if we consider 3D points $X_\perp$ 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: reconstructed 3D points are much closer to the pseudo ground truth with our method. Note that points on the top left and top right part 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)

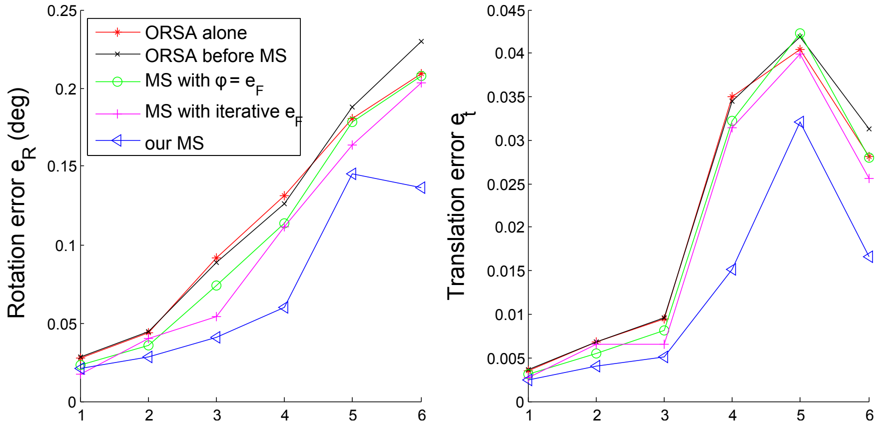

Fig. 1. Image are ordered by increasing rotation error for ORSA alone. Left: rotation error e_R. Right: translation error e_t. Lines:
-$\ast$-: ordinary RANSAC (actually ORSA) alone,
-$\times$-: MS preceded by ORSA to first clean up input matches,
-$\circ$-: MS using distance to epipolar line as score function ϕ,
-$+$-: MS using iterated distance to epipolar line and $r_{\min} = 0.4$,
-$\lhd$-: our MS method.

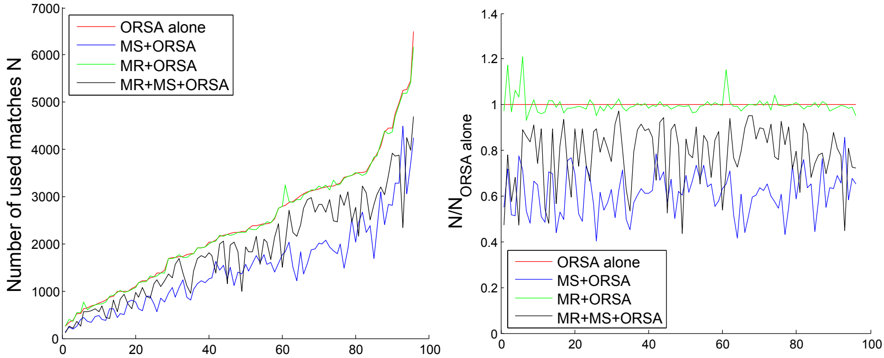

Fig. 2. Left: Number of matches used to compute motion for image pairs in Strecha et al.'s dataset. Right: Proportion of matches used to compute motion for image pairs in Strecha et al.'s dataset. Image pairs are ordered by increasing number of matches ORSA alone.

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

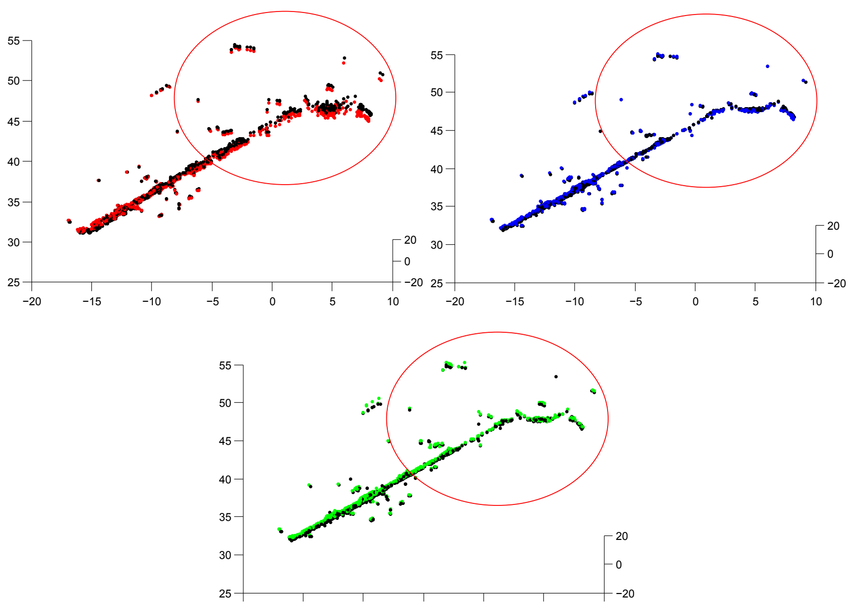

Fig. 4. View from above of the 3D points reconstructed from image pair in Figure 3. Color **black**: pseudo ground truth; red: using ORSA alone. blue: using match selection (MS) before ORSA. green: using our method, i.e., match refinement followed by match selection (MR+MS).

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

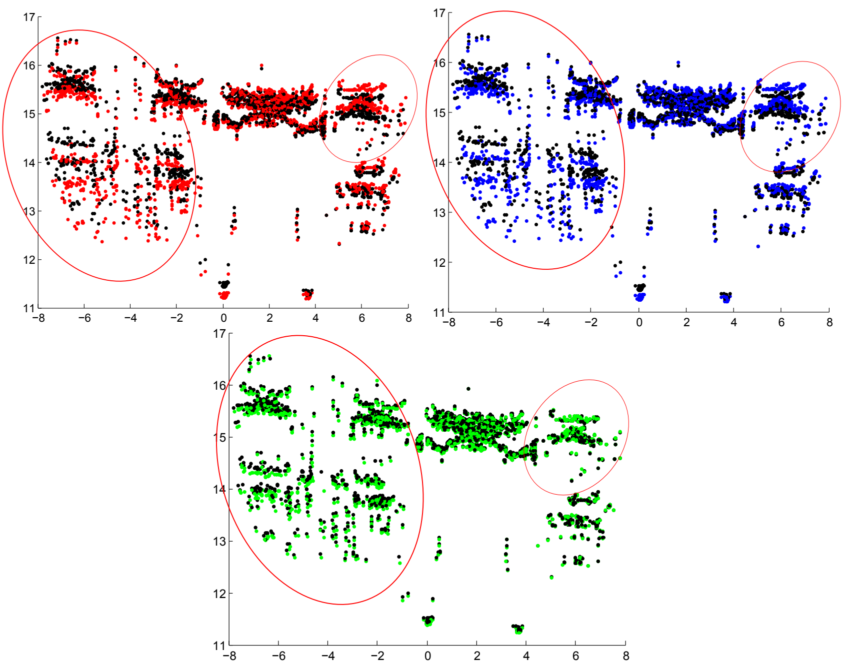

Fig. 6. Front view of the 3D points reconstructed from image pair in Figure 5. Color **black**: pseudo ground truth; red: using ORSA alone; blue: using match selection (MS) before ORSA; green: using our method, i.e., match refinement followed by match selection (MR+MS).