Robust relative pose estimation with integrated cheirality constraint

Wei Xu and Jane Mulligan
Department of Computer Science
University of Colorado
Boulder, CO 80309-0430 USA

Abstract

The **cheirality constraint***, which requires that reconstructed point correspondences lie in front of the cameras, has not typically been integrated into traditional RANSAC-based pose estimators. We have developed a new RANSAC-based relative pose estimator which incorporates the cheirality constraint not only to preempt invalid epipolar geometry hypotheses, but also as a criterion in identifying inliers for image feature correspondences. Because the application of the cheirality constraint is tightly related to the estimation of epipolar geometry, integrating them inside RANSAC can prevent inliers being falsely identified as cheirality outliers. The result is a more consistent and stable estimation which leaves denser feature correspondences for subsequent processing. Experimental comparison between the usual RANSAC-based approach and the proposed approach is performed.*

1. The Relative Pose Problem

The relative pose problem requires estimation of valid camera configurations based on correspondences from 2 or more images of the same scene. The geometric foundation of this problem is the epipolar geometry - the geometry between two different viewpoints in which the fundamental matrix F or the essential matrix E is estimated from image feature correspondences. The relative pose between the two viewpoints can further be derived from the estimated fundamental matrix F up to a projective distortion or from the estimated essential matrix E up to a Euclidean scaling. Widely used algorithms for estimation of the fundamental matrix F include the traditional 8-point algorithm [6] and its variant the 7-point algorithm [4].

For calibrated cameras we can assume the intrinsic parameters are determined beforehand via a separate camera calibration process. The camera matrix K can be assumed to be known a priori and time-invariant. Given K, the essential matrix E can be computed from the fundamental matrix by $E = K^T F K$. Another way to compute E is to normalize the coordinates of the feature points first based on K, then estimate E from five normalized point correspondences [1].

With E computed, one can treat the first viewpoint as the origin of the system and write its pose simply as $P1 = [I|0]$, where $P1$ is a 3x4 matrix and I is the 3x3 identity matrix, and then recover the relative pose of the second viewpoint $P2 = [R|t]$ from E. Here $P2$ is a 3x4 matrix, R is 3x3 rotation matrix and t is a 1x3 vector representing the translation between the two viewpoints (i.e. the baseline). As described in [7], this is done by first computing an SVD decomposition of E, which will produce four possible solutions for the relative pose of the second view, but only one of them is correct. The ambiguity of the solutions can be removed by applying the cheirality constraint, i.e. the fact that the scene points corresponding to the image feature points should be in front of both cameras [3].

2. Pose Estimation and Cheirality

In practice initial correspondence data are frequently noisy and contaminated by outliers (i.e. false feature correspondences). Least squares estimation of the epipolar geometry (F or E) over all correspondences does not work well in this case since outliers will have disproportionate influence, instead a robust estimation algorithm has to be used. One such estimator widely used to estimate the epipolar geometry is RANdom SAmple Consensus (RANSAC) [2].

RANSAC is based on the hypothesis-and-test architecture and was first used to estimate the epipolar geometry by Torr and Murray [10]. First, a hypothesis of the epipolar geometry is generated from a sample of the minimum number m of data points (i.e. image feature correspondences) randomly selected from a total of N data points. The size of the subset, m ($m < N$), depends on whether the system is estimating F or E, and which approach (hypothesis generator) is used to compute it. For example, $m = 7$ if the 7-point algorithm

978-1-4244-2175-6/08/$25.00 ©2008 IEEE

is used to estimate F, while $m = 5$ if the 5-point algorithm is used to estimate E. Next, the hypothesis is tested over all data points. Data points that agree with this hypothesis, i.e. points whose reprojection error under the estimated epipolar geometry is smaller than a threshold, are regarded as inliers, otherwise they are discarded as outliers. This kind of hypothesis-and-test operation is repeated many times, each time on a different random sample of data points and thus produces a different inlier set. Finally, epipolar geometry hypothesis with the largest inlier set is selected as the optimal output of RANSAC, and the epipolar geometry is re-estimated from this inlier set.

Note that most RANSAC-based systems find the optimal inlier set only with respect to the epipolar geometry and make no use of the cheirality constraint. Although it has been stated in the literature that the cheirality constraint can be used to identify outliers in combination with the epipolar geometry, in many papers the implementation details are omitted [3, 13, 12]. In practice we find the application of the cheirality constraint is not as simple as it appears to be. Actually, in the few papers that do provide the details on this point the approaches used are quite different.

In Torr's Structure and Motion (SaM) toolbox [11], the cheirality constraint is applied in a post-processing step after the RANSAC process to purify the cheirality of the inliers and to recover the pose R, t from E. A voting scheme based on the cheirality of RANSAC inliers for each of the 4 possible camera configurations is adopted. The configuration with the largest subset of inliers in front of the cameras, is regarded as the true relative pose. The problem with Torr's approach is that the application of the cheirality constraint is separated from the epipolar geometry estimation - in the RANSAC step itself there is no check of the cheirality of the hypotheses (i.e. the cheirality distribution among the random sample that generated the hypothesis), so the selection of the inliers may be biased if the hypothesis it relies upon is inconsistent in cheirality. Moreover, it may happen that a hypothesis inconsistent in cheirality produces more "inliers" than a consistent hypothesis because inliers are identified based only on the epipolar geometry reprojection error. In this case, the later voting step may not be valid with respect to the true cheirality configuration since the inlier selection is biased.

Nister [7] proposed an efficient solution to the 5-point relative pose problem in which the cheirality constraint is applied to recover the pose $[R, t]$ from E. This algorithm is embedded in a RANSAC framework as an epipolar geometry hypothesis generator. Comparing to Torr's approach, Nister does put the cheirality check inside the RANSAC framework, more specifically inside the hypothesis generator. However, in his algorithm one of the five points is arbitrarily selected to represent the cheirality of the whole sample that generates the hypothesis (i.e. the cheirality of the hypothesis). This may cause a problem if the selected point happens to be an outlier, or it's not an outlier but there are more outliers than inliers inside the hypothesis subset: in either case it is not representative. Thus, in Nister's approach the output of RANSAC, i.e. the identified inlier set may also be biased - the same problem as in Torr's approach.

On the estimation of the fundamental matrix F within a RANSAC framework, Chum [8] suggests eliminating hypotheses inconsistent in cheirality at the hypothesis generating stage to avoid wasting time testing over invalid hypotheses. Every time a hypothesis is generated its cheirality is checked immediately. Unless all the points in the sample have the same cheirality, the hypothesis is discarded and will not be used to select inliers. This measure guarantees the hypotheses themselves are consistent in cheirality, which is an improvement over Torr's and Nister's approaches. However, the motivation of Chum's hypothesis preemption scheme is to save computation time, while we find the same operation can be extended to the inlier identification stage to preempt data points which disagree with the hypothesis on cheirality, and produce a possibly larger inlier set and more consistent pose estimates.

3. Integrating the Cheirality Constraint

Previous work suggests that integrating the cheirality constraint into the RANSAC framework will improve relative pose estimates. Because both the epipolar geometry constraint and the cheirality constraint are geometric facts of the real world, applying them together can lead to pose estimates more likely to conform to the physical reality. We further suggest that if the cheirality constraint is integrated into a RANSAC-based pose estimator it can be applied both to eliminate incoherent hypothesis and as a criterion to identify inliers in combination with the reprojection error. The former guarantees we have a coherent and unbiased hypothesis, the latter guarantees that the identified inlier set tightly complies with the real world.

We thus propose the following improved RANSAC-based relative pose estimator:

1. Compute feature correspondences between two images taken from different viewpoints.
2. Repeat for n hypothesis-and-test trials. In each trial do the following:
    (a) Select a random sample of the minimum number of correspondences; use a hypothesis generator to generate a hypothesis of the epipolar

geometry (F or E);

(b) Check the cheirality of the sample points. If points disagree on cheirality discard hypothesis, goto (a). Otherwise record the cheirality of the sample/hypothesis C_H.

(c) Reconstruct all the correspondences and for each reconstructed point compute the reprojection error: $e_i^2 = \sum_i (d(p_1, \hat{p}_1)^2 + d(p_2, \hat{p}_2)^2)$ where d is the Euclidean distance in the image between the original correspondence points (p_1, p_2) and the projections of the reconstructed point $(\hat{p}_1, \hat{p}_2)$.

(d) Count the number of points with distance e_i less than some threshold T. These points are regarded as inliers consistent with the epipolar geometry hypothesis.

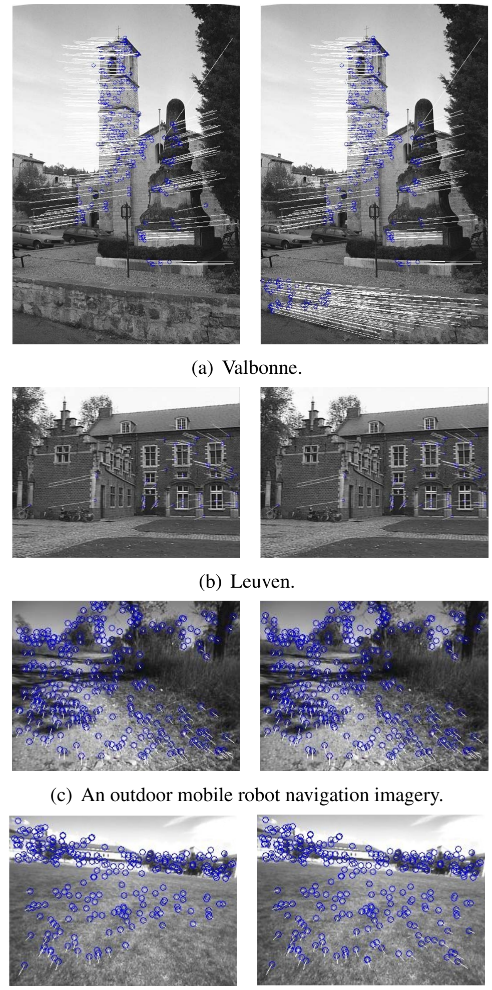
(e) For each inlier, we check its cheirality C_i. If $C_i = C_H$, it's an inlier consistent with the hypothesis in both epipolar geometry and cheirality. Record the number of cheirality-verified inliers.

3. Select the largest refined inlier set consistent with its generating hypothesis in both epipolar geometry and cheirality.
4. Use all the refined inliers to re-estimate the epipolar geometry (F or E).
5. Derive E from F (if necessary), and re-compute the relative motion $[R, t]$ from the re-estimated E.

In Chum's scheme [8] the cheirality constraint is expressed in the oriented projection geometry framework [9] and is integrated with the epipolar geometry into the so-called oriented epipolar constraint [13]. In contrast, Torr and Nister describe their work in the traditional and popular epipolar geometry framework. For generality considerations, we use Nister's approach [7] to evaluate the cheirality of the correspondences. Given an estimated E, the cheirality C_i of a correspondence has four possible configurations: $[R_a|t_u]$, $[R_a|-t_u]$, $[R_b|t_u]$ and $[R_b|-t_u]$ where $R_a = UDV^T$ and $R_b = UD^TV^T$; $[U, \Sigma, V] = \text{svd}(E)$ with $det(U) > 0$ and $det(V) > 0$; D is a predefined matrix and t_u is the third column of U. The selection of the cheirality is done by reconstructing the correspondence first, and then performing a series of checks on combinations of the 3D point coordinates (see [7] for details).

4. Experimental Results

We tested the proposed pose estimator on several well-known wide-baseline stereo pairs and on imageries from real outdoor robot navigation applications. We used the normalized 8-point algorithm [5] on SIFT features as the hypothesis generator for RANSAC. The number of RANSAC trials is set to 500, and the reprojection error threshold for identifying inliers is set to 1.2 times the mean reprojection error of the 8 points that generate the hypothesis. The focal length of the camera is set to 255 if its true value is unknown. Fig.1 shows the inliers identified by Torr's standard approach (i.e. RANSAC with post-processing on cheirality [11]) and by the proposed approach. The standard approach mistakenly identifies some inliers as cheirality outliers because it does not consider the epipolar geometry and the cheirality constraint coherently. The proposed approach can alleviate this problem and often finds more



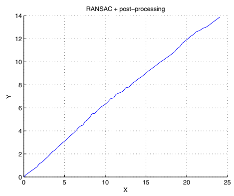
(a) Valbonne.

(b) Leuven.

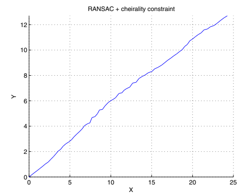
(c) An outdoor mobile robot navigation imagery.

(d) Another outdoor mobile robot navigation imagery.

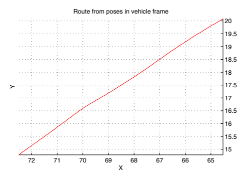
Figure 1. Inliers identified by standard RANSAC pose estimator (left side) and the proposed approach (right side). Our approach generates a larger inlier set in (a), (c) and (d), while in (b) both approaches generate the same inlier set.



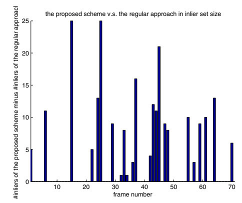
(a) Standard RANSAC.



(b) RANSAC with cheirality.



(c) Robot pose estimate.



(d) Inlier difference per frame.

Figure 2. (a) Route estimate using standard RANSAC approach. (b) Route estimate using proposed approach. Frame to frame distance in both (a) and (b) is scaled by distances from the robot GPS. (c) Ground-truth GPS route. (d) Increase in inliers for the proposed approach over standard RANSAC at each frame pair.

inliers than does the standard approach.

We also test the proposed approach on a short video sequence from a real outdoor mobile robot navigation application. The video has 71 frames recorded by a camera mounted on a mobile robot while the robot goes along an outdoor path. The camera is pre-calibrated so the camera matrix K is known. RANSAC parameters are the same as in the above experiments. Fig.2(a-c) shows the estimate of the robot's route using the standard approach and the proposed approach respectively, and the ground-truth route (from GPS device mounted on the robot). The routes in Fig.2(a,b) are computed from a concatenation of the robot's local pose estimates scaled by absolute frame to frame distance from GPS. Both computed routes are very close to the ground-truth and both are good candidates for later global adjustment (e.g. bundle adjustment). However, statistics in Fig.2(d) shows the proposed approach also generates more inliers than does the standard approach in 24 of the 70 frame pairs. The average increase over all frame pairs is 1.73% and the maximum increase is 10.61%.

5. Conclusions

In this paper we propose a new RANSAC-based relative pose estimator that integrates the cheirality constraint in the estimation of the epipolar geometry. This integrated approach can produce a lager and more consistent inlier set with respect to both the epipolar geometry and the cheirality constraint, which is useful for higher-level applications based on dense feature correspondences. This is demonstrated by our experimental comparison between a standard approach and the proposed approach where we observe an average 1.73% increase in inliers.

Acknowledgments

We gratefully acknowledge the equipment and financial support provided by the DARPA LAGR Program (DOD AFRL award no. FA8650-07-C-7702), NSF CNS-0430593 and by NIDRR RERC RecTech.

References

[1] O. Faugeras. *Three-Dimensional Computer Vision: a Geometric Viewpoint*. MIT Press, 1993.

[2] M. A. Fischler and R. C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Comm. of the ACM*, 24:381C395, June 1981.

[3] R. Hartley. Cheirality invariants. In *DARPA'93*, pages 745–753, 1993.

[4] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, New York, NY, USA, 2nd edition edition, 2003.

[5] R. I. Hartley. In defense of the eight-point algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19:580–593, June 1997.

[6] C. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293:133–135, Sep. 1981.

[7] D. Nister. An efficient solution to the five-point relative pose problem. In *Proc. 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition(CVPR'03)*, volume 2, pages 195–202, June 2003.

[8] T. W. Ondrej Chum and J. Matas. Epipolar geometry estimation via ransac benefits from the oriented epipolar constraint. In *Proc. 17th IEEE International Conference on Pattern Recognition (ICPR'04)*, volume 1, pages 112–115, Aug. 2004.

[9] J. Stolfi. *Oriented Projective Geometry: A Framework for Geometric Compuatatioins*. Academic Press, Inc., 1250 Sixth Avenue, San Diego, CA, USA, 1991.

[10] P. Torr and D. Murray. Outlier detection and motion segmentation. *Proceedings of SPIE, Sensor Fusion VI, P.S. Schenker ed.*, 2059:432–443, Aug. 1993.

[11] P. H. Torr. Structure and motion toolkit in matlab: Interactive adventures in s and m. Technical Report MSR-TR-2002-56, Microsoft Research Cambridge, June 2002.

[12] T. Werner. Constraint on five points in two images. In *Proc. 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition(CVPR'03)*, volume 2, pages 203–208, June 2003.

[13] T. Werner and T. Pajdla. Oriented matching constraints. In *Proc. 12th British Machine Vision Conference (BMVC'01)*, pages 441–450, Sep. 2001.