A FAST AND ROBUST SOLUTION TO THE FIVE-POINT RELATIVE POSE PROBLEM USING GAUSS-NEWTON OPTIMIZATION ON A MANIFOLD

Knut Hüper[†]

National ICT Australia Ltd
Systems Engineering and Complex Systems Program
Locked Bag 8001
Canberra ACT 2601, Australia
and
Department of Information Engineering, RSISE
The Australian National University,
Canberra ACT 0200, Australia
Email: Knut.Hueper@nicta.com.au

Michel Sarkis, Klaus Diepold[*]

Technische Universität München
Institute for Data Processing
Munich, Germany
Emails: [michel, kldi]@tum.de

ABSTRACT

Extracting the motion parameters of a moving camera is an important issue in computer vision. This is due to the need of numerous emerging applications like telepresence and robot navigation. The key issue is to determine a robust estimate of the (3x3) essential matrix with its five degrees of freedom. In this work, a robust technique to compute the essential matrix is suggested under the assumption that the images are calibrated. The algorithm is a combination of the five-point relative pose problem using an optimization technique on a manifold, with the random sample consensus. The results show that the proposed method delivers faster and more accurate results than the standard techniques.

Index Terms— Differential Geometry, Iterative Methods, Machine Vision.

1. INTRODUCTION

Multi-view 3D reconstruction is a well established problem in computer vision. Based on the common features of a sequence of images, it is possible to obtain the scene structure along with the camera positions [1]. In the case where the intrinsic parameters of the camera are known, the problem summarizes in determining the essential matrix (EM) between the consecutive views since it encapsulates all the information needed to extract the rigid motion, i.e. rotation matrix and translation vector.

Several techniques have been developed to compute this entity, starting from the celebrated eight-point algorithm that was formulated by Longuet-Higgins in [2] arriving at the seven-, six-, and five-point algorithm [1,3–5] and Gauss-Newton or Newton iterative techniques [6, 7].

The five points algorithm have given better results than the eight, seven, and six points algorithms in terms of accuracy and performance under noise; moreover, the obtained solution conforms to the properties of the EM and does not require a least-squares fit [8]. Nevertheless, this algorithm results in ten possible solutions. Thus, each one has to be tested in order to determine the best solution.

Since the matches are usually not ideal, the five points algorithm has to be implemented in conjunction with the random sample consensus (RANSAC). In each iteration, it is required to recover the rotation matrix and the translation vector for each of the ten solutions and then triangulate at least one point match for disambiguation [5]. In addition, to account for the imperfectness of the matches, a non-linear triangulation technique, which is usually computationally expensive, has to be employed to minimize the error [1].

In a telepresence scenario, a mobile robot or teleoperator is usually equipped with a stereo camera which might be installed at a remote location. The operator who is usually placed at a different location has to be updated about the 3D structure and pose information as fast as possible. Thus, fast and accurate techniques that compute a robust estimate of the essential matrix has to be used. This criterion might not be met by the robust version of the five-point algorithm since its computational overhead is relatively high.

In this work, a fast and robust algorithm that computes the EM is presented. It is based on optimization techniques over a manifold where a cost function has to be defined and minimized such that the minimum is the required EM [6]. In addition, the proposed technique is compared in terms of speed and performance to the five points algorithm [8]. Results show that the technique is faster and delivers more accurate results.

Section 2 presents some preliminaries needed to understand the work. Section 3 describes the iterative technique used to compute the EM using the optimization technique over a manifold assuming ideal matches. Section 4 establishes the robust technique used to compute a robust estimate of the EM. Section 5 presents the five point algorithm using ideal matches. Section 6 shows an analysis and a comparison of the proposed technique with the five-point algorithm. Finally, conclusions are drawn in Section 7.

2. PRELIMINARIES

It is well known that any two calibrated point matches $\mathbf{p}$ and $\mathbf{q}$ in two images satisfy the epipolar constraint if

$$\mathbf{q}^T \cdot \mathbf{E} \cdot \mathbf{p} = 0, \tag{1}$$

[*]This research is sponsored by the German Research Foundation (DFG) as a part of the SFB 453 project, High-Fidelity Telepresence and Teleaction.

[†]National ICT Australia is funded by the Australian Government's Department of Communications, Information Technology and the Arts and the Australian Research Council through *Backing Australia's Ability* and the ICT Research Centre of Excellence programs.

holds, where $\mathbf{E}$ is the 3x3 essential matrix. The essential matrix $\mathbf{E}$ can be written as

$$\mathbf{E} = [\mathbf{t}]_\times \mathbf{R} = \mathbf{U} \cdot \mathbf{E}_0 \cdot \mathbf{V}^T \quad (2)$$

where $[.]_\times$ is the skew symmetric matrix operator defined below, $\mathbf{t}$ is the 3x1 translation vector, $\mathbf{R}$ is the 3x3 rotation matrix between the two views, and $\mathbf{U}$ and $\mathbf{V}$ are 3x3 orthogonal matrices obtained via singular value decomposition (SVD), and $\mathbf{E}_0 = \left[\begin{smallmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{smallmatrix}\right]$. From this equation, it is easy to see that $\mathbf{E}$ has 6 degrees of freedom (DOF), encapsulated in the elements of $\mathbf{R}$ and $\mathbf{t}$ or in $\mathbf{U}$ and $\mathbf{V}$ respectively. Since the scaling factor is not important, only 5 DOF are usually considered [1]. Therefore, a 3×3-matrix is called an essential matrix if it has singular values equal to $\{1, 1, 0\}$.

From the algorithms that are mainly used to compute the essential matrix, the five-point and the Newton-type algorithms are the ones that are able to enforce the constraints in their derivations [5–7]. The eight-point algorithm does not enforce any constraint on the solution EM; thus, the result has to be corrected by an SVD to obtain a least squares solution [2]. The seven- and six- point algorithms, satisfy the singularity constraint in their derivations; however, the obtained EM has to be corrected to satisfy the singular values constraint [1,3,4]. Consequently, this fact creates another motivation for comparing the behavior of these two algorithms.

3. THE GAUSS-NEWTON ITERATIVE TECHNIQUE

The main issue here is to extract the EM from the calibrated point matches. Since the essential matrix $\mathbf{E}$ is a 3x3 matrix with 9 entries, the first step is to find a suitable parametrization of the set of essential matrices that takes into account these 5 DOF. In [6], the correct geometry and the differential manifold structure of the set of essential matrices was exploited. In the sequel, we will call this manifold of all essential matrices, the essential manifold ξ.

The rotational transformations in $\Re^3$ are represented by the elements of the special orthogonal group SO_3, which consists of all 3×3 orthogonal matrices of determinant equal to one. The set SO_3 is a 3 dimensional Lie group and its associated Lie algebra $\mathbf{so}_3$ is the set of 3×3 skew symmetric matrices which can be considered as the tangent space of SO_3 at the identity. There is a well known isomorphism that allows to identify $\mathbf{so}_3$ with $\Re^3$, defined as:

$$[\ ]_\times : \Re^3 \longrightarrow \mathbf{so}_3,$$
$$\begin{bmatrix} \omega_1 \\ \omega_3 \\ \omega_3 \end{bmatrix}_\times = \begin{bmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{bmatrix}. \quad (3)$$

Let $\mathbf{U}$,$\mathbf{V}\in SO_3$ with $\mathbf{E} = \mathbf{U}\mathbf{E}_0\mathbf{V}^T$, one possible smooth and local parametrization for ξ around an essential matrix $\mathbf{E}$ is as

$$\mu_{(\mathbf{U},\mathbf{V})} : \Re^5 \longrightarrow \xi, \mu_{(\mathbf{U},\mathbf{V})}(x) = \mathbf{U}e^{\Omega_1(x)} \cdot \mathbf{E}_0 \cdot e^{-\Omega_2(x)} \cdot \mathbf{V}^T, \quad (4)$$

where Ω_1 and Ω_2 are mappings defined respectively as:

$$\mathbf{\Omega_1} : \Re^5 \longrightarrow \mathbf{so}_3,$$
$$\begin{bmatrix} \omega_1 \\ \vdots \\ \omega_5 \end{bmatrix}_\times := \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & -\frac{\omega_3}{\sqrt{2}} & \omega_2 \\ \frac{\omega_3}{\sqrt{2}} & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{bmatrix} = \mathbf{\Omega_1}(\omega)$$

$$\mathbf{\Omega_2} : \Re^5 \longrightarrow \mathbf{so}_3,$$
$$\begin{bmatrix} \omega_1 \\ \vdots \\ \omega_5 \end{bmatrix}_\times := \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & \frac{\omega_3}{\sqrt{2}} & \omega_5 \\ -\frac{\omega_3}{\sqrt{2}} & 0 & -\omega_4 \\ -\omega_5 & \omega_4 & 0 \end{bmatrix} = \mathbf{\Omega_2}(\omega).$$

Note that the $\sqrt{2}$ factors were introduced for subsequent Riemannian geometry interpretations [6]. This parametrization of ξ ensures that one actually deals with the correct number of DOF, i.e. with five DOF, the dimension of ξ. Consequently, only five point matches in general position are required to compute the solution. Furthermore, the computed essential matrix will satisfy the singularity constraint and will have exactly two singular values equal to one.

In order to compute the required $\mathbf{E}$, a suitable cost has to be derived from Equation (1). This cost then has to be minimized taking into account the parametrization defined in (4). This is achieved by employing an optimization technique. In this work, Gauss-Newton optimization will be used to optimize over the smooth manifold of essential matrices since it has a local quadratic convergence to the solution. Thus, in each iteration a linear system has to be solved which involves both, gradient and Hessian evaluation [6][1].

4. THE PROPOSED ROBUST ITERATIVE TECHNIQUE

In order to minimize a cost derived from Equation (1), point matches between the candidate images are required. These are obtained by detecting some features in the images, e.g. using Harris feature detector, and then matching the points using any similarity measure, e.g. correlation. However, this will lead to errors in the results since a lot of false matches will be introduced. To overcome this problem, the Gauss-Newton optimization will be applied together with the Random Sample Consensus (RANSAC) algorithm [9].

Random samples of at least five points each are taken from all the primary matches found. For each sample i, an essential matrix is generated and a robust error measure is computed. Then, the hypothesis with the most number of points and the lowest distance measure is retained. The steps of the algorithm are outlined in Table 1. To increase the accuracy, the final result can then be refined using bundle adjustment [10].

Table 1. The Proposed Robust Iterative Technique

Step 1: Choose a random sample of at least 5 point matches and set the distance threshold d
Step 2: Reject the sample points that are in the critical configuration by testing the condition of the Hessian matrix
Step 3: Compute the Essential Matrix estimate using the algorithm of Section 3
Step 4: Compute the distance ϵ_i for each point match as in (6). Reject the points with distances larger than d
Step 5: Repeat Steps 1 to 4 until convergence

4.1. The Distance Measure

The distance measure that is implemented in this algorithm is the reprojection error ϵ_i, which is defined as:

$$\epsilon_i = d(\mathbf{p}_i, \hat{\mathbf{p}}_i)^2 + d(\mathbf{q}_i, \hat{\mathbf{q}}_i)^2. \quad (5)$$

[1]The Matlab implementation of this algorithm can be downloaded from *http://www.ldv.ei.tum.de/page169*

$\mathbf{p}_i$ and $\mathbf{q}_i$ are the measured point matches to be tested, while $\hat{\mathbf{p}}_i$ and $\hat{\mathbf{q}}_i$ are the true correspondences that satisfy (1) exactly.

This error when minimized results in the Maximum Likelihood estimate of the essential matrix assuming that the noise in the point matches measurements follow a Gaussian distribution. Equation (5) can be simplified by taking the first-order Sampson approximation:

$$\tilde{\epsilon}_i = \frac{\left(\mathbf{q}^T \cdot \mathbf{E} \cdot \mathbf{p}\right)^2}{\left(\mathbf{E}\mathbf{p}_i\right)_1^2 + \left(\mathbf{E}\mathbf{p}_i\right)_2^2 + \left(\mathbf{E}^T\mathbf{q}_i\right)_1^2 + \left(\mathbf{E}^T\mathbf{q}_i\right)_1^2}, \tag{6}$$

where $(.)_j$ denotes the jth entry of the corresponding vector [11].

4.2. Critical Configuration

In order to avoid the occurence of a critical configuration, the chosen point samples must not be coplanar. This can be solved by increasing the amount of points in the chosen sample since this will minimize the probability of coplanarity of the points [3]. It is easily noticed in experiments that if the chosen points are coplanar or close to coplanarity, the Hessian matrix $\mathbf{H}$ that is computed tends to be ill-conditioned; moreover, its inverse might not exist. Consequently, a critical configuration can be avoided by testing the condition number of the Hessian $\mathbf{H}$. If it is too small, the sample is rejected. From a mathematical point of view one needs at least five points which are in *general position*, a term used in algebraic geometry. But this is a property which is analytically hard to verify, a solution to this problem is therefore just to use more than five points.

5. THE FIVE-POINT ALGORITHM

When the camera intrinsic parameters are known, five points are theoretically enough to compute the essential matrix since it only has five DOF. This is one of the motivations that gives the importance to the five-point pose algorithm. This technique has been dealt with thoroughly in the literature [10, 12, 13]. Recently, an efficient variant of the five-point algorithm has been proposed and successfully applied to compute the essential matrix [5]. Like all the other techniques, the five-point algorithm uses the epipolar constraint (1) as a starting point to construct the equations. To enforce the properties of the essential matrix into the solution, the following constraint has to be included:

$$\mathbf{E}\mathbf{E}^T\mathbf{E} - \frac{1}{2}\mathrm{trace}\left(\mathbf{E}\mathbf{E}^T\right)\mathbf{E} = 0. \tag{7}$$

Since this constraint is cubic, it will give rise to a cubic polynomial through which $\mathbf{E}$ has to be determined. Thus, up to ten solutions will be obtained and each has to be tested in order to choose the best one. This is done by choosing a robust error function as in (6), and then choosing the essential matrix with the least error and the most number of points. The final solution can then be refined using bundle adjustment. Nevertheless, due to the large number of solutions that is obtained with each sample, the essential matrix in the two view case will be susceptible to errors.

To overcome this ambiguity, the authors of [5] used three views instead since the solution obtained will be unique even if the sample point matches used are coplanar (critical configuration). However, this will lead to a large computational overhead since a lot of operations have to be repeated, e.g. the SVD of each essential matrix has to be computed and then triangulation of at least one point for disambiguitation needs to be done as well.

6. ANALYSIS AND COMPARISON

The five-point algorithm had a very good precision in the results and better performance under noise when compared to the eight-, seven-, and six-point algorithms [8]. Thus, it will be used in the comparisons done with the Gauss-Newton algorithm. To obtain a robust estimate, both techniques will be used in conjunction with RANSAC as described in Section 4. Since the extraction of the rigid motion is not the goal of this work, no bundle adjustment was conducted in this analysis.

The data used for testing is formed by generating random matrices and translation vectors, forming the corresponding EMs by employing Equation (2). Then, random 3D points are generated from a field of view of 60° and then projected onto two 512×512 image planes using the previously generated rotations and translations. Unless otherwise specified, all the used data were generated using a zero-mean, unit-variance normal distribution. The analysis was repeated 50000 times to ensure the correctness of the results. All the tests were performed on a 3 GHz Pentium-4 machine using Matlab. In all of the tests, the maximum distance d allowed was set to 10^{-3} to obtain accurate results.

First the performance of both of the algorithms with respect to the noise will be tested. Thus, the sample matched points were perturbated with a zero mean white Gaussian noise, while varying the standard deviation up to one pixel error in each of the images. For each run, we computed the error measure defined by:

$$\mathrm{error} = \min\left(\frac{\mathbf{E}}{||\mathbf{E}||} - \frac{\hat{\mathbf{E}}}{||\hat{\mathbf{E}}||}, \frac{\mathbf{E}}{||\mathbf{E}||} + \frac{\hat{\mathbf{E}}}{||\hat{\mathbf{E}}||}\right), \tag{8}$$

where $\mathbf{E}$ is the true essential matrix and $\hat{\mathbf{E}}$ is the estimated one. This was done since the estimated result is defined up to a scale. The result of this test is depicted in Figure 1. Note that the number of points chosen in each RANSAC sample was set to the minimal case of 5 points. As seen in the results, the behavior of the two algorithms is almost identical; however, the average error of the proposed algorithm is lower. This result was also noticed for a higher number of points in each sample size. Taking Figure 2 for example, the number of points that was used in each RANSAC iteration is 15, which is more than enough even for the eight-point algorithm [1]. This shows that the proposed technique is more robust to noise than the calibrated five-point pose algorithm.

Another entity that needs to be tested is the dependency on the number of match points used in each sample of the RANSAC algorithm. This result can be directly depicted by looking at Figure 1 and Figure 2. It is easily noticed that the average error becomes lower for both of the algorithms.

Finally the time needed to produce the estimate of the essential matrix will be tested. In this analysis, the minimum number of 5 point matches per sample was used to show the functionality of each algorithm under the worst condition. In Figure 3, the computation time versus noise is plotted. It can be seen that the proposed algorithm is faster. The average time that was taken by the proposed method is 0.054 seconds for each of the fifty thousand cases made, while that of the calibrated five-point algorithm is 0.062 seconds. Thus, the amount of gain in time is about 14.8%.

The analysis that has been done in this section, can be easily extended to the case of three views as in [5]. The sensitivity of both of the algorithms will not change with respect to the noise. However, in each iteration, the proposed algorithm, has to perform a SVD for one essential matrix followed by triangulation of at least one point

for disambiguation. In the case of the calibrated five-point pose algorithm, these operations must be performed to every real solution of the essential matrix. For example, when the five-point algorithm finds only two real solutions, *one extra* SVD operation and triangulation must be performed in comparison to the proposed algorithm, while *nine extra* operations must be conducted in each iteration when ten real solution essential matrices are found. Thus, the proposed algorithm, will save all of these extra computations made which will introduce a gain in the overall speed. This gain mostly depends on the triangulation algorithm used and the SVD operations.

7. CONCLUSION

A fast and robust technique that extracts the essential matrix was proposed. The method is based on the Gauss-Newton optimization on a manifold where it is exploited that the set of essential matrices forms a smooth manifold. The technique is comparable to the five point pose algorithms since it preserves all the properties of the essential matrix and requires also five point matches to compute the result. In conjunction with RANSAC, the algorithm was able to deliver faster and more accurate results even in the worst case analysis: minimum number of 5 point matches per sample and high noise level. In addition, it was shown that the proposed algorithm is able to save a lot of operations, and hence computation time, that accompany the calibrated five-point algorithm in the multi-view case due to the uniqueness of its solution.

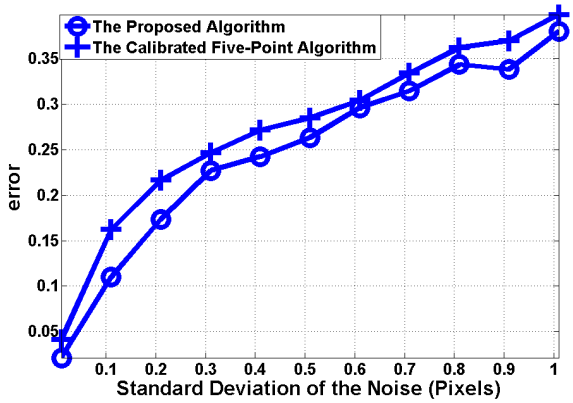

Fig. 1. Plot of the estimation error versus the noise using 5 point matches in each RANSAC sample.

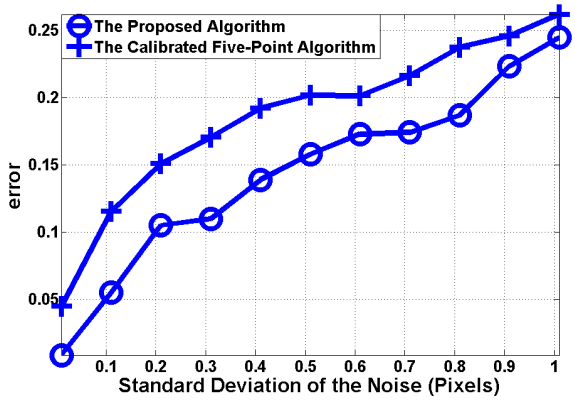

Fig. 2. Plot of the estimation error versus the noise using 15 point matches in each RANSAC sample.

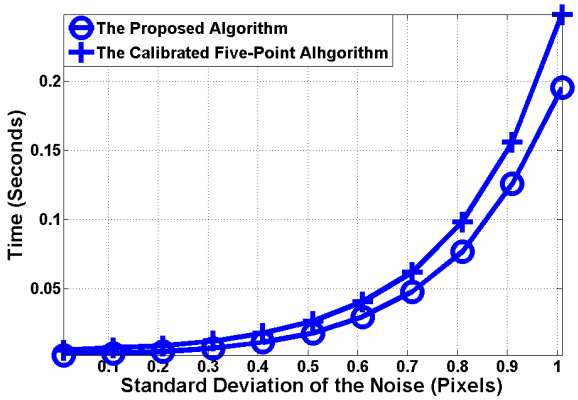

Fig. 3. Plot of the time in seconds versus the noise using the minimal case of 5 point matches in each RANSAC sample.

8. REFERENCES

[1] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, second edition, 2004.

[2] H. Longuet-Higgins, "A computer algorithm for reconstructing a scene from projections," *Nature*, vol. 293, 1981.

[3] J. Philip, "Critical point configurations of the 5-, 6-, 7-, and 8-point algorithms for relative orientation," in *TRITA-MAT-1998-MA-13*, 1998.

[4] O. Pizarro, R. Eustice, and H. Singh, "Relative pose estimation for instrumented, calibrated platforms," in *VIIth Digital Image Computing: Techniques and Applications*, 2003.

[5] D. Nistér, "An efficient solution to the five-point relative pose problem," *IEEE T. Pattern Analysis, Machine Intelligence*, vol. 26, no. 6, 2004.

[6] U. Helmke, K. Hüper, P.Y. Lee, and J.B. Moore, "Essential matrix estimation using Gauss-Newton iterations on a manifold," *Int. J. Computer Vision*, 2006, accepted for publication.

[7] Y. Ma, J. Košecká, and S. Sastry, "Optimization criteria and geometric algorithms for motion and structure estimation," vol. 44, no. 3, pp. 219–249, 2001.

[8] H. Stewénius, C. Engels, and D. Nistér, "Recent developments on direct relative orientation," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 60, no. 4, 2006.

[9] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, 1981.

[10] B. Triggs, "Routines for relative pose of two calibrated cameras from 5 points," Tech. Rep., INRIA, 1996.

[11] Z. Zhang, "Determining the epipolar geometry and its uncertainty - a review," *Int. J. Computer Vision*, vol. 27, no. 2, 1998.

[12] O. Faugeras and S. Maybank, "Motion from point matches: Multiplicity of solutions," *Int. J. Computer Vision*, vol. 4, no. 3, 1990.

[13] J. Philip, "A non-iterative algorithm for determining all essential matrices corresponding to five point pairs," *Photogrammetric Record*, vol. 15, no. 88, 1996.