A NOVEL LINEAR TECHNIQUE TO ESTIMATE THE EPIPOLAR GEOMETRY

Ebroul Izquierdo(*) *and Valia Guerra*(**)

(*) **Department of Electronic Engineering**
Queen Mary, University of London
London E1 4NS, United Kingdom

(**) **Group of Numerical Methods**
Institute of Math, Cybernetics and Physics
Havana, Cuba

ABSTRACT

The accurate reconstruction of the 3D scene structure from two different projections and the estimation of the camera scene geometry is of paramount importance in many computer vision tasks. Most of the information about the camera-scene geometry is encapsulated in the Fundamental Matrix. Estimating the Fundamental Matrix has been an object of research for many years and continues to be a challenging task in current computer vision systems. While nonlinear iterative approaches have been successful in dealing with the high instability of the underlying problem, their inherent large workload makes these approaches inappropriate for real-time applications. In this paper practical aspects of highly efficient linear methods are studied and a novel low-cost and accurate linear algorithm is introduced. The performance of the proposed approach is assessed by several experiments on real images.

1. INTRODUCTION

Recovering the 3D scene structure from its 2D projections onto perspectively different image planes is a major task in computer vision. It finds application in several technologies including 3D videoconferencing, image based rendering, medical imaging, augmented reality, robotics, immersive telepresence, etc. The basic problem consists of estimating the parameters governing the camera-scene geometry, i.e., the parameters describing the global position of the cameras with respect to the 3D scene, from a set of given corresponding points in the 2D image planes. If image correspondences and camera parameters are known, arbitrary views of the scene can be synthesised. Furthermore, using this information the 3D scene structure can be inferred by triangulation. The whole process is recursive and involves four processing steps:

1. Estimation of few correspondences with high accuracy
2. Extraction of the camera-scene geometry using the correspondence estimates from step 1
3. Estimation of dense or almost dense disparity fields using the Epipolar constraint to reduce the search complexity and to achieve better estimation accuracy
4. Intermediate views synthesis or extraction of the relative 3D structure using disparity information and camera parameters.

Regarding the first processing step a high accuracy solution to the correspondence problem for few image points has been proposed recently in [5]. The development of a reasonably good real-time approach for the second step is envisaged in this work. The third and fourth processing steps have been studied in other works [2], [4].

As mentioned before, in this work we are concerned with the estimation of the camera-scene geometry assuming that a set of few correspondence estimates is available. Due to the high instability and ill-posed nature of the calibration problem, this is perhaps the hardest of the four processing steps outlined above. The first attempt to deal with this problem was reported by Longuet-Higgins [7] in 1982. He discovered that if eight or more corresponding points in two different images are known, the camera-scene geometry can be encapsulated in a 3x3 matrix called the *essential matrix*. If the cameras are uncalibrated, i.e., if the intrinsic camera parameters are unknown, the same process can be applied to recover the scene structure up to a projective transformation [1], [2]. In this case the 3x3 matrix encoding the camera-scene geometry or Epipolar geometry is called *the fundamental matrix*. Longuet-Higgins's algorithm for the estimation of the essential matrix leads to a system of linear equations in which the nine elements of the fundamental matrix form the vector of unknowns that should be estimated. This classic method is best known as "The Eight-Point Algorithm" (EPA). The most relevant property of this approach is its linearity, and consequently its low complexity, which makes it very suitable for real-time applications.

Longuet-Higgins's work started a wave of research activities about structure from stereo and camera calibration within the computer vision community [1], [2], [8]. On the one hand most of these papers deal with fundamental theoretical aspects of the problem. On the other hand several experiments have shown that although Longuet-Higgins's linear scheme is optimal from a computational point of view, it is unstable when applied to real data. More recent papers, including the works of Xu and Zhang [8], introduce nonlinear constraints and robust statistics to deal with the extreme distortion susceptibility to non-Gaussian noise and outliers of the linear model. The major drawback of these methods is that the linearity of the original model gets lost and the iterative solution of a high non-linear model becomes necessary. This brings with it an essentially higher workload and consequently the use of these methods in real-time applications becomes unrealistic.

In 1997 Hartley [3] presented a careful and sound analysis of the reasons for the high instability of the EPA. He showed that this fact mainly originated in non rigorous implementations in which aspects of linear algebra and numerical analysis are totally ignored. In his paper Harley states: "The poor performance of the EPA can probably be traced to implementation that does not take sufficient account of numerical considerations, most specifically the condition of the set of linear equations being involved". In this statement the word "condition" refers to the condition number of the matrix of the linear system to be solved. Furthermore, Hartley [3] showed that by applying a kind of normalization to the input data, the EPA delivers results comparable with computationally more expensive non-linear methods but at very low computational cost. Basically, Hartley's "Normalized Eight-Point Algorithm" (NEPA) consists of a modification of the reference co-ordinate system for image points via a simple translation, followed by a straightforward scaling in order to considerably reduce the condition number of the matrix of the underlying linear system.

In this paper a novel accurate and extremely low-complexity linear approach for camera-scene registration is introduced. The introduced approach reinforces Hartley's statement about the cause of the poor performance of the EPA by using efficient numerical schemes and some fundamental results on data normalization and preconditioning. Our work is close in spirit to Hartley's work.

We are aware of the fact that neither Hartley's approach nor the extensions proposed in this paper outperform the best non-linear iterative algorithms. On the other hand several computer experiments and comparisons show that the results obtained by applying the improved linear approaches only slightly differ from those obtained with iterative techniques. Our objective has been to develop and implement a stable method showing good performance, i.e., delivering good results at very low computational cost and suitable for real-time applications.

2. FUNDAMENTALS

Given two stereo images the relationship between two corresponding points $X_l = (x_l, y_l, 1)^T$ and $X_r = (x_r, y_r, 1)^T$ can be formulated as:

$$Z_r M_r^{-1} X_r = Z_l R M_l^{-1} X_l + \Psi, \qquad (1)$$

with *M* the camera intrinsic matrices, *R* the rotation matrix from left to right camera coordinate systems and Ψ the translation vector from the origin of the left camera coordinate system to the origin of the right camera coordinate system. Taking $U_l = (u_l, v_l, 1)^T = M_l^{-1} X_l$ and $U_r = (u_r, v_r, 1)^T = M_r^{-1} X_r$ (the normalized image coordinates), the cross-product of (1) with the translation vector $\Psi = (t_1, t_2, t_3)^T$ followed by the inner product with U_r leads to

$$U_r E U_l = 0, \qquad (2)$$

with $E = \tau \cdot R$, $\tau = \begin{pmatrix} 0 & -t_3 & t_2 \\ t_3 & 0 & -t_1 \\ -t_2 & t_1 & 0 \end{pmatrix}$. Equation (2) is called the epipolar equation. This equation can be written as an homogeneous linear equation in the nine unknown elements of the essential matrix *E:* $h^T \hat{E} = 0$, with $h = (u_l u_r, u_r v_l, u_r u_l v_r, v_l v_r, v_r u_l, v_{l,} 1)^T$ and $\hat{E}$ the nine-dimensional vector, whose elements are the coefficients of the essential matrix. For *n* corresponding points we obtain a system of linear equations of the form:

$$H_n \hat{E} = 0. \qquad (3)$$

Since $\hat{E}$ is defined up to a scale factor eight corresponding points are sufficient to calculate the essential matrix. Moreover, $rank(H_n) = k$ and *k<9*, because in the other case, i.e., when k=9, equation (3) only poses the trivial solution. Most methods for solving (3) assume that the errors in the input data are Gaussian distributed with zero mean, and use the $L_2 - norm$ to solve the problem. Unfortunately, the least squares method is extremely unstable because the distribution of errors is not Gaussian and the initial data can contain outliers. Thus, the straightforward linear EPA method is extremely sensitive to noise in the input data. This fact led to the belief that the linear approach fails when real data is used. Consequently, iterative methods in which high nonlinear constraints are used to deal with the instability of the linear model became widely spread [8].

The first attempt to challenge (3) using real data and an EPA-like linear model was done by Hartley [3]. He demonstrated that the linear model can perform as well as iterative methods if linear algebra tools and a rigorous numerical analysis is used in the algorithm design. Basically, Hartley's algorithm can be outlined as follows:

1. Transform the image coordinates according to transformations $\tilde{U}_r = T_r U_r$ and $\tilde{U}_l = T_l U_l$
2. Find the essential matrix $\tilde{E}$ corresponding to the transformed coordinates
3. Set $E = T_r^T \tilde{E} T_l$.

Although Hartley's approach opened a new way to solve the calibration problem without using complex non-linear techniques, it still applies the standard and expensive SVD. This fact together with the complexity of the performed transformation make it difficult to run Hartley's algorithm in real-time using standard equipment. In the next section we establish the most important results concerning Hartley's proposed transformations, and use that analysis to derive more efficient algorithms. A new kind of transformation is proposed to solve the linear problem maintaining stability and accuracy but extremely reducing the complexity.

3. THE PROPOSED ALGORITHM

In the following some important results derived from our study are outlined. These results not only add value to the rigor of the analysis concerning the camera-scene registration problem, but they demonstrate the ideas behind a more efficient algorithm. For the sake of clarity, and perhaps lack of space, these assertions will be presented without proof. Readers interested in the theorem's proofs and additional results are referred to [6].

The objective of the next two theorems is to give mathematical rigor to Hartley's scaling approach. It is well-known that the matrix H_n of the system usually has rank eight. Notice that only in rare cases in which the configuration of the image points has very specific structure, the rank of H_n is less than eight [2]. This fact is used to prove that Harley's algorithm (NEPA) leads to the same solution (up to a scale factor) of the original linear system (3).

Theorem 1: *Let E be the essential matrix obtained by solving (3) using the original (non-transformed) data. If* H_n *has rank* *eight, then solving (3) using NEPA leads to a matrix* $\tilde{E}$ *satisfying*

$$E = \lambda \tilde{E}, \quad \text{for } \lambda \neq 0.$$

Hartley proposes two different scaling techniques: isotropic and non-isotropic. The next theorem states a very important property of these scalings.

Theorem 2: *Hartley's non-isotropic scaling is optimal in the sense of reducing the condition number of the system.*

Using the analytical study presented above, we have devised a new and essentially more simple implementation of Hartley's scaling approach. Three basic points tackled in the new approach are:

- The use of a more efficient numerical algorithm to solve the underlying linear system
- To perform the scaling directly on the matrix H_n rather than on the input data
- To avoid products of badly conditioned matrices.

The NEPA scheme is based on the estimation of the least eigenvector of the matrix obtained from $H_n^T.H_n$. This is achieved by performing SVD of this matrix product. We found that there are essentially more efficient methods to estimate the least eigenvector. Most of these methods have been developed in the context of the estimation of the condition number of a matrix. For this reason they are known as techniques of *condition estimator* type. Basically, we propose to use inverse iteration algorithms or alternatively the CCVL. A detailed description of these techniques can be found in [6]. Using the inverse iteration technique in conjunction with a QR-decomposition leads to a highly efficient implementation since only few iterations are necessary to obtain an accurate approximation of the least eigenvector. The major advantage of this technique compared with the SVD is the low computational cost. The cost of both methods measured in flops is given in the second row of table 1. To give more numerical stability to the new approach and further reduce the computational cost, products of type $H_n^T.H_n$ are avoided. Notice that Harley's technique begins by doing this product. Since $cond(H_n^T.H_n) = [cond(H_n)]^2$, this leads to an unnecessary worsening of the condition of the system. In this context the proposed algorithm is more efficient compared with NEPA. The computational difference between the two techniques is summarized in table 1.

Using the NEPA schema two scaling matrices should be estimated, one for each set of input points on the right and left images respectively. In contrast to that, in the sequel an alternative comprising just one scaling matrix is proposed. For a given 9x9 scaling matrix *S* the introduced technique consists of the following algorithmic steps:

- Define $\tilde{E} = S^{-1}\hat{E}$
- Solve the transformed linear system $H_n S\tilde{E} = 0$
- Find $\hat{E} = S\tilde{E}$
- Estimate the Essential Matrix *E* from the vector $\hat{E}$

The question now is how to find the "best matrix *S*" in the sense of improving the condition number of $H_n S$. Following the argumentation that leads to Hartley's non-isotropic scaling, one should be tempted to choose $S = R^{-1}$, with *R* the matrix taken from the *QR*-decomposition of H_n. In this case it is easy to see that:

$$cond(H_n S) = cond(H_n R^{-1}) = cond(Q) = 1.$$

Consequently, *R* is optimal in the sense of improving the condition of $H_n S$. Unfortunately, things are not as simple as that. In this case the system $\tilde{E} = S^{-1}\hat{E}$ implicit in the algorithm has the same bad condition properties as the original system (3). To avoid this dilemma we should look for a scaling matrix that on the one hand reduces the condition of the system and on the other hand is numerically stable when it is inverted. These two conditions can be achieved by using diagonal matrices. Let us define *S* as:

$$S = D^{1/2} = diag(d_1^{1/2}, d_2^{1/2}, \dots, d_9^{1/2}), \quad d_j = \left(\|H_{nj}\|_2\right)^2,$$

for *j=1,2,…,9*, (4)

where H_{nj} is the j-th column of H_n and $\|.\|_2$ is the L_2-norm. Since $H_n S^{-1}$ has columns of unit length it is straightforward to prove that among all the diagonal scaling this scaling minimizes the condition of the system. Furthermore, the estimation of $\hat{E}$ from $\tilde{E}$ is numerically stable as the calculation of S^{-1} is trivial. Additionally, the amount of operations is lower compared with the scaling used in the NEPA approach.

One question remains open: does the Essential Matrix obtained in this way correspond to the original Essential Matrix of the system (3)? This question is answered by the following statement

Theorem 3: *Let S be a non-singular matrix,* $\hat{E}$ *and* $\tilde{E}$ *the solutions of the systems* $H_n\hat{E} = 0$ *and* $H_n S\tilde{E} = 0$ *respectively. If* H_n *has rank eight, then*

$\hat{E} = \lambda S\tilde{E}$ *for* $\lambda \neq 0$.

This means the original Essential Matrix and that obtained using the proposed technique are the same up to a scaling factor.

Algorithmic Step	#flops in NEPA	#flops in this Method
Matrix Multiplication	$O(n^2)$	*None*
System Solution	*324n+5832*	*162n-314*

Tab. 1: Computational cost in flops for Hartley's method and the proposed method for *n* corresponding points

Assuming that *n* corresponding points are known, the preceding arguments led to the design of the following algorithm for the estimation of the Epipolar Geometry:

- Build the matrix H_n as given by (3)
- Define the matrix *S* as given in (4)
- Use the CCVL technique to estimate $\tilde{E}$ from $H_n S\tilde{E} = 0$
- Find $\hat{E} = S\tilde{E}$ and build the Matrix E' from the vector $\hat{E}$
- Select the matrix of rank 2 closest to E' as the Essential Matrix.

This linear algorithm provides accuracy in the results at low computational cost. The proposed scaling acts directly on the system to be solved (3). It uses a preconditioning matrix *S*, which is optimally defined in the sense of improving the condition of the system. Since the computational cost is extremely low, the method can be easily run in real-time.

Fig 1. Epipolar lines estimated for the scene GWEN.

4. SELECTED RESULTS

The proposed technique has been evaluated by estimating the epipolar geometry for real images. Several comparisons with results obtained using other previously reported methods have been conducted. In these comparisons both performance and complexity have been evaluated. A comprehensive report of this comparative evaluation is in preparation [6]. The evaluation includes results obtained using different techniques from the literature [8] as well as a direct comparison with the NEPA scheme [3]. Although, the methods introduced in [8] were designed to satisfy physical constraints inherent to the calibration problem, the new algorithm supplies similar results. For reasons of space, in this article only few results obtained for the scene GWEN can be reported. The epipolar lines obtained with the proposed algorithm are shown in the image at the bottom of Fig. 1. In this representation both stereo images are shown in the background. Superimposed on the left image (top) the matched points are highlighted, on the right image (bottom) the epipolar lines corresponding to these points are drawn.

5. CONCLUSIONS

A novel, accurate and simple linear approach for camera-scene registration has been presented. Some fundamental results that give mathematical rigor to Hartley's view about numerical implementation of the classic EPA scheme are proven. The introduced algorithm uses techniques of condition estimator type to approximate the least singular value and its corresponding singular vector, instead of explicitly performing an essentially more expensive singular value decomposition. To improve the condition of the system a diagonal scaling on the matrix H_n is introduced. Another important feature of the proposed algorithm is to strive for a direct solution to the underlying linear system $H_n E = 0$, avoiding the use of products of type $H_n^T . H_n$. The performance of the method has been assessed by estimating the Epipolar Geometry from stereoscopic images.

5. REFERENCES

[1] O. D. Faugeras, Q. T. Luong and S. J. Maybank, Camera self-calibration: Theory and experiments, in *Proc. 2nd European Conf. On Computer Vision (ECCV)*, pp. 321-334, Springer-Verlag, Berlin, Heidelberg, New York, 1992.

[2] O. D. Faugeras, *Three-dimensional computer vision*, MIT Press, Cambridge, 1993.

[3] R. I. Hartley, "In defense of the eight-point algorithm", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, 1997, pp. 580-593.

[4] E. Izquierdo, "Disparity/Segmentation Analysis: Matching with Adaptive Windows and Depth Driven Segmentation", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 9, No. 4, Jun. 1999, pp. 589-607.

[5] E. Izquierdo, "High accuracy solution to the correspondence problem", *Electronic Letters*, Vol. 36, No. 11, May 2000, pp. 948-949.

[6] E. Izquierdo and V. Guerra, "On the estimation of the epipolar geometry by linear algorithms", in preparation.

[7] H. C. Longuet-Higgins, "A computer algorithm for reconstructing a scene from two projections", *Nature 293,* 1981, pp. 133-135.

[8] G. Xu and Z. Zhang, "Epipolar Geometry in Stereo, Motion and Object Recognition", *Kluwer Academic Publishers*, London, 1996.