Five-Point Motion Estimation Made Easy

Hongdong Li and Richard Hartley

RSISE, The Australian National University.
Canberra Research Labs, National ICT Australia.*
Email: firstname.lastname@anu.edu.au

Abstract

Estimating relative camera motion from two calibrated views is a classical problem in computer vision. The minimal case for such problem is the so-called five-point problem, for which the state-of-the-art solution is Nistér's algorithm [1][2]. However, due to the heuristic nature of the procedures it applies, to implement it needs much effort for non-expert user. This paper provides a simpler algorithm based on the hidden variable resultant *technique. Instead of eliminating the unknown variables one by one (i.e, sequentially) using the Gauss-Elimination as in [1], our algorithm eliminates many unknowns at once. Moreover, in the equation solving stage, instead of back-substituting and solve all the unknowns sequentially, we compute the minimal singular vector of the coefficient matrix, by which all the unknown parameters can be estimated simultaneously. Experiments on both simulation and real images have validated the new algorithm.*

1. Introduction

This paper studies the classical problem of estimating relative camera motion from two views. Particularly, we are interested in the *minimal case problem*. That is, to estimate the rigid motion from minimal *five* corresponding points of two views. Since the relative geometry between two view is faithfully captured by an essential matrix E, which is an real 3 by 3 homogeneous matrix, the task is therefore equivalent to estimating the essential matrix from five points.

The classical way of estimating E is the eight-point algorithm [3]. As it is a linear algorithm and its accuracy is reasonably good, it is widely adopted as a benchmark algorithm. Admitting this, why do we need a five-point algorithm? The justifications lie in both theoretic and practical aspects. In theory, the significance of having a minimal-case solver is quite obvious, which enable us a deeper understanding to the problem itself. In practice, a five-point algorithm also offers many benefits in reality. As demonstrated in [12], (1)this five-point algorithm suffers fewer types of "critical surface". For example, a arbitrary plane is not dangerous for the 5-pt algorithm; (2)when a 5-pt algorithm is used as a hypothesis-generator for RANSAC, its needs much less samples than by the 8-pt algorithm; (3) the estimation accuracy by the 5-pt estimation is also higher than by a 8-pt algorithm, because the minimal solver has better exploited all available geometric constraints of the problem.

Previously, we proposed a simple resultant-based algorithm ([5]) to solve the six-point focal-length problem [6]. In that paper we argued that the resultant technique was *not* an individual success, but can be generally applicable. This paper substantializes such argument by showing that with the same technique the five-point problem can also be solved simply and elegantly.

*NICTA is funded through the Australian Government's Backing Australia's Ability Initiative, in part through the ARC.

2. Previous Work

It has been known for long time that from five corresponding points of two calibrated images one can estimate the relative motion between the two views ([8], [9], [4]). However, despite the theoretical result, there was very few practical algorithmic implementations until very recently ([10] [1]).

The state-of-the-art algorithm is Nistér's five-point algorithm proposed in [1] or [2], which is based on a modified Gaussian-Jordan elimination procedure. His algorithm is founded on Philip's previous work [11] but made significantly improvements. However, the particular elimination sequences he proposed is *ad hoc* and heuristic. For example, different sequences of elimination may lead to different computer programs. A careful reader may notice that in Nister's two versions of implementation (see section-4,cf.[1] and [2]) there are subtle differences. Stèwenius later revised this algorithm by porting the problem into the $\mathbb{Z}_p$ domain [12]. The Gröbner basis technique is used to find suitable elimination sequences. This method is interesting,

Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06)
0-7695-2521-0/06 $20.00 © 2006 IEEE

promising and quite general.

This paper provides a (hidden-variable) resultant-based algorithm. It is so simple that is almost self-explained. Compared with existing algorithms, our new algorithm is easier to implement and to understand. Non-expert user can apply it with comfort. Rather than eliminating irrelevant unknowns one by one (i.e., sequentially), our algorithm eliminates all the irrelevant unknowns all at once. Furthermore, in the equation solving stage, we show that by computing the minimal singular vector of the coefficient matrix, all the solutions can be found at once.

3. Theory of Motion Estimation

The reader is assumed to be familiar with camera calibration and epipolar geometry (or, is referred to [3]). To save space we simply list some fundamental results without explanation.

Consider a camera with constant intrinsic matrix $\mathtt{K}$ observing a static scene. Two corresponding image points $\mathbf{m}$ and $\mathbf{m}'$ are then related by a fundamental matrix $\mathtt{F}$:

$$\mathbf{m}'^{\mathbf{T}}\mathtt{F}\mathbf{m} = 0. \tag{1}$$

A valid $\mathtt{F}$ must satisfy the following *cubic* singularity condition:

$$\det(\mathtt{F}) = 0. \tag{2}$$

If the camera is fully-calibrated, then the fundamental matrix is reduced to an *essential matrix*, denoted by $\mathtt{E}$, and the relationship becomes:

$$\mathtt{K}^{-\mathbf{T}}\mathtt{E}\mathtt{K}^{-1} = \mathtt{F}. \tag{3}$$

Since an essential matrix $\mathtt{E}$ is a faithful representation of the motion (translation and rotation, up to a scale), it has only five DOFs. Consequently, to be a valid essential matrix $\mathtt{E}$, it must further satisfy two more constraints, which are characterized by the following result:

$$2\mathtt{E}\mathtt{E}^{\mathbf{T}}\mathtt{E} - tr(\mathtt{E}\mathtt{E}^{\mathbf{T}})\mathtt{E} = 0. \tag{4}$$

This actually gives nine equations in the elements of $\mathtt{E}$, but only two of them are algebraically independent. Given five corresponding points, there are five epipolar equations eq.(1), plus the above nine equations and the singularity condition eq.(2), one therefore has enough equations to estimate the essential matrix.

4. Overview of Nistér's 5pt Algorithm

Nistér's 5pt algorithm proceeds as follows (based on the CVPR version [1], rather than the PAMI version[2], as the former is more straightforward).

1. Writing down the epipolar equation eq.(1) for the five points, one can get a null-space representation $\mathtt{F} = x\mathtt{F}_0 + y\mathtt{F}_1 + z\mathtt{F}_2 + w\mathtt{F}_3$, where $\mathtt{F}_i, i = 0,1,2,3$ are the null-space bases. Using the fact that $\mathtt{F}$ is homogeneous, without loss of generality, let $w = 1$.
2. Using the nine equations of eqs.(4), form a 9×20 coefficient matrix (PAMI version uses a 10×20 matrix [2]) corresponding to a monomial vector:
$$\begin{aligned}&[x^3, y^3, x^2y, xy^2, x^2z, x^2, y^2z, y^2, xyz, xy,\\ &\quad xz^2, xz, x, yz^2, yz, y, z^3, z^2, z, 1].\end{aligned} \tag{5}$$
Then apply Gaussian-Jordan elimination to the 9×20 matrix, reduce it to an upper triangle form.
3. Use some *ad hoc* procedures to extract the determinants of two 4×4 matrices (PAMI version uses one 3×3 matrix [2]), followed by a second stage of elimination. Eventually a 10-th degree polynomial is obtained, from which one then obtains 10 roots of z.
4. Back-substituting those real roots, one can solve other unknowns sequentially.
5. Recover the essential matrix, and extract the corresponding motion vectors of rotation and translation ([15]).

5. Derivation of Our New 5pt Algorithm

5.1 Hidden Variable Resultant

Our algorithm is based on the *hidden variable* technique([13]), which is probably the best known resultant technique for algebraic elimination, and very easy to implement. The purpose of this technique is to eliminate variables from a multivariate polynomial equation system. Its basic idea is as follows.

Given a system of M homogeneous polynomial equations in N variables, say, $p_i(x_1, x_2, ..., x_N) = 0$, for $i = 1, 2, ..., M$. If we treat one of the unknowns (for example, x_1) as a *parameter* (that is, we *hide* the variable x_1), then by some simple algebra we can re-write the equation system as a matrix equation: $\mathtt{C}(x_1)\mathbf{X} = 0$, where the coefficient matrix $\mathtt{C}$ will depend on the *hidden variable* x_1, and the $\mathbf{X}$ is a vector space consisting of the homogeneous monomial terms of all other N-1 variables (say, $x_2, x_3, \cdots, x_N$). If the number of equations equals the number of monomial terms in the vector $\mathbf{X}$ (i.e. the matrix $\mathtt{C}$ is square), then the equation system will have non-trivial solutions *if and only if* $\det(C(x_1)) = 0$.

By such procedures, one thus eliminates N-1 variables all at once. For more information the interested reader is referred to [13].

5.2 Algebraic Derivation

Notice eq.(4) and eq.(2)again. They are all cubic in x, y, z. For a moment let us treat the unknown z as a parameter (i.e., a *hidden variable*), and collect an **coefficients matrix** $\mathtt{C}$ with respect to the other **two** variables x, y. The monomials we obtain span a vector space of:

$$\mathbf{X} = [x^3, y^3, x^2y, xy^2, x^2, y^2, xy, x, y, 1]^{\mathbf{T}}. \tag{6}$$

Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06)
0-7695-2521-0/06 $20.00 © 2006 IEEE

Note its dimensionality is only ten.

Combining these 9 equations with the singularity condition (eq.2), we now have in total 10 equations in the above ten-dimensional monomial vector. Then we have a generalized linear matrix equation:

$$\mathtt{C}(z)\mathbf{X}(x,y)=0.$$

Note that we have explicitly included the dependent variable z in the coefficient matrix and the monomial vector.

Recall that this matric equation will have non-trivial solutions *if and only if* the **determinant** of the coefficient matrix vanishes. That is:

$$\det(\mathtt{C}(z))=0. \tag{7}$$

This determinant is better known as a *hidden-variable resultant*, which is an univariate polynomial of the hidden variable.

Inspecting the hidden-variable-resultant closely, we find that terms whose degree is greater than 10 have been precisely cancelled-out, and a precisely 10-th degree polynomial is left. As a result, at most ten solutions to the five-point problem can be obtained. This accords precisely with previous result, but we achieve this via a more straightforward way.

There are many techniques for solving a univariate resultant equation, for example, the companion matrix method, or Sturm's bracketing method. In our experiments we use the first one, for it is easier to implement than the second.

6. Solve Other Unknowns using SVD

Once z is computed by solving the resultant equation, the other two unknowns are usually solved by back-substituting the solved z. Notice that this back-substitution needs to be done multiple times, as in general we will have multiple roots of z. This approach has been adopted by previous work. This paper suggests an easier way of solving other unknowns.

We notice that the monomial vector of eq.(6) has exhausted all the bi-variate combinations of variables x and y up to order three. Therefore, the solutions of x and y simply as the right null-space of $\mathtt{C}(z)$. This null-space can be effectively computed through a numerical SVD decomposition. In comparison, the back-substitution approach in [2][1] is more involved.

Once all x, y, z have been solved, one can find the essential matrix $\mathtt{E}$. And the motion parameters can be extracted easily using method reported elsewhere [3].

6.1 Algorithm Outlines

1. Write down the five epipolar equations of the five points.
2. Compute the null-space of the essential matrix.
3. Compute the symbolic determinant of the coefficient matrix $\mathtt{C}(z)$. Solving the determinant equation, one then find the solution z.
4. Back-substitute the real roots of z. Compute the null-space of $\mathtt{C}(z)$ and extract x, y from the null-space.
5. Recover the essential matrix, and extract the motion vectors.

7. Experimental Validation

7.1 Synthetic Data

We generate synthetic image and essential matrix using Torr's Matlab SFM Toolbox [16]. To resemble the real situation, the synthetic image size is set to be 512×512. Camera motions between two views are randomly drawn from a uniform distribution. No special attention has been paid to avoid the degenerate motion. The 3D points are located in general positions (though can be on a single plane).

In section 5.2, we have shown theoretically the determinant of $\mathtt{C}(z)$ is a 10-th degree polynomial in z. Now Our first experiment is used to validate this result.

From five corresponding points, after applying the proposed five-point algorithm, we obtain for example the following 10-th degree determinant equation:

$$\begin{aligned}&\det(C(z))=-.2996e^{-5}z^{10}-.3233e^{-4}z^{9}\\&+.9819e^{-3}z^{8}-.1547e^{-2}z^{7}+.4625e^{-3}z^{6}\\&+.1496e^{-4}z^{5}+.9100e^{-4}z^{4}-.1060e^{-2}z^{3}\\&\quad+.3477e^{-3}z^{2}+.8115e^{-3}z-.3587e^{-5}\end{aligned}$$

Solving this equation using the companion matrix method, we then obtain ten complex roots.

To measure the estimation precision, we adopt the formula of [1], which is

$$\epsilon_{\mathtt{E}}=\min_{i}\min\left(\left\|\frac{\hat{\mathtt{E}}_i}{\|\hat{\mathtt{E}}_i\|}-\frac{\mathtt{E}_i}{\|\mathtt{E}_i\|}\right\|,\left\|\frac{\hat{\mathtt{E}}_i}{\|\hat{\mathtt{E}}_i\|}+\frac{\mathtt{E}_i}{\|\mathtt{E}_i\|}\right\|\right) \tag{8}$$

Fig-1 display the $\log_{10}$ error (using eq.(8)) distribution results. The top row figures are results by the 8-pt linear algorithm (with Hartley's normalization and rank(2)modification) for noise-free case (in the left) and for $noise=1.0$ pixels case (in the right). The bottom row is our results. Both are the averages of 100 independent tests. We also compare the estimation errors under different noise conditions by linear algorithm and our algorithm, as shown in fig-2-Left. The results are the average of 100 times independent experiments. It is clear in both experiments that our algorithm is more accurate than the linear algorithm. Finally, we verify that the five-point algorithm works well for planar scene. we synthesize a single plane and test our algorithm again. A result is shown in fig-2-Right.We further

Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06)
0-7695-2521-0/06 $20.00 © 2006 IEEE

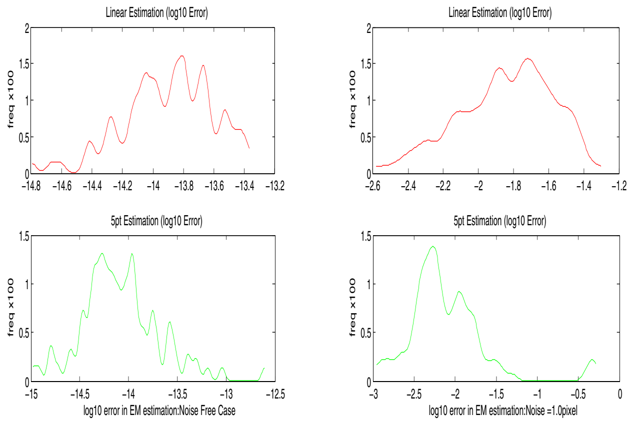

Figure 1. log10 error distribution (Left: noise free case; Right: noise=1.0 pixels; Top row: result by 8-pt algorithm; Bottom: by the proposed 5-pt algorithm).

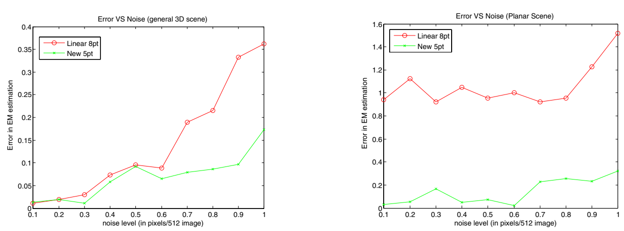

Figure 2. estimation error vs. noise level. (Left: a general 3D scene; Right: a planar scene; Red curve: by 8-pt algorithm; Green curve: by the proposed 5pt algorithm)

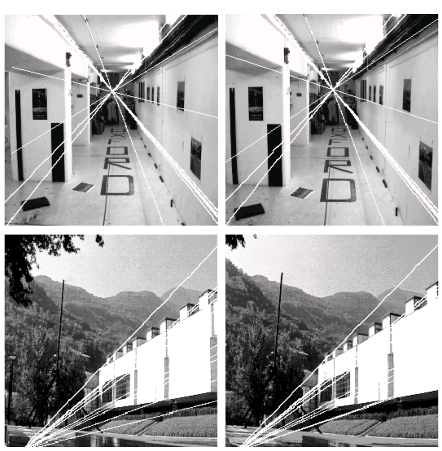

Figure 3. Some standard test images with estimated epipolar lines (by using the 5-pt algorithm) superimposed on.

compare the numerical performance of our new algorithm with Nistér's original algorithm, but have not found significant difference.

7.2 Tests on Real Images

We test our algorithm on some standard real images with known calibration information (see fig-3, courtesy of Oxford VGG and INRIA). Good results are obtained by our algorithm. Fig-3 shows the estimated epipolar lines superimposing on the images. In these experiments known camera calibration information is assumed. To use the 5pt algorithm more effectively, it is highly recommended to combine it with the RANSAC scheme [1].

8. Conclusion

We have proposed a new algorithm for solving the five-point motion estimation problem. This algorithm follows the simple idea of hidden-variable resultant, and is easy to implement. We believe that the simplicity actually suggests a deeper understanding to the problem. We hope this new algorithm, combining with the RANSAC scheme, will be a useful tool in vision motion estimation.

References

[1] D.Nistér, An efficient solution to the five-point relative pose problem, in Proc. IEEE-CVPR-2003, 2003.
[2] D.Nistér, An efficient solution to the five-point relative pose problem, IEEE-T-PAMI, 26(6), 2004.
[3] R.Hartley, A.Zisserman,Mutiview Geometry in computer vision, 2nd Edition, Cambridge University Press, 2004.
[4] R.Hartley, Estimation of Relative Camera Positions for Uncalibrated Cameras, In Proc. 2nd ECCV, 1992.
[5] Hongdong Li, A simpe solution to the six-point focal-length problem, In proc. ECCV-2006, May,2006.
[6] H.Stewénius, et al, A minimal solution for relative pose with unknown focal length, In Proc.CVPR-2005, 2005.
[7] M.Demazure, Sur Deux Problemes de Reconstruction, Technical Report No 882, INRIA, France, 1988.
[8] Faugeras, S. Maybank, Motion from Point Matches: Multiplicity of Solutions, IJCV, 4(3):225-246, 1990.
[9] A. Heyden, G. Sparr, Reconstruction from Calibrated Cameras - a New Proof of the Kruppa-Demazure Theorem, JMIV,vol-10:1-20, 1999.
[10] B. Triggs, Routines for Relative Pose of Two Calibrated Cameras from 5 Points, INRIA-Technical Report,INRIA, France, 2000.
[11] J. Philip, A Non-Iterative Algorithm for Determining all Essential Matrices Corresponding to Five Point Pairs, Photogrammetric Record, 15(88):589-599, 1996.
[12] David Nister and Henrik Stewenius, Using Algebraic Geometry for Solving Polynomial Problems in Computer Vision, short course at ICCV-2005, Beijing, 2005.
[13] D.Cox, J.Little and D.O'shea, *Using Algebraic Geometry*, 2nd Edition, Springer, 2005.
[14] L. Quan, Z.D.Lan, Linear N-Point Camera Pose Determination,IEEE PAMI-21(8): 774-780, 1999.
[15] Yi Ma, S. Soatto, J. Kosecka, and S. Sastry,An Invitation to 3-D Vision: From Images to Geometric Models, Springer-Verlag,2003.
[16] P.H.S. Torr, and D.W. Murray, The Development and Comparison of Robust Methods for Estimating the Fundamental Matrix,IJCV, pp.271–300, v-24,1997.

Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06)
0-7695-2521-0/06 $20.00 © 2006 IEEE