Approach of Camera Relative Pose Estimation Based on Epipolar Geometry

Chao Li and Yue Zhao
Institute of Mathematics and Statistics, Yunnan University, 2 Cuifu Northern Road, 650091, Kunming, Republic of China

Abstract: Estimation of camera relative positions is very important in computer vision. Assuming that the camera has been calibrated, this study presents two algorithms of camera relative pose estimation which one is to recover the camera motion parameters by essential matrix and the other by epipolar geometry. Algorithm 1 is commonly used in three-dimensional (3D) reconstruction. In algorithm 2, firstly linearize coplanar condition equation through Taylor's expansion and solve elements of relative orientation by iteration procedure, finally recover camera motion parameters. Compared with the above two algorithms, algorithm 1 requires at least eight pairs of matching points and consider degenerate condition, algorithm 2 simply requires five pairs of matching points. Experimental data reveals that the two algorithms are feasible and the results of them are close which have the higher robustness and can satisfy the requirement of 3D reconstruction.

Key words: Computer vision, relative orientation, essential matrix, photogrammetry

INTRODUCTION

The development of modern digital photogrammetry is closely linked with the research of computer vision (Pu et al., 2011). The research aim of computer vision is for the computers to acquire the ability through two-dimensional image to understand geometry information of object in the 3D setting, including its shape, position, posture, motion, etc. (Kaavaase et al., 2011; Murthy and Jadon, 2011). Simply put, computer vision is to utilize the computer to simulate the human eye so as to identify and understand the scene object. From this point of view, there are so many similarities between computer vision and digital photogrammetry. And of course, the basic categories of photogrammetry are to determine the geometric and physical properties of the measured object (Hussain et al., 2007).

Using the image correspondingly to estimate the camera’s position, posture and spatial scene structure is one of the main tasks in computer vision and photogrammetry (Yakar et al., 2009). The relative position estimation between two or more cameras is called the extrinsic parameters assumption in computer vision which is also called the relative orientation in photogrammetry. The algebraic description of motion parameters among the sequence images was first suggested by Longuet-Higgins (1981) which could be described by the essential matrix, E=[T], R, [T], is symmetric matrix of translation matrix T, T= denotes a difference of one scale factor and rotation matrix R can be determined by the decomposition of E but the results are not the only.


Spatial straightnesses S,M, S,M, S,S, are coplanar and coplanarity constraint as shown in Fig. 1. Obviously, the coplanarity condition is also true in binocular stereo vision model, as shown in Fig. 2, so as to establish coplanarity constraint equation to get relative orientation elements. By a review of the relationship between relative orientation elements and camera motion parameters, camera motion parameters can be recovered.

In Fig. 3, when air survey camera photographs the ground, the angle α is the primary optical axis of photographic lens deviation from the plumb line SN which is called aerial photograph angle of slope. The current aerial photography technology keeps α within 3 degrees. In Fig. 4, air survey camera must be
adjacent spacing camera to a minimum which usually
can not keep in computer vision.
This study mainly considers difference and
connection between computer vision and
photogrammetry and discusses typical methods of
recovering camera motion parameters and make theoretical
analysis and comparative experiments.

RETRIEVING CAMERA MOTION PARAMETERS
FROM ESSENTIAL MATRIX

Essential matrix: As shown Fig. 1, left and right
camera coordinate systems are \(O-XYZ\), \(O'-X'Y'Z'\),
respectively. Let \((R, T)\) be motion parameters of
the second camera relative to the first one \((R \text{ and } T\) are rotation matrix and translation vector).
Assuming camera intrinsic parameters matrix \(K\), \(K'\) is
known, two images can be normalized transformation
\(m_n = K^{-1}m\), \(m'_n = K'^{-1}m\) and two new images \(\{l_n, l'_n\}\) will
be obtained which are called the normalized image of
the original image. Fundamental matrix between original
images is \(F = K^{-1}K'^{-T}\). RK'\(^{-1}\), so epipolar constraint
equation between \(l_n\) and \(l'_n\) must be \(m'^T_n [l]\). \(Rm'_n - 0\). The
equation with essential matrix \(E\) can be determined as
follows:

\[
E = [l], \quad R = K'^T FK
\]  

(1)

REtrieving camera motion parameters from essential
matrix: Assuming camera calibration matrix \(K, K'\)
is known, from Eq. 1, the fundamental matrix \(F\) is
gotten and then essential matrix \(E\) can be obtained
(Zhao and Lv, 2012). The fundamental matrix \(F\)
requires at least 8 pairs of matching points which
is obtained by RANSAC algorithm (Zhang et al.,
1995; Wang et al., 2009; Guo et al., 2011).
Matching points can be obtained by Tomasi and
Kanade (1991), Harris and Stephens (1988) and
Zhang et al. (1995). Using the essential matrix, we
can further retrieve the camera relative motion
parameters. Let:
$W = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

Through decomposing $E$ by SVD, one has $E = \text{Udial}(1 \ 1 \ 0)V^T$, where $\ast$ stands for equality up to scale, translation vector is $t = [u_1, u_2, u_3]^T$, rotation matrix is $R = \text{UWV}$ or $R = \text{UW}^TV^T$.

Assuming that the first camera coordinate system is world coordinate system and then projection matrix is $P_1 = K[I \ 0]$ in the first camera, projection matrix has four possible solutions in the second camera $P_2 = K[R_{21}, t_1]$, $P_2 = K[R_{21}, -t_1]$, $P_2 = K[R_{21}, t_2]$, $P_2 = K[R_{21}, -t_2]$.

From the geometric analysis of solutions in cameras, it is found that only one set of solutions are reasonable which can be determined by reconstructing of a pair of corresponding points. When $Z$ coordinates of the reconstructed points is negative value in the two camera coordinate systems, it is the solution we need.

**RELATIVE ORIENTATION DETERMINE THE MOTION PARAMETER OF CAMERA**

In Fig. 2, $S_l$, $S_r$ are the optical centers of two cameras, respectively, $m_l$, $m_r$ denote the image of space-point $M$ under the left and right camera and $S_l m_l$, $S_r m_r$ denote the ray through the optical centers of two cameras, respectively. They are coplanar with the baseline $S_l S_r$ of cameras and this plane can be denoted by mixed product of three vectors $R_l, R_r$ and $B_r$.

$$B_r \cdot (R_l \times R_r) = 0$$

In the above formula, switching coordinates on the type, its third-order determinant is equal to zero:

$$F = \begin{vmatrix} B_x & B_y & B_z \\ -X_l & Y_l & Z_l \\ X_r & Y_r & Z_r \end{vmatrix} = 0 \quad (2)$$

where, $[X_l, Y_l, Z_l]$, $[X_r, Y_r, Z_r]$ denote the coordinates of $m_l$, $m_r$ under the two camera coordinate systems, respectively. Equation 2 is called coplanar condition equations.

Retrieving motion parameters: As a camera coordinate system, one may usually take one of the world coordinate system in the binocular stereo vision. Assuming that the first camera coordinate system is world coordinate system, $X_l$, $Y_l$, $Z_l$ is known and $B_x = B \cdot \cos \phi \cdot \cos \mu$, $B_y = B \cdot \cos \phi \cdot \sin \mu$, $B_z = B \cdot \sin \phi$ where $B$ is set to 1. Beyond this, assume that $\phi$, $\omega$, $\kappa$ denote the diagonal elements of the two cameras.

**Proposition 1:** Coplanar condition equation can be linearized by Taylor expansion.

**Proof:** Since Eq. 2 is nonlinear function, the first-order term can be approximately expanded by Taylor formula about multiple-variable functions, we have:

$$F_l = E_l = \frac{\partial F}{\partial \phi} \cdot \delta \phi + \frac{\partial F}{\partial \theta} \cdot \delta \theta + \frac{\partial F}{\partial \kappa} \cdot \delta \kappa + \frac{\partial F}{\partial B_x} \cdot \delta B_x + \frac{\partial F}{\partial B_y} \cdot \delta B_y + \frac{\partial F}{\partial B_z} \cdot \delta B_z = 0 \quad (3)$$

If we want to solve the partial derivative:

$$\frac{\partial F}{\partial \phi} \frac{\partial F}{\partial \theta} \frac{\partial F}{\partial \kappa}$$

in Eq. 3, we must first solve the partial derivative:

$$\frac{\partial X_l}{\partial \phi} \frac{\partial X_r}{\partial \theta} \frac{\partial X_r}{\partial \kappa}$$

When $\phi$, $\omega$, $\kappa$ are small-angle, coordinate transformation type can quote small rotation matrix:

$$\begin{bmatrix} X_l \\ Y_l \\ Z_l \end{bmatrix} = \begin{bmatrix} 1 & -\kappa & -\phi \\ \kappa & 1 & -\eta \\ \phi & \eta & 1 \end{bmatrix} \begin{bmatrix} x_l \\ y_l \\ z_l \end{bmatrix}$$

The derivative of $\phi$, $\omega$, $\kappa$ can be obtained, respectively:

$$\begin{bmatrix} \delta x_l \\ \delta y_l \\ \delta z_l \end{bmatrix} = \begin{bmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_l \\ y_l \\ z_l \end{bmatrix}$$

Using the above equations solve the five partial derivatives of Eq. 3 and substitute into Eq. 3, we obtain:

$$\begin{bmatrix} B_x \\ B_y \\ B_z \end{bmatrix} + \begin{bmatrix} \frac{\partial F}{\partial \phi} \\ \frac{\partial F}{\partial \theta} \\ \frac{\partial F}{\partial \kappa} \end{bmatrix} \begin{bmatrix} x_l \\ y_l \\ z_l \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad (4)$$
Proposition 2: Let:

\[ q = \frac{EN}{B_2Z_1} \]

can establish system of linear equations about the exterior orientation, in which:

\[ N = \frac{B_2Z_1 - B_2X_1}{X_1Z_2 - Z_1X_2} \]

is the projection coefficient to translate image point \( m_1 \) into point \( M \) in the space.

**Proof:** On the one hand:

\[ \begin{bmatrix} B_x & B_y & B_z \\ X_1 & Y_1 & Z_1 \end{bmatrix} q = \begin{bmatrix} [B_2Z_1 - B_2X_1] / X_1Z_2 - Z_1X_2 \\ [B_2Z_1 - B_2X_1] / X_1Z_2 - Z_1X_2 \end{bmatrix} \]

where, \( N \) is the projection coefficient of image point \( m_1 \):

\[ N = \frac{B_2Z_1 - B_2X_1}{X_1Z_2 - Z_1X_2} \]

On the other hand, both sides of Eq. 4 are divided by \( B_x \), respectively and neglect quadratic or more terms. Rearranging the result, we have:

\[ Y_1, d\phi + (Y_1 - Z_1) d\omega - x_2 Z_2 d\kappa + (Z_2 X_2 - X_2 Z_2) d\mu + (X_2 Y_2 - X_2 Y_2) d\nu = 0 \]

(5)

Only considering the simple term, \( x_2, y_2 \) in Eq. 5 can be replaced by \( X_2, Y_2 \) and approximately regard as:

\[ \begin{bmatrix} Y_1 - Y_2 \\ Z_2 - Z_1 \end{bmatrix} = \begin{bmatrix} B_x \end{bmatrix} N \]

(6)

Substitute Eq. 6 into 5, \( \frac{N'}{Z_2} \) times all terms, we get:

\[ q = -\frac{x_2 N' d\phi - (Z_2 + \frac{Y_2}{Z_2}) N' d\omega + X_2 N' d\kappa + B_2 d\mu - \frac{Y_2}{Z_2} B_2 d\nu}{Z_2} \]

(7)

Proposition 3: Determining the relative orientation elements \( \mu, \nu, \phi, \omega, \kappa \) is equivalent to determining the camera motion parameters.

**Proof:** Equation 6 is calculating formula for analytical method of continuous relative orientation. Through the solution of Eq. 6, relative orientation elements \( \mu, \nu, \phi, \omega, \kappa \) can be set which determine the rotation matrix \( R \) and translation vector \( T \), with:

\[ T = [B_x B_y B_z] \]

(8)

\[ R = \begin{bmatrix} \cos \phi \cos \kappa - \sin \phi \sin \omega \cos \kappa - \sin \phi \cos \omega & \cos \phi \sin \kappa - \sin \phi \sin \omega \cos \kappa & -\sin \phi \cos \omega - \sin \phi \sin \omega \cos \kappa \\ -\sin \phi \cos \kappa - \cos \phi \sin \omega \cos \kappa & \cos \phi \sin \kappa - \sin \phi \sin \omega \cos \kappa & \cos \phi \cos \omega - \sin \phi \sin \omega \cos \kappa \\ -\sin \kappa \cos \omega + \cos \phi \sin \omega \cos \kappa & -\sin \kappa \sin \omega + \cos \phi \sin \omega \cos \kappa & \sin \kappa \cos \omega + \cos \phi \sin \omega \cos \kappa \end{bmatrix} \]

(9)

Outline of the algorithm: In the Eq. 7, \( d\phi, d\omega, d\kappa, d\mu, d\nu \) are five unknown parameters and each group matching points can solve an equation, so at least five pairs of matching points solve the equation \( Ax = L \) which is composed of Eq. 7. Since result on Eq. 7 is condition Eq. 2 after linearization, solving the relative orientation elements is a step iterative process. In practice, when all correction value is less than limit \( 0.3\times 10^{-4} \), the iteration ends. The steps of the algorithm are as follows:

**Step 1:** Let relative orientation elements \( \mu = \nu = \phi = \omega = \kappa \)

**Step 2:** Calculate \( R, T \) according to Eq. 8 and 9

**Step 3:** Calculate projection coefficient \( N, N' \) and solve \( q \) by Eq. 7

**Step 4:** Establish equation \( AX = L \)

**Step 5:** Solve above equation to get its iterative solution

**Step 6:** See if all unknown number and correction are less than threshold, output \( R, T \), else return Eq. 2

**EXPERIMENTS RESULTS**

Simulation result: In Fig. 5, a cube is shown in simulation experiment. And take a corner and three mutually perpendicular edges as the origin and coordinate axes in the world coordinate system, respectively. The edge length of the cube is 3000 from Fig. 5. Figure 6a and b are the images of cubic, in which the cameras extrinsic parameters are:
which each surface is composed of the 9 small squares. Without considering the repeat points of adjacent small squares, there are 9×4 corners in a surface altogether, so the total number of corner points is 108. By these corners, the experiment results are shown as Fig. 7 (Rij, i, j = 1, 2, 3 denotes the ith row and the jth column of the rotation matrix and Tij, i = 1, 2, 3 denotes the ith element of translation vector).

From the experimental results, it is obvious that the precision of Algorithm 1 is a little higher than Algorithm 2 (Fig. 7). However, Algorithm 1 requires eight pairs of matching points at least and considers the deterioration (multiple points are collinear or coplanar). Algorithm 2 only needs five pairs of matching points. The choice of the first 30 data points is inconsistent in the Algorithm 1 and 2. Algorithm 1 selects the non-coplanar data points (considering the deterioration), algorithm 2 chooses coplanar data points. When the number of data points is more than 30, two methods also choose the same data points for the experiment. From the above results, the result of algorithm 2 is edging closer to the precise value with the increasing data points. Especially there are the non-coplanar points in the data points (that is, the data points are more than 30), the precisions of result improve significantly. With regard to algorithm 2, although the data points are all coplanar points, the relative error is also small. In a word, algorithm 1 has slightly higher precision and the applicability of algorithm 2 is a little better.

In addition, in order to test and compare the robustness of two algorithms, add the Gaussian noise (the mean value is zero, the mean square error is from 0.3 pixel to 3 pixel) on the position of pixels. 50 independent experiments can be executed for each noise level and solve the averages. The following Fig. 8 shows the changing situation of the motion parameters R, T along with the noise.

**Experiment with real images**: The factual data can be used for comparing and testing two algorithms which are proposed in this paper. The kind of camera is CCD digital camera. The size of image is 320×240. Figure 9 and 10 are two photos from different visual angles. Firstly, (Harris and Stephens, 1988) corner detection is executed (Elatta et al., 2004; Zhang et al., 1995), then 20 pairs of matching points can be get though matching, where the red circles indicate the coordinates of the selective image points, the blue letters is the serial number of matching position.

First of all, the camera intrinsic parameters are obtained by Zhang (2000) calibration algorithm as follows:
Fig. 7: Accuracy of the proposed methods concerning the number of points, Rij: Element in the ith row and jth column of R, Ti: Element in the ith row of T

\[
K = \begin{bmatrix}
387.755726 & 0 & 167.452919 \\
0 & 383.109923 & 181.310536 \\
0 & 0 & 1
\end{bmatrix}
\]

Then, the calculation of parameters is made by Algorithm 1, 2 and the above-mentioned 20 pairs of matching points. The results are as follows:

\[
R_1 = \begin{bmatrix}
0.092396 & 0.016082 & 0.361531 \\
0.004415 & 0.999230 & -0.039249 \\
-0.360644 & 0.003188 & 0.921534
\end{bmatrix},
T_1 = [-0.872123 0.137812 0.469177],
R_2 = [-0.206239 0.999728 -0.011614],
T_2 = [-0.360123 0.141825 0.489974]
\]

R_1, T_1 and R_2, T_2 show the relatively motion parameters from Algorithm 1 and 2. It is evident that the results of two proposed algorithms are very close.

1207
Fig. 8: Accuracy of the proposed methods concerning the noise level, $R_{ij}$: Element in the $i$th row and $j$th column of $R$, $T_i$: Element in the $i$th row of $T$.

Fig. 9(a-b): Photos from different visual angles.
CONCLUSIONS

Two algorithms of recovering camera motion parameters are proposed on the assumption that the camera has been calibrated. Algorithm 1 recovers the camera motion parameters from the essential matrix. Algorithm 2 recovers camera motion parameters by coplanarity equation. Algorithm 1 and 2 are commonly used in computer vision and photogrammetry, respectively. Comparatively speaking, the precision of Algorithm 1 is a little higher than Algorithm 2. However, Algorithm 1 requires eight pairs of matching points at least and considers the deterioration. Algorithm 2 only needs five pairs of matching points for recovering camera motion parameters. More importantly, Algorithm 2 doesn’t have the degenerate case, so Algorithm 2 is more suitable. In this study, the derivation of two algorithms is revealed in detail. Finally, simulation and real experiments results show that the two algorithms are feasible and very close. Meanwhile, two algorithms also have the higher robustness and can satisfy the requirements of 3D reconstruction.

ACKNOWLEDGMENTS

This study is supported by the Natural Foundation of Yunnan Province, China (2011FB017), the Scientific Research Foundation of Yunnan Education Department of China (2010Y245) and the Scientific Research Foundation of Yunnan University (2010YB021).

REFERENCES


