# Two-step pose estimation method based on five reference points

**Zimiao Zhang (张子森)*, Changku Sun (孙长库), and Peng Wang (王 鹏)**

*State Key Laboratory of Precision Measuring Technology and Instruments, Tianjin University, Tianjin 300072, China*

**Corresponding author: zzm19850126@yahoo.com.cn*

Received December 19, 2011; accepted January 10, 2012; posted online March 28, 2012

A two-step method for pose estimation based on five co-planar reference points is studied. In the first step, the pose of the object is estimated by a simple analytical solving process. The pixel coordinates of reference points on the image plane are extracted through image processing. Then, using affine invariants of the reference points with certain distances between each other, the coordinates of reference points in the camera coordinate system are solved. In the second step, the results obtained in the first step are used as initial values of an iterative solving process for gathering the exact solution. In such a solution, an unconstrained nonlinear optimization objective function is established through the objective functions produced by the depth estimation and the co-planarity of the five reference points to ensure the accuracy and convergence rate of the non-linear algorithm. The Levenberg-Marquardt optimization method is utilized to refine the initial values. The coordinates of the reference points in the camera coordinate system are obtained and transformed into the pose of the object. Experimental results show that the RMS of the azimuth angle reaches 0.076° in the measurement range of 0°–90°; the root mean square (RMS) of the pitch angle reaches 0.035° in the measurement range of 0°–60°; and the RMS of the roll angle reaches 0.036° in the measurement range of 0°–60°.

*OCIS codes:* 150.0155, 140.1135, 330.4060.
*doi: 10.3788/COL201210.071501.*

Pose estimation, which can be widely applied in the fields of robot navigation[1], light pen measurement[2], electro-optic aiming[3,4], vehicle quality inspection[5], and aerospace science[6−8], among others, has an important value. Research in this area has been more active in recent years. The existing approaches to generate solutions fall in two distinct categories: analytical solutions and iterative solutions. Earlier studies have proved that four reference points can achieve a linear solution, but four points may yield more than one solution[9−11]. Tang *et al.* presented a linear algorithm with under condition of five feature points[12]. DeMenthon *et al.* proposed a POSIT algorithm, which obtains the initial value of the solution using the scaling and orthography projection model to approximate the perspective projection model[13]. Zhang *et al.* estimated an object pose based on geometrical constraints[14]. Chen *et al.* solved the rotation and translation matrix using the least squares method according to orthonormal constraints; the method used has many points (10×10)[13]. Wang *et al.* achieved the object pose non-linearly on the basis of five reference points using the least squares approach. Wildey *et al.* proposed a positioning method based on three points, but their approach had a smaller measurement range (0°–5°).

Both analytical and iterative solutions have their scopes of application and disadvantages. Compared with iterative solutions, analytical solutions generally take less computation time for the same accuracy. On the other hand, analytical solutions are extremely susceptible to noise. Non-linear optimization problems are normally solved with variation on gradient descent or Gauss–Newton methods. Iterative solutions have drawbacks, namely, they need ideal initial values of the true solution, and they are time-consuming.

Given these considerations, a two-step method for pose estimation based on five co-planar reference points is studied in this letter. In the first step, the pose of the object is estimated by a simple analytical solving process. The pixel coordinates of the reference points on the image plane are extracted through image processing. Then, by utilizing affine invariants of the reference points with certain distances between each other, the coordinates of the reference points in the camera coordinate system are solved. In the second step, the results obtained in the first step are used as the initial values of an iterative process for gathering the exact solution. In such a solution, an unconstrained nonlinear optimization objective function is established through the objective functions produced by the depth estimation and the co-planarity of the five reference points to ensure the accuracy and convergence rate of the non-linear algorithm. Then, the Levenberg–Marquardt optimization method is utilized to refine the initial values. The coordinates of the reference points in the camera coordinate system are obtained and transformed into the pose of the object. The advantages of the two-step method are as follows. It improves the linear method in the situation wherein the method is extremely susceptible to noise; and it yields a preferable initial value to the iterative process in the second step. The proposed algorithm processes each image frame separately unlike in Ref. [18], thereby eliminating the accumulation of calculation errors.

A target pattern with five reference points is designed for pose estimation, as shown in Fig. 1. The No. 0, No. 1, and No. 2 points are on the same straight line. The No. 3 and No. 4 points can be distinguished according to the distance between the point and the straight line. The No. 2 point is farthest from the straight line composed of No. 3 and No. 4 points. The No. 0 and No. 1 points could be distinguished according to their distances from

 

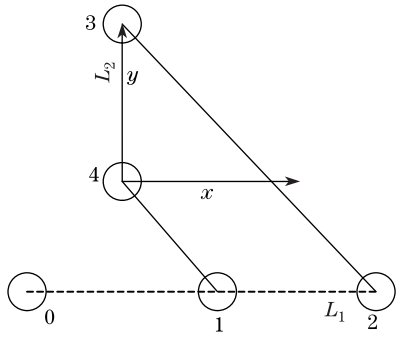


Fig. 1. Measurement target with five feature points.

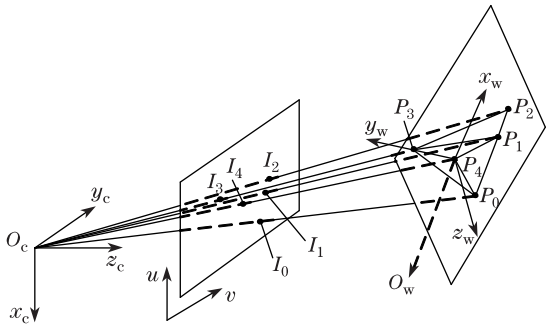


Fig. 2. Perspective projection model with five points.

No. 2. The No. 4 point is set as the origin of the coordinate system, and the connection line between No. 4 and No. 3 points is in the $y$-axis direction of the coordinate system. The direction parallel to the connection line between No. 2 and No. 0 points, starting from the No. 4 point, is the $x$-axis direction. The $z$-axis is determined in accordance with the principle of the right hand. In this setup, the world coordinate system (target coordinate system) is constructed.

To achieve the solution to the object pose, the coordinates of reference points in the camera coordinate system must be solved firstly. The perspective projection model with five reference points is shown in Fig. 2. $o_{\rm c}-x_{\rm c}y_{\rm c}z_{\rm c}$ is the camera coordinate system, which denotes the camera frame. $uv$ is the CCD image plane, with the original point at the image center of the CCD plane.

The intrinsic parameters of the camera, which include the focus of the lens $f$, the radial distortion coefficients of camera lens $k_1$ and $k_2$, the tangential distortions $p_1$ and $p_2$, the center pixels of the computer image $u_0$ and $v_0$, and the uncertainty image factor $s_x$, could be obtained through camera calibration according to the calibration method presented by Zhang *et al.* $d_x$ and $d_y$ are center-to-center distances between pixels in the row and column directions, respectively. On account of the radial distortion and tangential distortion of the lens, the undistorted coordinate point on the image plane could be obtained by the transformation formula of the ideal and actual image coordinates.

The coordinates of reference points in the camera coordinate system are represented by $\mathbf{P}_{ci}=(x_{ci},y_{ci},z_{ci})$, and those in the world coordinate system follow $P_{{\rm w}i}=(x_{{\rm w}i},y_{{\rm w}i},z_{{\rm w}i})$. The corresponding ideal image coordinates are as follows: $\mathbf{I}_{ui}$=$(x_{ui},y_{ui},\ 1)^{\rm T}(i$=$0\cdots4)$. The relationship of $P_{ci}$ and $I_{ui}$ could be described as

$$\begin{cases} z_{ci}\mathbf{I}_{ui}=\begin{bmatrix} s_xf/d_x & 0 & u_0\\ 0 & f/d_y & v_0\\ 0 & 0 & 1\end{bmatrix}\\ \mathbf{P}_{ci}=\begin{bmatrix} f_x & 0 & u_0\\ 0 & f_y & v_0\\ 0 & 0 & 1\end{bmatrix}\mathbf{P}_{ci}=\mathbf{K}\mathbf{P}_{ci}\\ \mathbf{P}_{ci}=\lambda_i\mathbf{K}^{-1}\mathbf{I}_{ui}(\lambda_i=z_{ci}) \end{cases}. \tag{1}$$

The coordinates of the reference points in the camera coordinate system could be obtained by solving for $\lambda_i$.

$\overrightarrow{P_1P_4}$ is parallel to $\overrightarrow{P_2P_3}$. Thus, two affine invariants are introduced as

$$\begin{cases} k=\parallel\mathbf{P}_{\rm w4}-\mathbf{P}_{\rm w1}\parallel/\parallel\mathbf{P}_{\rm w3}-\mathbf{P}_{\rm w2}\parallel\\ \mathbf{P}_{\rm w4}=k(\mathbf{P}_{\rm w3}-\mathbf{P}_{\rm w2})+\mathbf{P}_{\rm w1}\\ \quad=[\mathbf{P}_{\rm w1},\ \mathbf{P}_{\rm w2},\ \mathbf{P}_{\rm w3}]\,[1,\ -k,\ k]^{\rm T} \end{cases}. \tag{2}$$

The transformation from the world coordinate system to the camera coordinate system is an affine transformation; thus, we have

$$\mathbf{P}_{\rm c4}=k(\mathbf{P}_{\rm c3}-\mathbf{P}_{\rm c2})+\mathbf{P}_{\rm c1}=[\mathbf{P}_{\rm c1},\mathbf{P}_{\rm c2},\mathbf{P}_{\rm c3}]\,[1,-k,k]^{\rm T}. \tag{3}$$

Given $\mathbf{P}_{\rm c}=[\mathbf{P}_{\rm c1},\mathbf{P}_{\rm c2},\mathbf{P}_{\rm c3}]$, $\mathbf{P}_{\rm c}^{-1}\mathbf{P}_{4{\rm c}}$ could be represented by

$$\begin{cases} \mathbf{P}_{\rm c}=[\mathbf{P}_{\rm c1},\mathbf{P}_{\rm c2},\mathbf{P}_{\rm c3}]\\ \quad=\mathbf{K}^{-1}{\rm diag}(\lambda_1,\lambda_2,\lambda_3)[\mathbf{I}_{\rm u1},\mathbf{I}_{\rm u2},\mathbf{I}_{\rm u3}]\\ \mathbf{P}_{\rm c}^{-1}\mathbf{P}_{4{\rm c}}={\rm diag}(1/\lambda_1,1/\lambda_2,1/\lambda_3)[\mathbf{I}_{\rm u1},\mathbf{I}_{\rm u2},\mathbf{I}_{\rm u3}]^{-1}\\ \mathbf{K}(\lambda_4\mathbf{K}^{-1}\mathbf{I}_{\rm u4})={\rm diag}(\lambda_4/\lambda_1,\lambda_4/\lambda_2,\lambda_4/\lambda_3)\\ \qquad\qquad\cdot[\mathbf{I}_{\rm u1},\mathbf{I}_{\rm u2},\mathbf{I}_{\rm u3}]^{-1}\mathbf{I}_{\rm u4}. \end{cases}. \tag{4}$$

Given $[\eta_1,\eta_2,\eta_3]=[\mathbf{I}_{\rm u1},-\mathbf{I}_{\rm u2},\mathbf{I}_{\rm u3}]^{-1}\mathbf{I}_{\rm u4}$, Eq. (5) is a derivate.

$$\mathbf{P}_{\rm c}^{-1}\mathbf{P}_{\rm c4}={\rm diag}(\lambda_4/\lambda_1,\lambda_4/\lambda_2,\lambda_4/\lambda_3)[\eta_1,-\eta_2,\eta_3]. \tag{5}$$

According to Eqs. (3) and (5), the distance between each pair of reference points is known by referring to the world coordinate system constructed above, and the image coordinates of reference points are extracted through image processing. Thus, $\lambda_i$, which represents the depth information of every reference point, could be calculated as

$$\begin{cases} \lambda_1=\eta_1\lambda_4,\lambda_2=\eta_2\lambda_4/k,\lambda_3=\eta_3\lambda_4/k\\ \lambda_4=\eta_4\lambda_4(\eta_4=1)\\ k=\parallel\mathbf{P}_{\rm w4}-\mathbf{P}_{\rm w1}\parallel/\parallel\mathbf{P}_{\rm w3}-\mathbf{P}_{\rm w2}\parallel\\ \parallel\mathbf{P}_{\rm w3}-\mathbf{P}_{\rm w2}\parallel=\parallel\mathbf{P}_{\rm c3}-\mathbf{P}_{\rm c2}\parallel\\ \quad=\lambda_4\parallel\eta_3\mathbf{K}^{-1}\mathbf{I}_{\rm u3}/k-\eta_2\mathbf{K}^{-1}\mathbf{I}_{\rm u2}/k\parallel \end{cases}. \tag{6}$$

The vector from the optical center $O_{\rm c}$ to each reference point could be calculated through

$$\begin{cases} \overrightarrow{O_{\rm c}P_{ci}}=\eta_i\mathbf{K}^{-1}\mathbf{I}_{ui}\parallel\mathbf{P}_{\rm w4}-\mathbf{P}_{\rm w1}\parallel\\ /\parallel\eta_3\mathbf{K}^{-1}\mathbf{I}_{\rm u3}-\eta_2\mathbf{K}^{-1}\mathbf{I}_{\rm u2}\parallel(i=1,4)\\ \overrightarrow{O_{\rm c}P_{\rm ci}}=\eta_i\mathbf{K}^{-1}\mathbf{I}_{ui}\parallel\mathbf{P}_{\rm w3}-\mathbf{P}_{\rm w2}\parallel\\ /\parallel\eta_3\mathbf{K}^{-1}\mathbf{I}_{\rm u3}-\eta_2\mathbf{K}^{-1}\mathbf{I}_{\rm u2}\parallel(i=2,3) \end{cases}. \tag{7}$$

With the results of $\lambda_i(i$=1, 2, 3, 4), the equation of the target plane $\pi^{\rm T}\mathbf{X}=0(\pi=[\pi_1,\ \pi_2,\ \pi_3,\ \pi_4]^{\rm T})$ could be constructed.

On the basis of the results of $\lambda_i$ ($i$=1, 2, 3, 4), the rotation and translation matrix from the target coordinate

system to the camera coordinate system could be solved by

$$\begin{cases} \mathbf{R}_1 = \overrightarrow{O_cP_{c2}} - \overrightarrow{O_cP_{c1}} / \parallel \overrightarrow{O_cP_{c2}} - \overrightarrow{O_cP_{c1}} \parallel \\ \mathbf{R}_2 = \overrightarrow{O_cP_{c3}} - \overrightarrow{O_cP_{c4}} / \parallel \overrightarrow{O_cP_{c3}} - \overrightarrow{O_cP_{c4}} \parallel \\ \mathbf{R}_3 = [\pi_1/r, \pi_2/r, \pi_3/r], \quad r = \sqrt{\pi_1^2 + \pi_2^2 + \pi_3^2} \\ \mathbf{R} = \begin{bmatrix} R_{1x} & R_{1y} & R_{1z} & 0 \\ R_{2x} & R_{2y} & R_{2z} & 0 \\ R_{3x} & R_{3y} & R_{3z} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \\ \mathbf{T} = \begin{bmatrix} 1 & 0 & 0 & -P_{c4x} \\ 0 & 1 & 0 & -P_{c4y} \\ 0 & 0 & 1 & -P_{c4z} \\ 0 & 0 & 0 & 1 \end{bmatrix} \\ \mathbf{P}_{w0} = \lambda_0 \mathbf{K}^{-1} \mathbf{I}_{u0} = \mathbf{R} \cdot \mathbf{T} \cdot \mathbf{P}_{c0} \end{cases}, \quad (8)$$

and depth information $\lambda_0$ of reference point $P_0$ could be obtained through the rotation and translation matrix.

As can be seen in Fig. 2, the spatial geometric model of the figure enclosed by the connective lines of five reference points is known according to the world coordinate system constructed before. The geometric figure includes 10 triangles, and each triangle has three sides. According to the vector from the optical center to the reference point each side of the triangle could be expressed as

$$\begin{aligned} E_1(i,j) = {} & \lambda_i^2 \parallel \mathbf{K}^{-1}\overrightarrow{O_cI_{ui}} \parallel^2 + \lambda_j^2 \parallel \mathbf{K}^{-1}\overrightarrow{O_cI_{uj}} \parallel^2 \\ & -2\lambda_i\lambda_j(\mathbf{K}^{-1}\overrightarrow{O_cI_{ui}}) \cdot (\mathbf{K}^{-1}\overrightarrow{O_cI_{uj}}) - \parallel \overrightarrow{P_{wj}P_{wi}} \parallel^2 . \end{aligned} \quad (9)$$

A total of 10 distance constraints exist aggregately.

Each triangle has three angles, and every four points can make up three pairs of vectors. Each pair of vectors could also constitute an angle. These angles are denoted by

$$\begin{cases} E_2(i,\ j,\ l) = (\lambda_j \mathbf{K}^{-1}\overrightarrow{O_cI_{uj}} - \lambda_i \mathbf{K}^{-1}\overrightarrow{O_cI_{ui}}) \\ \quad \cdot(\lambda_l \mathbf{K}^{-1}\overrightarrow{O_cI_{ul}} - \lambda_j \mathbf{K}^{-1}\overrightarrow{O_cI_{uj}}) \\ \quad / \parallel \lambda_j \mathbf{K}^{-1}\overrightarrow{O_cI_{uj}} - \lambda_i \mathbf{K}^{-1}\overrightarrow{O_cI_{ui}} \parallel \\ \quad \cdot \parallel \lambda_l \mathbf{K}^{-1}\overrightarrow{O_cI_{ul}} - \lambda_j \mathbf{K}^{-1}\overrightarrow{O_cI_{uj}} \parallel - \cos\theta_1 \\ E_3(i,j,l,n) = (\lambda_j \mathbf{K}^{-1}\overrightarrow{O_cI_{uj}} - \lambda_i \mathbf{K}^{-1}\overrightarrow{O_cI_{ui}}) \\ \quad \cdot(\lambda_l \mathbf{K}^{-1}\overrightarrow{O_cI_{ul}} - \lambda_n \mathbf{K}^{-1}\overrightarrow{O_cI_{un}}) \\ \quad / \parallel \lambda_j \mathbf{K}^{-1}\overrightarrow{O_cI_{uj}} - \lambda_i \mathbf{K}^{-1}\overrightarrow{O_cI_{ui}} \parallel \\ \quad \cdot \parallel \lambda_l \mathbf{K}^{-1}\overrightarrow{O_cI_{ul}} - \lambda_n \mathbf{K}^{-1}\overrightarrow{O_cI_{un}} \parallel - \cos\theta_2 \end{cases} . \quad (10)$$

A total of 45 angle constraints exist aggregately.

However, considering the distance and angle errors is not enough. To further ensure the shape of the geometric figure (i.e., preserve the rigidity of the target), other constraints need to be considered, namely, when each triangle makes up a plane, and when the remaining two points and the triangle are on the same plane.

$$\begin{aligned} E_4(i,\ j,\ l,\ n) = {} & [(\lambda_j \mathbf{K}^{-1}\overrightarrow{O_cI_{uj}} - \lambda_i \mathbf{K}^{-1}\overrightarrow{O_cI_{ui}}) \\ & \cdot (\lambda_l \mathbf{K}^{-1}\overrightarrow{O_cI_{ul}} - \lambda_i \mathbf{K}^{-1}\overrightarrow{O_cI_{ui}})] \\ & \cdot (\lambda_n \mathbf{K}^{-1}\overrightarrow{O_cI_{un}} - \lambda_i \mathbf{K}^{-1}\overrightarrow{O_cI_{ui}}). \end{aligned} \quad (11)$$

The objective function $F$ is constructed through Eqs. (9) and (10) as

$$\min F = \sum E_1(i,\ j) + \sum E_2(i,\ j,\ l) + \sum E_3(i,\ j,\ l,\ n). \quad (12)$$

A penalty factor $M$ is applied for the objective function $E_4(i,j,l,n)$, which converges significantly faster than the objective functions $E_1(i,j)$, $E_2(i,j,l)$, and $E_3(i,j,l,n)$. By multiplying the objective function $E_4(i,j,l,n)$ with $M$, which controls the co-planarity error of the five reference points, the unconstrained nonlinear optimization objective function of $\lambda_i$ is built, where $i = 0,\ 1,\ 2,\ 3,\ 4$. The Levenberg–Marquardt optimization method is used to solve for $\lambda_i$.

$$\begin{aligned} F = {} & \sum_{i=0}^{4}\sum_{j=i+1}^{4} E_1(i,j) + \sum_{i=0}^{4}\sum_{\substack{j=i+1 \\ l\neq i,\ j}}^{4} E_2(i,\ j,\ l) \\ & + \sum_{i=0}^{4}\sum_{\substack{j=i+1 \\ l\neq i,\ j \\ n\neq i,\ j,\ l}}^{4} E_3(i,j,l,n) + M\cdot\sum_{i=0}^{4}\sum_{\substack{j=i+1 \\ l\neq i,\ j \\ n\neq i,\ j,\ l}}^{4} E_4(i,j,l,n). \end{aligned} \quad (13)$$

The 3D coordinates of the reference points can be obtained by applying Eq. (13) iteratively along until $\lambda_i$ is stable and provides reasonable initial values of $\lambda_i$, which are acquired in the analytical solving process to ensure the accuracy and convergence speed of the nonlinear algorithm.

As shown in Fig. 3, the pose of the target from position 1 to position 2 is (**R, T**). According to the definition of the rotation and translation matrix, we deduce

$$\overrightarrow{P_{ci}^2P_{cj}^2} = \mathbf{R} \cdot \overrightarrow{P_{ci}^1P_{cj}^1}. \quad (14)$$

$\overrightarrow{P_{ci}^1P_{cj}^1}$ and $\overrightarrow{P_{ci}^2P_{cj}^2}$ confirm a 3×3 orthonormal matrix, respectively, and $\mathbf{R}$ is confirmed through these two orthonormal matrixes (the product of two orthonormal matrixes is still an orthonormal matrix, thereby ensuring the orthonormality of $\mathbf{R}$). The solving process is shown as

$$\begin{cases} \overrightarrow{h_1} = \lambda_2 \mathbf{K}^{-1}\overrightarrow{O_cI_{u2}} - \lambda_1 \mathbf{K}^{-1}\overrightarrow{O_cI_{u0}} \\ \quad / \parallel \lambda_2 \mathbf{K}^{-1}\overrightarrow{O_cI_{u2}} - \lambda_1 \mathbf{K}^{-1}\overrightarrow{O_cI_{u0}} \parallel \\ \overrightarrow{h_2} = \lambda_3 \mathbf{K}^{-1}\overrightarrow{O_cI_{u3}} - \lambda_4 \mathbf{K}^{-1}\overrightarrow{O_cI_{u4}} \\ \quad / \parallel \lambda_3 \mathbf{K}^{-1}\overrightarrow{O_cI_{u3}} - \lambda_4 \mathbf{K}^{-1}\overrightarrow{O_cI_{u4}} \parallel \\ \overrightarrow{h_3} = \overrightarrow{h_1} \times \overrightarrow{h_2} \\ \mathbf{S}_{A1} = \left[ \overrightarrow{h_1^{A1}} \quad \overrightarrow{h_2^{A1}} \quad \overrightarrow{h_3^{A1}} \right] \\ \mathbf{S}_{A2} = \left[ \overrightarrow{h_1^{A2}} \quad \overrightarrow{h_2^{A2}} \quad \overrightarrow{h_3^{A2}} \right] \\ \mathbf{R} = \mathbf{S}_{A1}^{-1} \cdot \mathbf{S}_{A2} \\ \mathbf{T} = P_{ci}^2 - \mathbf{R} \cdot P_{ci}^1 \end{cases} . \quad (15)$$

Considering that the experiment system is a three-axis stage, and the three axes do not usually intersect at a point, the measurement model should be constructed for such a situation, as shown in Fig. 3.

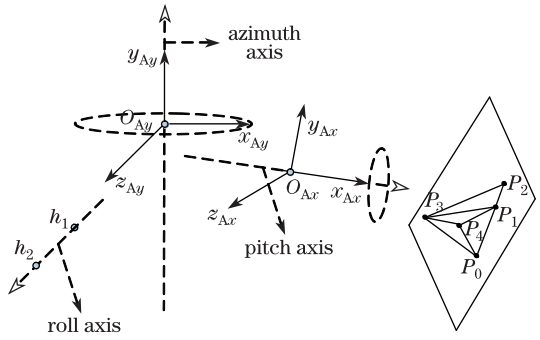


Fig. 3. Measurement model of the three-axis turntable.

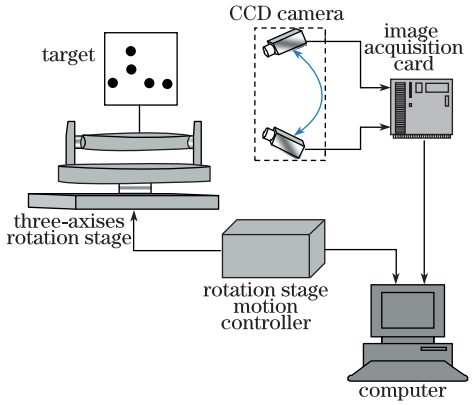


Fig. 4. Schematic diagram of the experimental system.

Figure 4 shows the experimental system, which consists of a target, a rotation and translation stage, two CCD cameras, and a computer. Images in the different locations are captured with a single CCD camera, and the pose of the target could be obtained by comparing the image in the current location with that in the initial location (zero location).

This work used Teli CSB4000F-20, with resolution of 2008 (h)×2044 (v) and pixel size of 0.006×0.006 (mm). The lens is a Pentax 25 mm. The rotation range of the stage is electronically controlled in the azimuth axis, pitching axis, and roll axis (±160°, ±80°, and ±45°, respectively). The positioning accuracy is less than 20. The calibration results of the intrinsic parameters of the camera are shown in Table 1.

The target pose measurement experiment is divided into two parts, namely, the digital simulation experiment and the practical measurement experiment.

To validate accuracy and noise immunity of the method in the present article, the proposed method is compared with the least squares method in Ref. [15] and the POSIT method in Ref. [13] during the digital simulation experiment process. In this process, the pinhole imaging model of the camera is simulated; thus, the reference points are transformed with perspective projection, and the simulated image coordinates of the reference points are acquired. By adding random Gaussian noises of different intensities to the image coordinates, the target pose is calculated using the proposed algorithm, the POSIT algorithm, and the least-squares algorithm. The results of the actual and calculation values, which represent the noise immunity ability of the algorithm, are shown in Fig. 5(a).

With the 5° interval, the 3-axis turntable takes the target to rotate at the three degrees of freedom (azimuth, pitch, and roll). The CCD camera captures an image at each location. The angle between the current position and the initial position (zero position) is calculated according to parts 2 and 3. The measuring range of the

**Table 1. Calibration Results**

| Parameter | $f_x$ | $f_y$ | $c_x$ | $c_y$ |
|---|---|---|---|---|
| Camera1 | 4366.406 | 4366.174 | 998.898 | 1005.258 |
| Camera2 | 4379.274 | 4379.074 | 1013.002 | 1022.227 |
| **Parameter** | $k_1$ | $k_2$ | $p_1$ | $p_2$ |
| Camera1 | −0.380 | 0.281 | 0.00179 | 0.00019 |
| Camera2 | −0.374 | 0.161 | 0.00146 | −0.00006 |

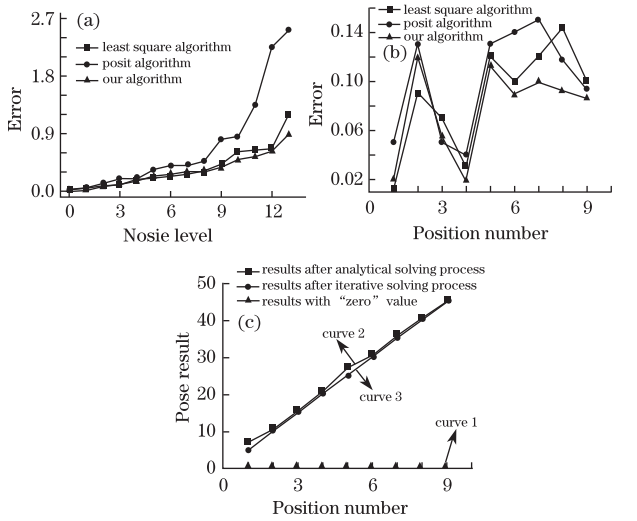


Fig. 5. Contrast experimental results.

**Table 2. Measurement Results of the Azimuth Angle**

| Serial No. | Position of Target (dey.) | Angle between Location $i$ and Location 0 (dey.) Actual Value (dey.) | Error (dey.) | Relative Error (%) |
|---|---|---|---|---|
| 0 | 0 | | | |
| 1 | 5 | 5.052 | 0.052 | 1.04 |
| 2 | 10 | 10.109 | 0.109 | 1.09 |
| 3 | 15 | 15.082 | 0.082 | 0.55 |
| 4 | 20 | 19.982 | −0.018 | −0.09 |
| 5 | 25 | 24.977 | −0.023 | −0.09 |
| 6 | 30 | 30.001 | 0.001 | 0.003 |
| 7 | 35 | 35.062 | 0.062 | 0.18 |
| 8 | 40 | 40.061 | 0.061 | 0.15 |
| 9 | 45 | 45.084 | 0.084 | 0.19 |
| 10 | 50 | 50.072 | 0.072 | 0.14 |
| 11 | 55 | 55.045 | 0.045 | 0.08 |
| 12 | 60 | 59.992 | −0.008 | −0.01 |
| 13 | 65 | 64.919 | −0.081 | −0.12 |
| 14 | 70 | 69.879 | −0.121 | −0.17 |
| 15 | 75 | 74.932 | −0.068 | −0.09 |
| 16 | 80 | 79.892 | −0.108 | −0.14 |
| 17 | 85 | 84.919 | −0.081 | −0.10 |
| 18 | 90 | 89.895 | −0.105 | −0.12 |

**Table 3. Measurement Results of the Pitch Angle**

| Serial No. | Position of Target (dey.) | Angle between Location $i$ and Location 0 (dey.) | | |
|---|---|---|---|---|
| | | Actual Value (dey.) | Error (dey.) | Relative Error (%) |
| 0 | 0 | | | |
| 1 | 5 | 4.994 | −0.006 | −0.12 |
| 2 | 10 | 10.006 | 0.006 | 0.06 |
| 3 | 15 | 14.953 | −0.047 | −0.31 |
| 4 | 20 | 19.932 | −0.068 | −0.34 |
| 5 | 25 | 24.946 | −0.054 | −0.22 |
| 6 | 30 | 29.964 | −0.036 | −0.12 |
| 7 | 35 | 34.999 | −0.001 | −0.003 |
| 8 | 40 | 39.967 | −0.033 | −0.08 |
| 9 | 45 | 44.930 | −0.070 | −0.16 |
| 10 | 50 | 49.922 | −0.078 | −0.16 |
| 11 | 55 | 54.959 | −0.041 | −0.07 |
| 12 | 60 | 60.035 | 0.035 | 0.06 |

**Table 4. Measurement Results of the Roll Angle**

| Serial No. | Position of Target (dey.) | Angle between Location $i$ and Location 0 (dey.) | | |
|---|---|---|---|---|
| | | Actual Value (dey.) | Error (dey.) | Relative Error (%) |
| 0 | 0 | | | |
| 1 | 5 | 5.013 | 0.013 | 0.26 |
| 2 | 10 | 10.034 | 0.034 | 0.34 |
| 3 | 15 | 15.045 | 0.045 | 0.30 |
| 4 | 20 | 20.063 | 0.063 | 0.32 |
| 5 | 25 | 25.079 | 0.079 | 0.32 |
| 6 | 30 | 30.099 | 0.099 | 0.33 |
| 7 | 35 | 35.115 | 0.115 | 0.38 |
| 8 | 40 | 40.049 | 0.049 | 0.12 |
| 9 | 45 | 45.009 | 0.009 | 0.02 |
| 10 | 50 | 50.011 | 0.011 | 0.02 |
| 11 | 55 | 55.016 | 0.016 | 0.03 |
| 12 | 60 | 60.021 | 0.021 | 0.04 |

azimuth angle is set to 0°–90°; the measuring range of the pitch angle is set to 0°–60°; the measurement range of the roll angle is set to 0°–60°. The results are shown in Table 2 (azimuth angle with RMS of 0.076°), Table 3 (pitch angle with RMS of 0.035°), and Table 4 (roll angle with RMS of 0.036°). The contrast experiments of the practical measurement of three methods are shown in Fig. 5(b). As shown in Fig. 5(c), if the pose is solved directly through the establishment of the objective function, the initial value for the pose is set as 0 (curve 1), and the final results (curve 3) are obtained through some iterations. When the present method is used, the initial value (curve 2) for the pose is solved with the analytical solution, then the final results (curve 3) are achieved through the establishment of the objective function with less iterations. Curve 2 is closer to curve 3 than to curve 1, and the first step (the analytical solution process) is easy to implement. Thus, the amount of iterations and calculation can be greatly reduced.

In conclusion, we present a two-step method for pose estimation based on five co-planar reference points. Experimental results show that the RMS of the azimuth angle reaches 0.076° in the measurement range of 0°–90°; the RMS of the pitch angle reaches 0.035° in the measurement range of 0°–60°; and the RMS of the roll angle reaches 0.036° in the measurement range of 0°–60°.

This work was supported by the Important National Science & Technology Specific Project (No.2009ZX04014-092).

## References

1. Y. Hao, Q. Wu, C. Zhou, S. Li, and F. Zhu, Robot **28,** 656 (2006).
2. F. Huang and H. Qian, Opto-Electron. Eng. (in Chinese) **34,** 69 (2007).
3. S. Hao, Z. Chai, Z. Zhang, and Y. Hu, Electron. Opt. Control (in Chinese) **9,** 19 (2002).
4. L. Xu and J. Wang, Fire Control and Command Control (in Chinese) **31,** 80 (2006).
5. S. Chen, "Study on the monocular vision measurement system of the position and orientation of object" PhD. Thesis (Tianjin University, 2007).
6. J. Zhang, C. Sun, and W. Cai, opt. tech. **36,** 187 (2010).
7. Z. Chao, G. Jiang, W. Huang, S. Song, and Q. Yu, Opt. Prec. Eng. (in Chinese) **18,** 2044 (2010).
8. T. Jin, H. Jia, W. Hou, R. Yamamoto, N. Nagai, Y. Fujii, K. Maru, N. Ohta, and K. Shimada, Chin. Opt. Lett. **8,** 601 (2010).
9. M. L. Liu and K. H. Wong, Pattern Recognition Lett. **20,** 69 (1999).
10. Z. Hu and F. Wu, IEEE Trans. Pattern Anal. Mach. Intell. **24,** 550 (2002).
11. P. Wang, C. Sun, and Z. Zhang, Chin. J. Scient. Instrum. (in Chinese) **32,** 1126 (2011).
12. J. Tang, W. Chen, and J. Wang, Applied Mathematics and Computation **205,** 628 (2008).
13. D. F. DeMenthon and L. S. Davis, International Journal on Computer Vision **15,** 123 (1995).
14. Z. Zhang, C. Sun, P. Sun, and P. Wang, Chin. Opt. lett. **9,** 081501 (2011).
15. S. Chen, T. Zhou, X. Zhang, and C. Sun, Chin. J. Sens. and Actuat. (in Chinese) **20,** 2011 (2007).
16. P. Wang, X. Xu, Z. Zhang, and C. Sun, Chin. Opt. Lett. **8,** 55 (2010).
17. C. Wildey, D. L. MacFarlane, A. Goyal, K. Gopinath, S. Cheshkov, and R. Briggs, Appl. Opt. **50,** 2088 (2011).
18. S. H. Or, W. S. Luk, K. H. Wong, and I. King, Image and Vision Computing, **16,** 353 (1998).
19. L. Tao, "Study on Key Technique for Color 3D Laser Scanning System" (in Chinese) PhD. Thesis (Tianjin University, 2005).