IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received January 17, 2018, accepted April 3, 2018, date of publication April 17, 2018, date of current version May 9, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2827565

Complete Initial Solutions for Iterative Pose Estimation From Planar Objects

KAI ZHOU[1], XIANGJUN WANG[1], ZHONG WANG[1], HONG WEI[2], AND LEI YIN[1]
[1]State Key Laboratory of Precision Measuring Technology and Instruments, Tianjin University, Tianjin 300072, China
[2]Department of Computer Science, University of Reading, Reading RG6 6AY, U.K.

Corresponding author: Xiangjun Wang (tjuxjw@126.com)

This work was supported by the Innovation Fund of Tianjin University under Grant C02014021.

ABSTRACT Camera pose estimation from the image of a planar object has important applications in photogrammetry and computer vision. In this paper, an efficient approach to find the initial solutions for iterative camera pose estimation using coplanar points is proposed. Starting with homography, the proposed approach provides a least-squares solution for absolute orientation, which has a relatively high accuracy and can be easily refined into one optimal pose that locates local minima of the according error function by using Gauss-Newton scheme or Lu's orthogonal iteration algorithm. In response to ambiguities that exist in pose estimation from planar objects, we propose a novel method to find initial approximation of the second pose, which is different from existing methods in its concise form and clear geometric interpretation. Thorough testing on synthetic data shows that combined with currently employed iterative optimization algorithm, the two initial solutions proposed in this paper can achieve the same accuracy and robustness as the best state-of-the-art pose estimation algorithms, while with a significant decrease in computational cost. Real experiment is also employed to demonstrate its performance.

INDEX TERMS Pose estimation, perspective-*n*-point problem, pose ambiguity.

I. INTRODUCTION

The PnP (Perspective-n-Point) problem was first proposed by Fischler and Bolles [1] in 1981, which can be defined as determining the 6 degrees of freedom of a camera's pose given its intrinsic parameters and a set of correspondences between 3D reference points and their 2D images. It is a basic problem in the fields of computer vision and photogrammetry, and has been extensively studied during the past decades. This paper focuses on the case in which the 3D points are coplanar. The applications of camera pose estimation from the image of coplanar points can be found frequently in camera calibration, augmented reality and pose tracking of a space target since most of the calibration boards and man-made targets have a planar structure.

In theory, the position and orientation of a camera can be calculated from four or more coplanar but noncollinear points, if the intrinsic parameters of the camera and the correspondences are known. Researchers have applied both iterative and non-iterative approaches to compute the pose solution for the case of coplanar points. Among non-iterative approaches, closed form solutions have been formulated for configurations of fixed number of points, among which there are three points and four coplanar points. The P3P problem (with three noncollinear points) can have as many as four possible solutions [2]–[6]. While the P4P problem has a single theoretical solution when the coplanar points are in an ordinary configuration [1], [7]–[10]. As for configurations of arbitrary number of points (N>4), almost all non-iterative methods are designed to cope with reference point sets with generic positions (not necessarily coplanar) as far as we know [11]–[18]. But these methods can also be used in the coplanar case and some of them have achieved excellent performance both in accuracy and computational cost [15]–[18]. In particular, the OPnP algorithm proposed by Zheng *et al.* [17] have reached an accuracy comparable with iterative methods, with O(n)- complexity. Generally speaking, non-iterative approaches are much faster than iterative ones, but more sensitive to additive noise and may lose accuracy more or less. What's more, when applied for coplanar case, they may not be robust enough. Indeed, even the most accurate state-of-the-art non-iterative algorithm may perform poorly in some configurations of specific pose due to their inherent defect, as will be shown in section 5 of this article.

A natural alternative to non-iterative approaches is iterative ones. The classical iterative approach used in photogrammetry is to formulate pose estimation as a nonlinear least-squares problem which minimizes a cost function such as reprojection error, and to solve it by nonlinear optimization algorithms, most typically, the Gauss-Newton method [19]. Lu *et al.* [20] formulated the relative pose estimation problem as a minimization of the object-space collinearity error, and proposed an orthogonal iterative algorithm (**LHM**) which is globally convergent. **LHM** is one of the most widely used iterative methods for the PnP problem in recent years. Based on minimization of a nonlinear cost function, iterative approaches can achieve high estimation accuracy. However they have some apparent drawbacks. In addition to large computational cost, these local optimization based methods suffer from the risk of getting trapped into local minimum, and provide poor results when they indeed do so. Therefore, a good initial guess is needed to converge to the correct solution. Lu *et al.* [20] used weak perspective approximation as the default initial guess for their orthogonal iterative algorithm, which is proven to be unstable when applied to coplanar cases. Oberkampf *et al.* [21] first discussed ambiguities that exist in pose estimation for coplanar feature points, and they developed their **POSIT** algorithm which uses scaled orthographic projection at each iteration step. **POSIT** starts from the two minima under orthography, maintains two alternative solutions, and iteratively refines up to two different poses. When applying reprojection errors for choosing the better one, planar **POSIT** achieves higher robustness than that of **LHM**. Schweighofer and Pinz [22] gave a comprehensive interpretation of ambiguities existing in the planar case and they enhanced the robustness of **LHM** by taking two distinct local minima into account. Based on the first local minima, they derived an analytical solution that locates the second minima, and then use it as an initial value for iterative algorithms to get the second pose. The **SP** method (method by Schweighofer and Pinz) is one of the most robust and accurate algorithms for the planar case up to date. However, it did not give a good initial guess for iteration in the first place, thus being time-consuming even when the noise level is low. What's more, its approach for obtaining the second initial guess is complicated.

In this paper, we propose a novel approach to find initial solutions for iterative pose estimation using coplanar points. In our approach, both initial solutions for the two poses that locate local minima are achieved. The first initial solution is derived from homography applying linear least square method, which is easy to manipulate and require almost negligible cost of computation. This initial solution is accurate enough so that iterative methods using it as a start point can always converge to the right solution if the image noise is relatively low. The method for searching for the second initial solution in this paper is different from that of Schweighofer and Pinz [22]. Our method is proposed based on the assumption that when the noise level is high, the solution will have a chance to flip to a "mirror" pose. The mirror pose is symmetric to the original pose about a known plane, which has a much more intuitive and concise expression. A comprehensive test on synthetic data was conducted in this study. The results have shown that compared with the method proposed by Schweighofer and Pinz, the aforementioned two initial guesses can reach a solution with the same accuracy and robustness when applying the same iteration algorithm, while with a dramatic decrease in computing time.

II. PROBLEM FORMULATION

Given n coplanar points $\mathbf{q}_i = [X_i\ Y_i\ 0]^{\mathrm{T}}, i = 1, 2, \ldots, n$, on XY-plane in object reference frame, and their corresponding projections on normalized image plane $\mathbf{p}_i = [u_i\ v_i\ 1]^{\mathrm{T}}$, the perspective imaging equation can be expressed as follows:

$$\lambda_i \mathbf{p}_i = \mathbf{R}\mathbf{q}_i + \mathbf{t}, \quad i = 1, 2, \ldots, n, \tag{1}$$

where λ_i denotes the depth factor of the i-th point. The rotation matrix $\mathbf{R}$ and the translation vector $\mathbf{t}$, accounting for camera orientation and position respectively, are the unknowns to be retrieved. Considering that the third element of $\mathbf{q}_i$ $(i = 1, 2, \ldots, n)$ is zero, equation (1) could be replaced by:

$$\begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = \frac{1}{\lambda_i} \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{t} \end{bmatrix} \begin{bmatrix} X_i \\ Y_i \\ 1 \end{bmatrix}, \quad i = 1, 2, \ldots, n, \tag{2}$$

where $[X_i\ Y_i\ 1]^{\mathrm{T}} (i = 1, 2, \ldots, n)$ can be defined as homogeneous coordinates of the coplanar points in the object reference frame. $\mathbf{r}_1$ and $\mathbf{r}_2$ are the first two columns of rotation matrix $\mathbf{R}$. It is clear that the map between points $[X_i\ Y_i\ 1]^{\mathrm{T}}$ and $[u_i\ v_i\ 1]^{\mathrm{T}} (i = 1, 2, \ldots, n)$ is a planar homography, which has the first two columns orthogonal and with the same norm.

In the following sections, we make an assumption that the centroid of coplanar points $\mathbf{q}_i$ $(i = 1, 2, \ldots, n)$ is aligned with the origin of the object reference frame for convenience.

III. THE FIRST INITIAL SOLUTION

The first initial solution is derived from planar homography. The approach is mainly inspired by the method in [23] by which the best rotation matrix is estimated from a general 3×3 matrix, but makes a variation to properly suit the algorithm for non-square matrix with a scale factor. It consists of two steps. The first step is to solve for the 3×3 homography matrix $\mathbf{H}$ between points on XY-plane in the object reference frame and their projections on the normalized image plane. Given a set of four or more point correspondences, $\mathbf{H}$ can be determined up to a non-zero scale factor applying DLT(Direct Linear Transformation) which is detailed in [19]. It should be noticed that the first two columns of $\mathbf{H}$ are not exactly orthogonal in practice due to the existence of random noise. The second step is to normalize $\mathbf{H}$ to the "closest" matrix which has the first two columns orthogonal and with the same norm. Thus the initial approximation of $\mathbf{R}$ and $\mathbf{t}$ can be obtained.

Let $\mathbf{Q}$ be the 3×2 matrix consisting of the first two columns of $\mathbf{H}$. The normalized matrix is expressed by:

$$k\mathbf{P} = k\left[\mathbf{r}_1 \;\; \mathbf{r}_2\right] \tag{3}$$

where $\mathbf{P}$ denotes a 3×2 matrix satisfying $\mathbf{P}^{\mathrm{T}}\mathbf{P} = \mathbf{I}_{2\times2}$, k is a positive scale factor. The second step is to solve the following problem:

$$\min_{k\mathbf{P}} \|k\mathbf{P}-\mathbf{Q}\|_{\mathrm{F}}^2 \text{ subject to } \mathbf{P}^{\mathrm{T}}\mathbf{P} = \mathbf{I}_{2\times2} \tag{4}$$

The objective function can be written as:

$$\begin{aligned}\|k\mathbf{P}-\mathbf{Q}\|_{\mathrm{F}}^2 &= trace\left((k\mathbf{P}-\mathbf{Q})^{\mathrm{T}}(k\mathbf{P}-\mathbf{Q})\right)\\ &= 2k^2 - 2k * trace\left(\mathbf{P}^{\mathrm{T}}\mathbf{Q}\right) + trace\left(\mathbf{Q}^{\mathrm{T}}\mathbf{Q}\right)\end{aligned} \tag{5}$$

Since $trace(\mathbf{Q}^{\mathrm{T}}\mathbf{Q})$ is constant, the problem is equivalent to the one of minimizing $2k^2 - 2k{*}trace(\mathbf{P}^{\mathrm{T}}\mathbf{Q})$. Note that if k is fixed, this expression is a monotonic decreasing function of $trace(\mathbf{P}^{\mathrm{T}}\mathbf{Q})$. Therefore the optimal $\mathbf{P}$ would be the one that maximize $trace(\mathbf{P}^{\mathrm{T}}\mathbf{Q})$.

Let the singular value decomposition of $\mathbf{Q}$ be $\mathbf{U}_{3\times3}\mathbf{S}_{3\times2}\mathbf{V}_{2\times2}^{\mathrm{T}}$, then

$$\begin{aligned}trace\left(\mathbf{P}^{\mathrm{T}}\mathbf{Q}\right) &= trace\left(\mathbf{P}^{\mathrm{T}}\mathbf{U}_{3\times3}\mathbf{S}_{3\times2}\mathbf{V}_{2\times2}^{\mathrm{T}}\right)\\ &= trace\left(\mathbf{V}_{2\times2}^{\mathrm{T}}\mathbf{P}^{\mathrm{T}}\mathbf{U}_{3\times3}\mathbf{S}_{3\times2}\right)\\ &= trace\left(\begin{bmatrix} v_{11} & v_{21}\\ v_{12} & v_{22}\end{bmatrix}\begin{bmatrix} r'_{11} & r'_{12} & r'_{13}\\ r'_{21} & r'_{22} & r'_{23}\end{bmatrix}\begin{bmatrix} s_1 & 0\\ 0 & s_2\\ 0 & 0\end{bmatrix}\right)\\ &= s_1\left(v_{11}r'_{11}+v_{21}r'_{21}\right) + s_2\left(v_{12}r'_{12}+v_{22}r'_{22}\right)\end{aligned} \tag{6}$$

Among which

$$\mathbf{P}^{\mathrm{T}}\mathbf{U}_{3\times3} = \begin{bmatrix} r'_{11} & r'_{12} & r'_{13}\\ r'_{21} & r'_{22} & r'_{23}\end{bmatrix} \tag{7}$$

It can be verified that the two rows of $\mathbf{P}^{\mathrm{T}}\mathbf{U}_{3\times3}$ are orthogonal and have norm 1. We can find

$$\begin{cases} r'_{11} = v_{11}\\ r'_{21} = v_{21}\\ r'_{12} = v_{12}\\ r'_{22} = v_{22}\end{cases}$$

namely

$$\begin{aligned}\mathbf{P}^{\mathrm{T}}\mathbf{U}_{3\times3} &= \begin{bmatrix} v_{11} & v_{12} & 0\\ v_{21} & v_{22} & 0\end{bmatrix}\\ &= \left[\mathbf{V}_{2\times2} \mid \mathbf{0}\right]\end{aligned} \tag{8}$$

such that $trace(\mathbf{P}^{\mathrm{T}}\mathbf{Q})$ reaches its maximum s_1+s_2. Then the optimal $\mathbf{P}$ is obtained:

$$\mathbf{P} = \mathbf{U}_{3\times3}\left[\mathbf{V}_{2\times2} \mid \mathbf{0}\right]^{\mathrm{T}} \tag{9}$$

Substituting $trace(\mathbf{P}^{\mathrm{T}}\mathbf{Q})$ by s_1+s_2in (5), we have

$$\|k\mathbf{P}-\mathbf{Q}\|_{\mathrm{F}}^2 = 2k^2 - 2k\left(s_1+s_2\right) + trace\left(\mathbf{Q}^{\mathrm{T}}\mathbf{Q}\right) \tag{10}$$

It is a quadratic function of k and reaches minimum at

$$k = \frac{s_1+s_2}{2} \tag{11}$$

In conclusion, the solution for (4) is

$$\begin{cases}\mathbf{P} = \mathbf{U}_{3\times3}\left[\mathbf{V}_{2\times2} \mid \mathbf{0}\right]^T\\ k = \dfrac{s_1+s_2}{2}\end{cases} \tag{12}$$

Accordingly the initial approximation of $\mathbf{R}$ and $\mathbf{t}$ are expressed as

$$\begin{cases}\mathbf{R} = \left[\mathbf{r}_1 \;\; \mathbf{r}_2 \;\; \mathbf{r}_1\times\mathbf{r}_2\right]\\ \mathbf{t} = \dfrac{\mathbf{h}_3}{k}\end{cases} \tag{13}$$

where $\mathbf{r_1}$ and $\mathbf{r_2}$ are the two columns of $\mathbf{P}$, and $\mathbf{h}_3$ is the third column of $\mathbf{H}$. Since this initial solution is derived from homography, we call it **HI** in the following sections.

It should be noticed that, since $\mathbf{t}(3)$ is always positive, $\mathbf{H}$ should be multiplied by -1 period to the second step if $\mathbf{H}(3,3)$ is negative, in order to ensure the correct solution of $\mathbf{R}$ and $\mathbf{t}$.

IV. THE SECOND INITIAL SOLUTION

To evaluate the performance of **HI**, we combine it with **LHM** [20], which is one of the most popular iterative approaches for pose estimation. Taking **HI** as the start point of **LHM**, we can refine up to one unique pose solution. During a full test for configuration of 10 coplanar points, **HI+LHM** achieves 100% correctness in pose estimation (The errors increase smoothly with the increase of image noise, no sharp changes.) when the image noise level is not greater than 6 pixels, as will be shown in section 5. When the number of coplanar points is small, or the image noise grows larger, wrong solutions arise at the place where the angle between image plane and object plane is extremely large. We record all of the wrong solutions that occur in 115600 experiments on synthetic data that traverse all pitching and yaw angles of the object plane from −80° to 80°, compare them with the corresponding true poses, and find that the translation vector $\mathbf{t}$ of each wrong pose is not so far different from its true value, and the real difference lies in the rotation matrix $\mathbf{R}$. We also noticed that if the origin of the object reference frame is on the optical axis of the camera, the relation between the rotation matrix of the wrong pose and that of the true pose is quite straightforward, it is summarized in the following approximate representation:

$$\underbrace{\begin{bmatrix} r_{11} & r_{12} & \mathbf{r_{13}}\\ r_{21} & r_{22} & \mathbf{r_{23}}\\ \mathbf{r_{31}} & \mathbf{r_{32}} & r_{33}\end{bmatrix}}_{\substack{\mathbf{R_1}\\ \textbf{Wrong Pose}}} \qquad \underbrace{\begin{bmatrix} r_{11} & r_{12} & \mathbf{-r_{13}}\\ r_{21} & r_{22} & \mathbf{-r_{23}}\\ \mathbf{-r_{31}} & \mathbf{-r_{32}} & r_{33}\end{bmatrix}}_{\substack{\mathbf{R_2}\\ \textbf{True Pose}}}$$

That is to say, with the existence of tiny residual error, $\mathbf{R_1}$ and $\mathbf{R_2}$ of the two poses can transform to each other by just inverting the first two entries of the third row and the first two entries of the third column, while the rest entries keep unchanged.

Considering that the rotation matrix may be written as:

$$\mathbf{R} = \begin{bmatrix} \mathbf{u} & \mathbf{v} & \mathbf{w} \end{bmatrix} \tag{14}$$

where **u**, **v**, **w** represents the basis of the object reference frame expressed in the camera coordinate system. Then the relationship between $\mathbf{R_1}$ and $\mathbf{R_2}$ can be interpreted in a visualized geometrical sense, as shown in Fig. 1.

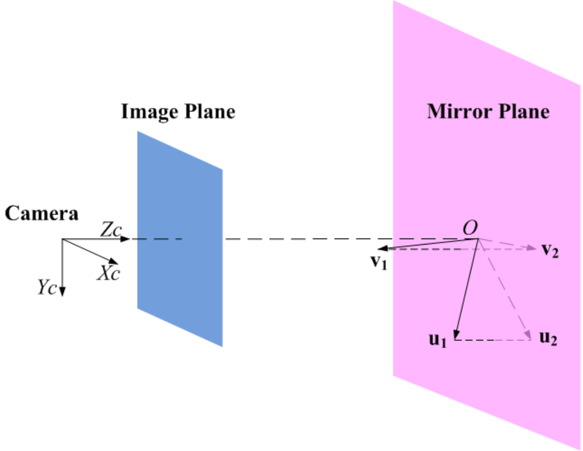


FIGURE 1. The relation between wrong pose and true pose in the case that the origin is on the optical axis of the camera.

In Fig. 1, $\mathbf{u_1}$, $\mathbf{v_1}$ and $\mathbf{u_2}$, $\mathbf{v_2}$ are the first two columns of $\mathbf{R_1}$ and $\mathbf{R_2}$ respectively. They represent the basis vectors of object planes corresponding to the wrong pose and the true pose. The two pairs of vectors are symmetric with each other about a plane passing through the origin of the object reference frame O and parallel to the image plane. We can understand Fig. 1 by considering that when the noise is large, the solution will have a chance to flip to a "mirror" pose.

In the case that O deviates from the optical axis of the camera in a large scale, the relation between the two poses is not easy to find at the first sight of rotation matrices. However, there is also regularity to follow. Indeed, from the geometric point of view, the mirror plane always passes through O, and is perpendicular to the translation vector **t**, namely the coordinate vector of O in the camera reference frame. This conclusion has been verified through experiments on ordinary positions of O with large noise introduced. It is shown in Fig. 2.

In summary, the relation between two poses can be written in a unified form, as shown in (15).

$$\begin{cases} \mathbf{u_2} = \mathbf{u_1} - 2\dfrac{\mathbf{t}\mathbf{t}^{\mathrm{T}}}{\mathbf{t}^{\mathrm{T}}\mathbf{t}}\mathbf{u_1} \\ \mathbf{v_2} = \mathbf{v_1} - 2\dfrac{\mathbf{t}\mathbf{t}^{\mathrm{T}}}{\mathbf{t}^{\mathrm{T}}\mathbf{t}}\mathbf{v_1} \\ \mathbf{t_2} = \mathbf{t_1} \end{cases} \tag{15}$$

where $\mathbf{u_2}$,$\mathbf{v_2}$ are the first and the second column of $\mathbf{R_2}$. The third column $\mathbf{w_2}$ can be calculated by the cross-product of $\mathbf{u_2}$ and $\mathbf{v_2}$. In practice, after the first pose solution is obtained, we can use (15) to get the initial solution for the second pose. In addition, since the true value of **t** is unknown, the calculated one $\mathbf{t_1}$ of the first pose is used to replace it in (15).

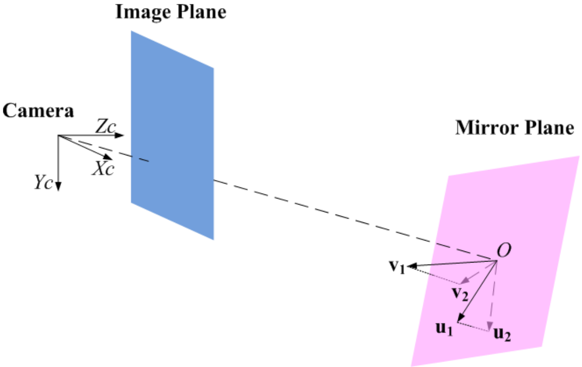


FIGURE 2. The relation between wrong pose and true pose in ordinary case.

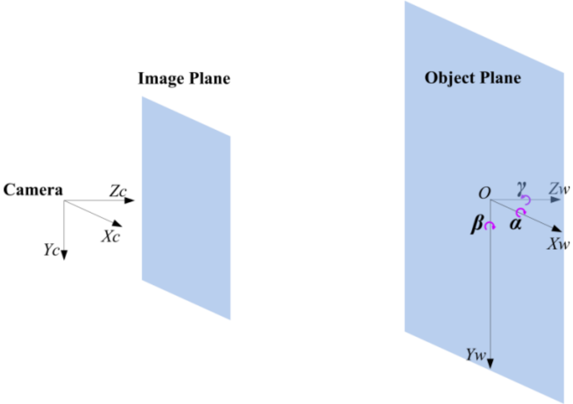


FIGURE 3. The definition of α, β and γ in experimental setup.

Compared with the method proposed by Schweighofer and Pinz [22] for calculating the second initial pose, which needs a series of coordinate transformations and a solution for a polynomial of degree four, the approach represented by (15) has a much more concise form, and correspond to a clear geometric interpretation as well.

The two initial solutions above lead us to our new pose estimation approach:

Step 1: Calculate the homography matrix between coplanar points in the object reference frame and their projections on the normalized image plane, then use (12) and (13) to get the first initial solution.

Step 2: Estimate a pose ($\mathbf{R_1}$, $\mathbf{t_1}$) using the first initial solution as a start point, applying any existing iterative pose estimation algorithm. In our experiments, the **LHM** proposed in [20] is adopted.

Step 3: Use (15) to obtain the second initial solution ($\mathbf{R_2}$, $\mathbf{t_2}$).

Step 4: Refine ($\mathbf{R_2}$, $\mathbf{t_2}$) to get the second pose by applying the same iterative algorithm as Step 2.

Step 5: Decide the final correct pose (**R**, **t**) based on the error function used by the iterative approach in Step 2 and Step 4. In our experiments, the object-space collinearity error is employed.

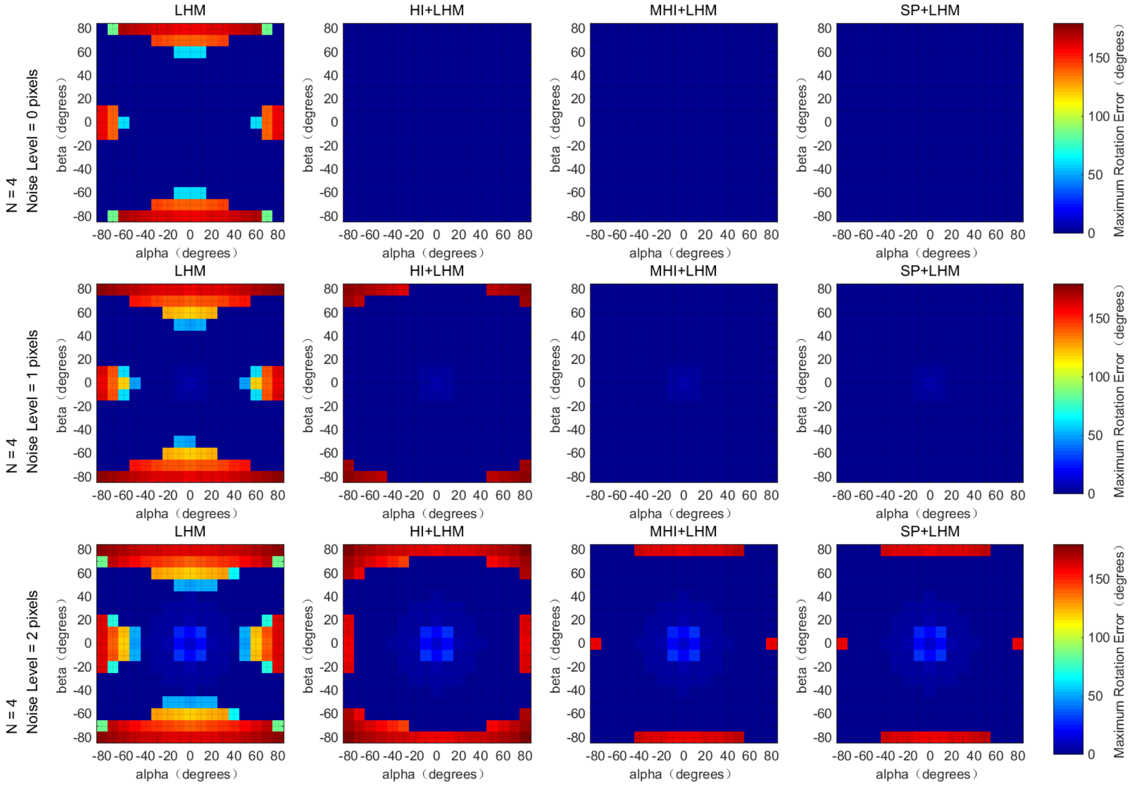


FIGURE 4. **The distribution of maximum rotation errors for the configuration of N=4.**

As a supplement, if the origin O is not aligned with the centroid of coplanar points, a translation $\mathbf{t_0}$ for alignment should be applied before Step 1. After obtaining ($\mathbf{R}$, $\mathbf{t}$) in step 5, we can replace $\mathbf{t}$ by $\mathbf{t}+\mathbf{Rt_0}$ to recover the final pose.

V. EXPERIMENT RESULTS

A. EXPERIMENTS WITH SYNTHETIC DATA

1) EXPERIMENTAL SETUP

In this section, we experimentally investigate our approach for pose estimation from planar objects, referred to as **MHI+LHM**, and compare it with state-of-the-art methods. For all experiments, we use an internally calibrated camera with focal length of *6mm* and pixel size of *3.75um×3.75um*.

In the tests demonstrated in 2) to 4), we evaluate the performance of our algorithm by using synthetic planar objects with 4 coplanar points and 10 coplanar points respectively. In the configuration of 4 coplanar points, the 4 points are located at the 4 corners of a square with a size of $1m\times1m$. In the configuration of 10 coplanar points, 4 points are also located at the corners of the square, and the other 6 points are positioned randomly inside this square. For each configuration, two cases of positions of camera relative to the object plane are considered. In the first case, the translation vector $\mathbf{t}$ is chosen as (0, 0, 5000)(*mm*), which indicates the square is located at the center of camera's vision field. In the second case, $\mathbf{t}$ is chosen as (2000, 1500, 5000)(*mm*), which indicates the location of square is close to the edge of the field. For each case, synthetic images are obtained using a number of pitching and yaw angles for the square and different levels of image noise. At noise level x, the coordinates of image points are disturbed by vertical and horizontal perturbations of $\pm x$ *pixels*. The roll angle γ is specified as 0° and 90°. The pitching angle α ranges from $-80°$ to 80° with a step length of 10° and yaw angle β ranges from $-80°$ to 80° with the same step length. Therefore there are $2\times17\times17$ triplets of γ, α and β. For each triplet of angles, 200 synthetic images are obtained with the same noise level. The camera poses are computed by different algorithms from synthetic images, and the results are compared with the actual camera poses. The mean and maximum errors as well as number of wrong solutions are recorded, respectively with different noise levels.

2) ANALYSIS OF ROTATION ERROR DISTRIBUTION

The performance of **LHM** [20], **HI+LHM** (**LHM** using only the first initial solution), **MHI+LHM** (**LHM** using two initial solutions) and **SP+LHM** [22] under different image noise levels are evaluated and presented in the form of colormaps, which show the distribution of maximum rotation errors against α and β, while γ is 0° and $\mathbf{t}$ is (0, 0, 5000)(*mm*). For the case of 4 coplanar points, three image noise levels of 0, 1 and 2(*pixels*) are considered. For the case of 10 coplanar points, the noise levels are 0, 3 and 6(*pixels*).

As can be seen in Fig. 4 and Fig. 5, under the noise free condition, **HI+LHM**, **MHI+LHM** and **SP+LHM** can all achieve zero rotation error in both configurations of N=4 and

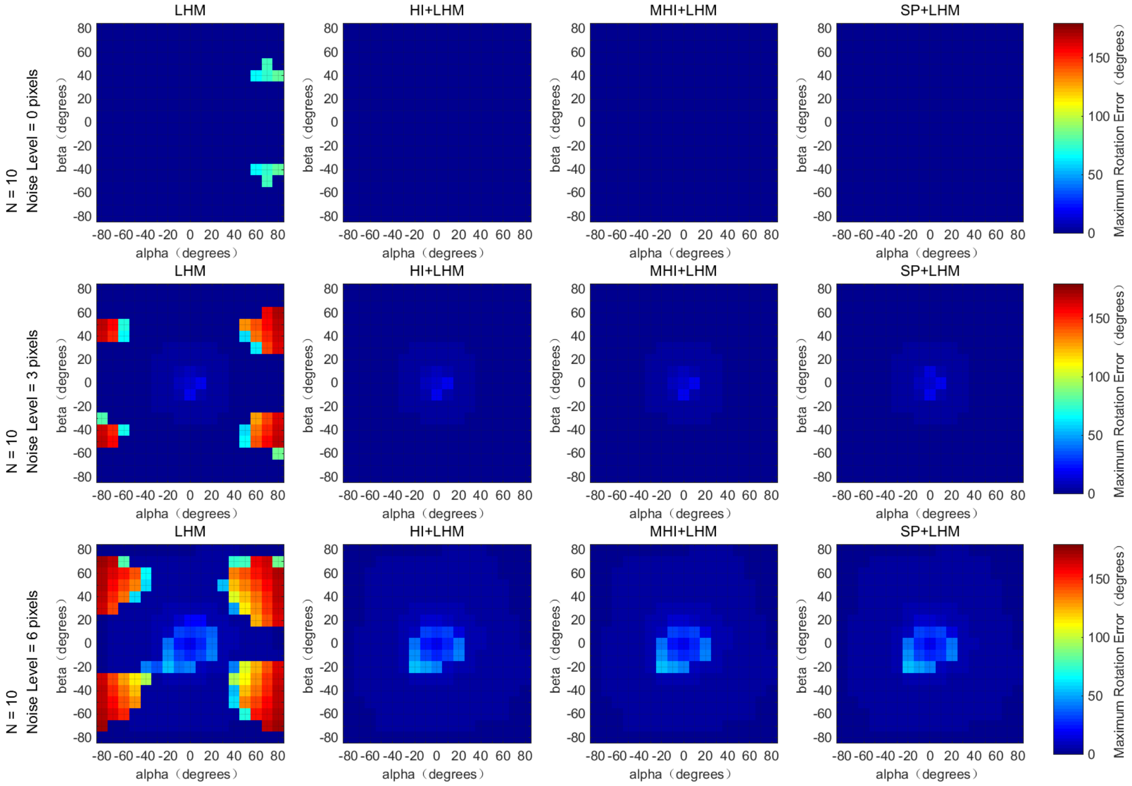


FIGURE 5. The distribution of maximum rotation errors for the configuration of N=10.

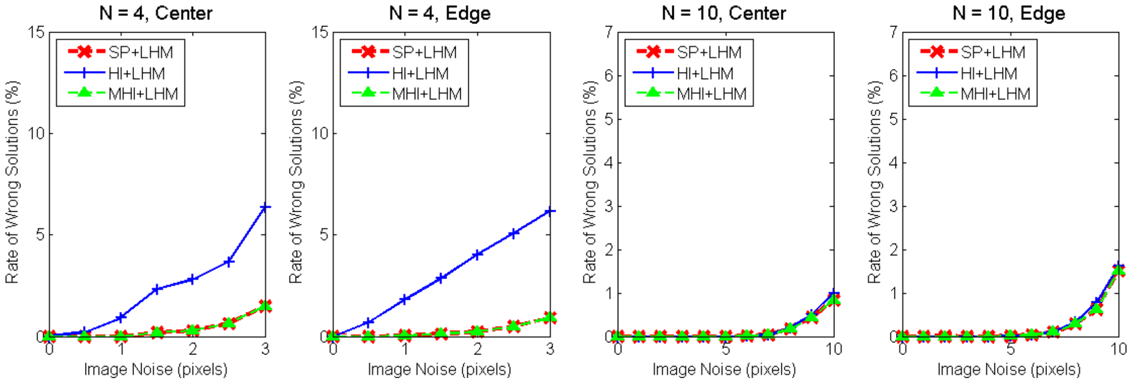


FIGURE 6. The rate of wrong solutions for N = 4 and N = 10.

N=10. As for the configuration of N = 4, **MHI+LHM** and **SP+LHM** perform better than **HI+LHM** with increase of image noise level. While for the configuration of N = 10, the three approaches perform equally well when the noise level is less than 6 *pixels*. It can also be noticed that wrong solutions always appear at the edge of color maps, which indicates that pose ambiguity would have a larger probability to arise in the case where the angle between image plane and object plane is extremely large(close to 90°).

3) RATE OF WRONG SOLUTIONS

Fig. 6 compares the rate of wrong solutions for our algorithm, namely **MHI+LHM** and **HI+LHM**, with **SP+LHM**, which is one of the most robust and accurate iterative algorithms for the pose estimation from planar objects. Four cases are considered, among which are N = 4, object at the center of vision field; N = 4, object near the edge of vision field; N = 10, object at the center of vision field; and N = 10, object near the edge of vision field, as described in 1). At each noise level, the rate of wrong solutions is calculated from 2×17×17×200 results. Without loss of generality, a wrong solution is defined as a solution with the rotation error over 45°.

In Fig. 7, we include into some excellent state-of-art non-iterative PnP algorithms, such as **RPnP** [16], **OPnP**[17] and **DLS**+ + +[18]. Only the configuration of 10 coplanar points is considered in Fig. 7. From Fig. 6 and Fig. 7, We can

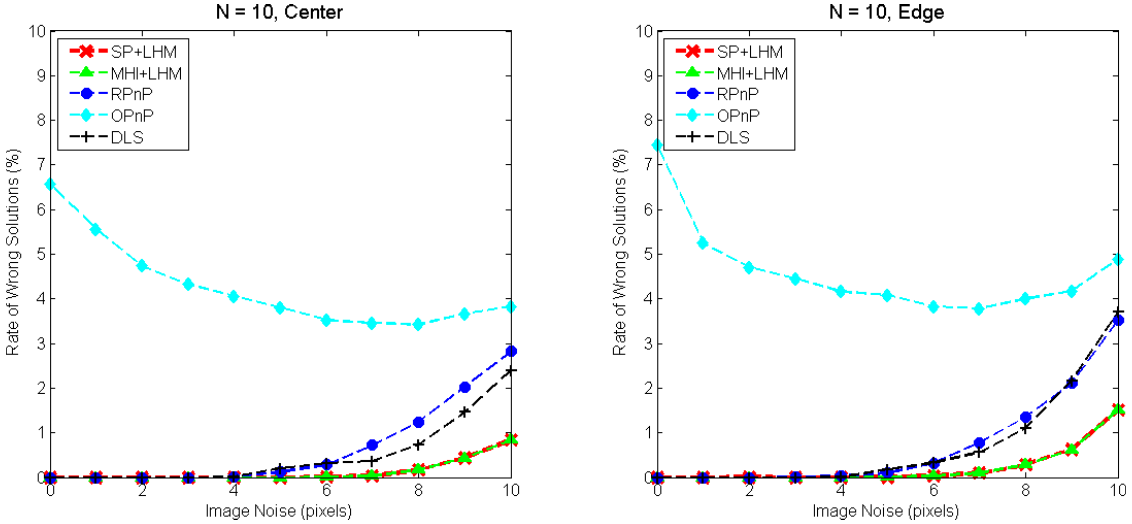


FIGURE 7. The rate of wrong solutions for N = 10.

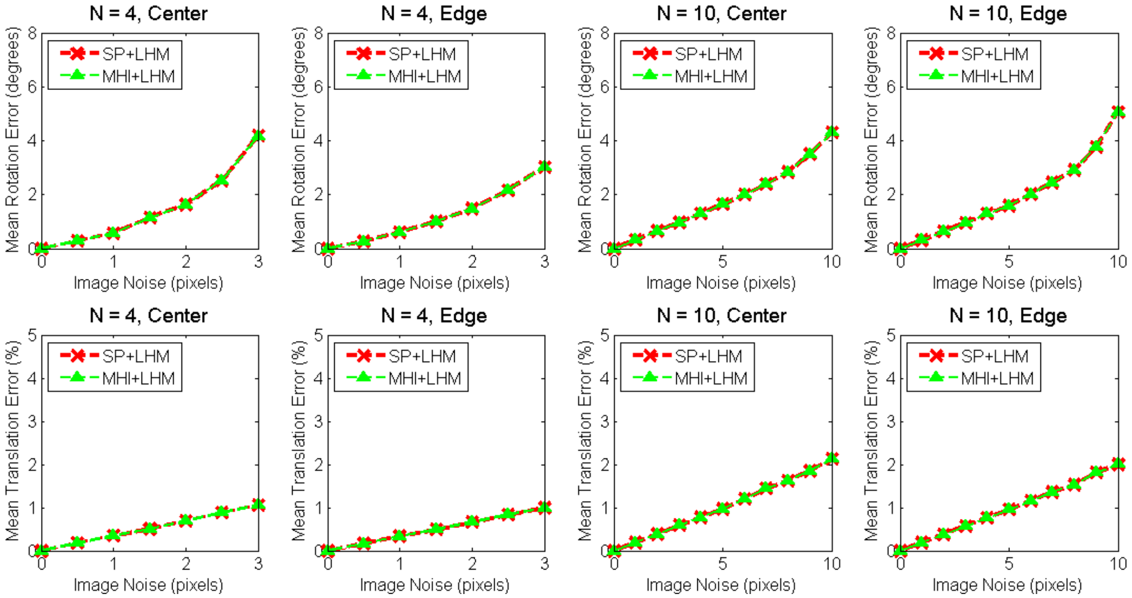


FIGURE 8. The mean rotation and translation error against noise.

observe that the robustness of **MHI+LHM** is identical with that of **SP+LHM** in all the four cases, and better than other PnP algorithms. It can also be noticed that, in the case of N=4, the rate of wrong solutions for **HI+LHM** increases more rapidly than the rate for **MHI+LHM** as image noise increases, while in the case of N = 10, the rates for **HI+LHM** and **MHI+LHM** keep at the same level. In Fig. 7, It is worth noting that **OPnP** always has a high rate of wrong solutions even under the noise free condition. From our observations, the solutions of **OPnP** tend to be unstable when α or β is close to zero.

4) MEAN ERROR AND COMPUTATIONAL TIME IN COMPARISON WITH THE SP METHOD

Fig. 8 shows the mean rotation error and mean translation error of **MHI+LHM** compared with those of **SP+LHM**. Fig. 9 compares average computational time of the two approaches. We run all codes in MATLAB on a desktop with 2.67GHz CPU and 6GB RAM.

It can be seen from Fig. 8 and Fig. 9 that **MHI+LHM** can achieve the accuracy similar to **SP+LHM**, while the average computational time is much less than that of **SP+LHM**. We can explain the computational efficiency of **MHI+LHM** in two aspects: (1): The second initial solution has a very concise expression. In theory, the computational cost required in obtaining the second initial solution is a negligible constant. By contrast, the **SP** method for calculating the second initial solution needs a series of coordinate transformations and a solution for a polynomial of degree four, which is more complicated. and (2): The first initial solution helps to improve the convergence speed for iterative algorithms to reach the first local minima, especially in the case of low noise. Moreover,

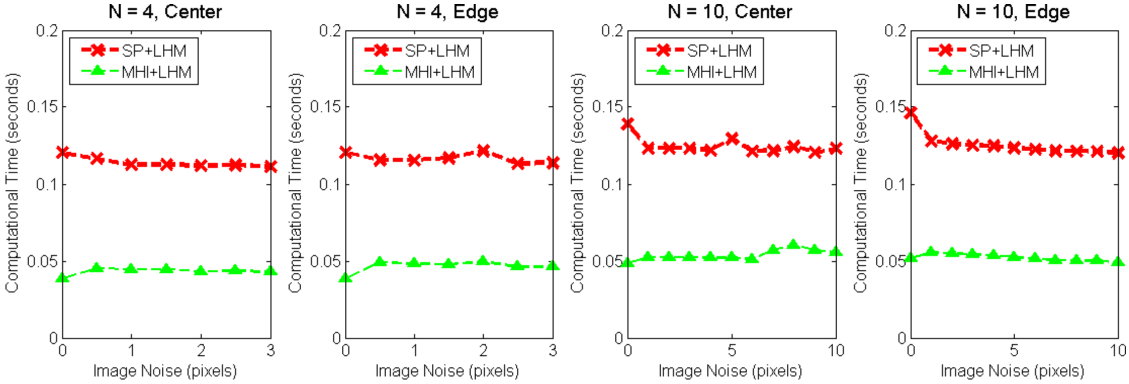


FIGURE 9. The average computational time.

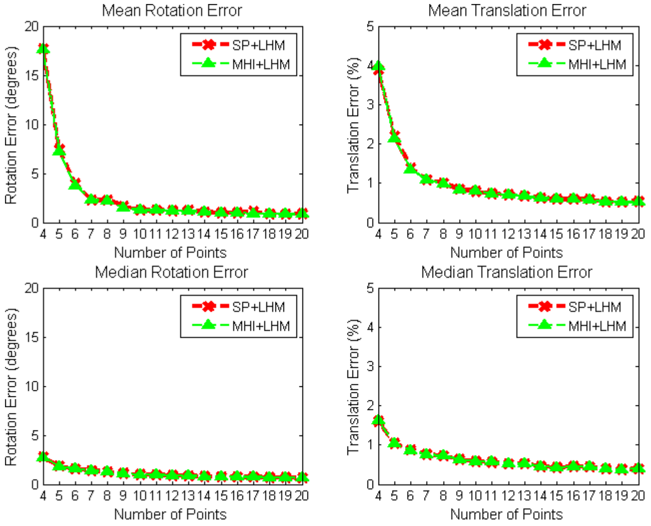


FIGURE 10. The mean and median errors against N.

the process for obtaining the first initial solution is approximately the process for solving a 2D homography in terms of time cost, which is computationally very fast.

5) MEAN AND MEDIAN ERRORS AGAINST NUMBER OF POINTS

A more general test has been made to evaluate the accuracy of our approach. In this test, the reference points are uniformly distributed in the range [−2000, 2000]×[−2000, 2000]×0 (*mm*) of object reference frame, while the rotation matrices are randomly generated and translation vectors are randomly chosen from the range [−500, 500]×[−500, 500]×[4000, 12000] (*mm*). Gaussian noise with $\sigma = 3$ *pixels* is added to the projected image points, and for each number of reference points, 1000 test data sets are generated. Fig. 10 shows the mean and median rotation errors and translation errors plotted against the number of reference points. We can see that the performance of **MHI+LHM** is always similar to **SP+LHM**, which further validates the two initial solutions we proposed.

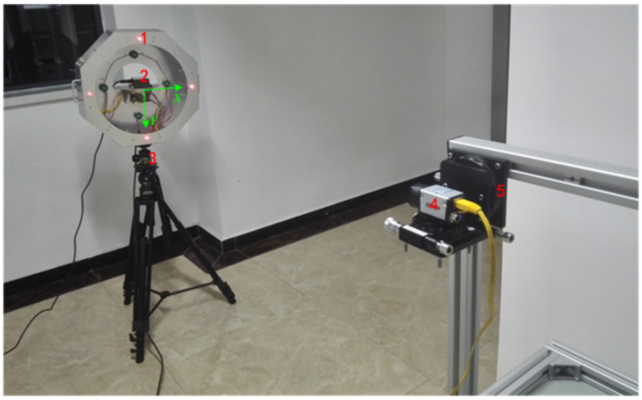


FIGURE 11. Experimental condition. (1) LED fixed on the disk target, (2) 2-axis inclinometer, (3) 3D rotation stage, (4) COMS camera, (5) 2D rotation stage for leveling.

B. EXPERIMENTS WITH REAL IMAGES

In this section, we use **MHI+LHM** to track the movement of a disk target. Four circular LEDs fixed on the disk target are used as the feature points to be extracted, which form the four corners of a square with diagonals of 440*mm* length. The object reference frame is built on the disk target with the origin located at the center of the square, as shown in Fig.11.

A 2-axis inclinometer with precision of ±0.01° and measurement range of ±30° is installed in the center of the disk target as a benchmark of rotation. A calibrated CMOS camera with a resolution of 1280×1024 and focal length of 6*mm* is applied in this experiment. Our algorithms run on a DSP platform connected to the camera. The calculated rotation matrix **R** is decomposed into three rotation angles, which are roll angle, pitching angle and yaw angle. The calculation results of DSP including rotation angles and translation vectors along with the outputs of inclinometer are sent to PC through RS-232 in real time.

We first adjust the attitude of the disk target and the COMS camera to the extent that the outputs of inclinometer as well as the three rotation angles calculated by DSP are close to 0°, then successively change the pitching angle and roll angle of

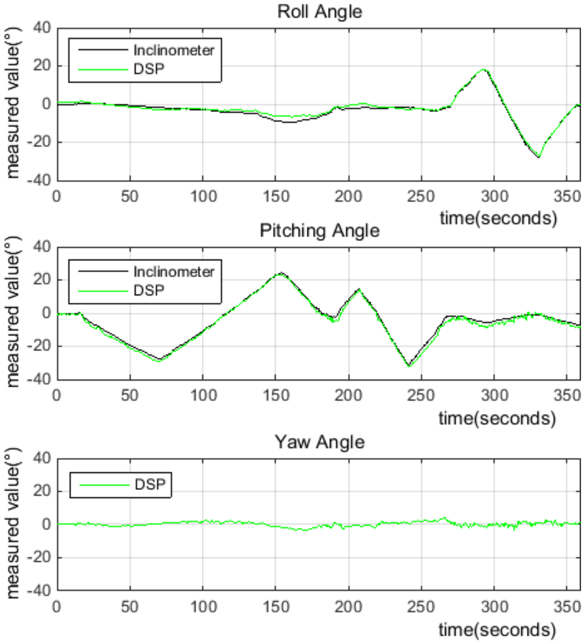


FIGURE 12. The calculation results of DSP for the three angles compared with the 2-axis inclinometer outputs.

the disk target, while recording the outputs of inclinometer and the calculation results of DSP at the same time. We try to ensure that when one of the three angles is being adjusted, the other two angles stay close to zero and keep unchanged. The recorded data is plotted in Fig. 12, which indicate that the DSP platform can stably output calculation results of the three rotation angles, among which the roll and pitching angles are comparable with the inclinometer outputs. It is also verified that when executing the algorithm on a hardware platform, the method for obtaining the two initial solutions is very computationally efficient.

VI. CONCLUSION

The robustness of iterative algorithms for PnP problems deeply relies on the choice of initial values. For planar case, we have presented two initial solutions that can be stably refined into the two poses that locate local minima. They both have simple forms that can be obtained effectively. The significance of the first initial solution mainly lies in two aspects: (1) it helps to improve the convergence speed for iterative algorithms to reach the first local minima; and (2) it can directly reach the global minima in the case of low image noise and large number of coplanar points. The second initial solution in this paper is proposed based on our new concept of "mirror" pose, which has a more concise form compared with the **SP** method, while the robustness of the corresponding final results is equally well. Besides their simple expressions, the two initial solutions proposed in this paper are not dependent on any specific iterative algorithm or error function. Those characteristics make them a good reference for choosing initial values in applications of iterative pose estimation where optimal accuracy and efficiency are required.

ACKNOWLEDGMENT

The authors would like to thank Bo Lv and Yin Guan for designing the experimental platform. They also appreciate Yu Cao and Yu Feng's contribution on DSP programming.

REFERENCES

[1] M. A. Fischler and R. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[2] R. J. Holt and A. N. Netravali, "Camera calibration problem: Some new results," *Comput. Vis. Graph. Image Process., Image Understand.*, vol. 54, no. 3, pp. 368–383, 1991.

[3] R. M. Haralick, C. Lee, K. Ottenburg, and M. Nolle, "Analysis and solutions of the three point perspective pose estimation problem," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Maui, HI, USA, Jun. 1991, pp. 592–598.

[4] W. J. Wolfe, D. Mathis, C. W. Sklair, and M. Magee, "The perspective view of three points," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 1, pp. 66–73, Jan. 1991.

[5] X.-S. Gao, X.-R. Hou, J. Tang, and H.-F. Cheng, "Complete solution classification for the perspective-three-point problem," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 8, pp. 930–943, Aug. 2003.

[6] L. Kneip, D. Scaramuzza, and R. Siegwart, "A novel parametrization of the perspective-three-point problem for a direct computation of absolute camera position and orientation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2011, pp. 2969–2976.

[7] Y. Hung, P.-S. Yeh, and D. Harwood, "Passive ranging to known planar point sets," in *Proc. IEEE Int. Conf. Robot. Autom.*, Mar. 1985, pp. 80–85.

[8] M. A. Abidi and T. Chandra, "A new efficient and direct solution for pose estimation using quadrangular targets: Algorithm and evaluation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 17, no. 5, pp. 534–538, May 1995.

[9] R. M. Haralick, "Determining camera parameters from the perspective projection of a rectangle," *Pattern Recognit.*, vol. 22, no. 3, pp. 225–230, 1989.

[10] J. S.-C. Yuan, "A general photogrammetric method for determining object position and orientation," *IEEE Trans. Robot. Autom.*, vol. 5, no. 2, pp. 129–142, Apr. 1989.

[11] L. Quan and Z. Lan, "Linear N-point camera pose determination," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 8, pp. 774–780, Aug. 1999.

[12] B. Triggs, "Camera pose and calibration from 4 or 5 known 3D points," in *Proc. IEEE Int. Conf. Comput. Vis.*, Sep. 1999, pp. 278–284.

[13] P. D. Fiore, "Efficient linear solution of exterior orientation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 2, pp. 140–148, Feb. 2001.

[14] A. Ansar and K. Daniilidis, "Linear pose estimation from points or lines," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 5, pp. 578–589, May 2003.

[15] V. Lepetit, F. Moreno-Noguer, and P. Fua, "EP*n*P: An accurate O(*n*) solution to the P*n*P problem," *Int. J. Comput. Vis.*, vol. 81, no. 2, pp. 155–166, 2008.

[16] S. Li, C. Xu, and M. Xie, "A robust O(n) solution to the perspective-n-point problem," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 7, pp. 1444–1450, Jul. 2012.

[17] Y. Zheng, Y. Kuang, S. Sugimoto, K. Åström, and M. Okutomi, "Revisiting the PnP problem: A fast, general and optimal solution," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 2344–2351.

[18] J. A. Hesch and S. I. Roumeliotis, "A Direct Least-Squares (DLS) method for PnP," in *Proc. IEEE Int. Conf. Comput. Vis.*, Nov. 2011, pp. 383–390.

[19] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2003, pp. 88–131.

[20] C.-P. Lu, G. D. Hager, and E. Mjolsness, "Fast and globally convergent pose estimation from video images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 6, pp. 610–622, Jun. 2000.

[21] D. Oberkampf, D. F. DeMenthon, and L. S. Davis, "Iterative pose estimation using coplanar feature points," *Comput. Vis. Image Understand.*, vol. 63, no. 3, pp. 495–511, May 1996.

[22] G. Schweighofer and A. Pinz, "Robust pose estimation from a planar target," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 12, pp. 2024–2030, Dec. 2006.

[23] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 11, pp. 1330–1334, Nov. 2000.

KAI ZHOU received the M.S. degree in control theory and control engineering from the Shandong University of Science and Technology, Qingdao, China, in 2013.

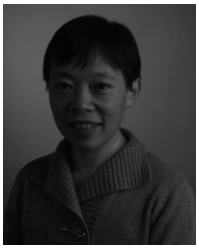
He is currently pursuing the Ph.D. degree with the College of Precision Instrument and Opto-Electronics Engineering, Tianjin University, Tianjin, China. His research interests include photogrammetry and 3-D reconstruction.

XIANGJUN WANG received the B.S., M.S., and Ph.D. degrees in precision measurement technology and instruments from Tianjin University, Tianjin, China, in 1980, 1985, and 1990, respectively.

He is currently a Professor and the Director of the Precision Measurement System Research Group with Tianjin University. His research interests include photoelectric sensors and testing, computer vision, image analysis, MOEMS, and MEMS.

ZHONG WANG received the M.S. degree in instrumentation science and technology from Tianjin University, Tianjin, China, in 1991.

He is currently a Professor with the College of Precision Instrument and Opto-Electronics Engineering, Tianjin University. His current research interests include laser and photoelectric measurement and in situ vision inspection.



HONG WEI received the Ph.D. degree from Birmingham University, U.K., in 1996. Then she was a Post-Doctoral Research Assistant on a Hewlett Packard sponsored project, high-resolution CMOS camera systems. She was also a Research Fellow on an EPSRC-funded Faraday project, model from movies.

She joined the University of Reading in 2000. Her current research interest includes intelligent computer vision and its applications in remotely sensed images, and face recognition (biometric).

LEI YIN received the M.S. degree in mechatronic engineering from the Changchun University of Science and Technology, Changchun, China, in 2014.

He is currently pursuing the Ph.D. degree with the College of Precision Instrument and Opto-Electronics Engineering, Tianjin University, Tianjin, China. His research interests include computer vision and 3-D reconstruction.