Effective Self-calibration for Camera Parameters and Hand-eye Geometry Based on Two Feature Points Motions

Jia Sun, Peng Wang, *Member, IEEE*, Zhengke Qin, and Hong Qiao, *Senior Member, IEEE*

***Abstract*—A novel and effective self-calibration approach for robot vision is presented, which can effectively estimate both the camera intrinsic parameters and the hand-eye transformation at the same time. The proposed calibration procedure is based on two arbitrary feature points of the environment, and three pure translational motions and two rotational motions of robot end-effector are needed. New linear solution equations are deduced, and the calibration parameters are finally solved accurately and effectively. The proposed algorithm has been verified by simulated data with different noise and disturbance. Because of the need of fewer feature points and robot motions, the proposed method greatly improves the efficiency and practicality of the calibration procedure.**

***Index Terms*—Camera calibration, hand-eye calibration, robot vision, two feature points.**

I. Introduction

In the area of robot vision, camera is usually used as visual sensor to aid the robot in performing specific tasks [1]−[3], such as grasping and assembling objects. The accuracy of robot vision system mainly depends on the results of calibration, including the camera calibration and hand-eye calibration. In this paper, we will focus on the eye-in-hand robot vision system that camera is attached to the end-effector of a robot. In this case, the intrinsic camera parameters (principal point, focal length and the aspect ratio) and the hand-eye transformation (the orientation and position of the camera in relation to the robot end-effector) should be calculated first.

Much work has been done in the research of camera calibration techniques [4]−[6]. These techniques can be classified mainly into two categories: traditional camera calibration (photogrammetric calibration) and self-calibration. Traditional methods applied for calibration use a set of 3D to 2D correspondences extracted with the help of a calibration object of known structure. According to different structures, the calibration objects can be divided into three categories: 3D calibration objects, 2D calibration objects and 1D calibration objects. 3D calibration object is a block with precisely known metric structure, often with some sort of grid pattern [7]. Using 3D calibration object, camera parameters with high accuracy can be obtained. 2D calibration object is usually a board with checker pattern, and calibration method with 2D calibration object is more flexible and easier to implement [8]−[11]. 1D object needs at least three collinear points with known relative positions [12]−[14]. By observing the certain motion of the object such as moving around a fixed point, the calibration can be performed.

However, these methods are not practical in some physical environments, especially in some robotic application systems. In these cases, the camera and hand-eye parameters are needed to calibrate frequently in a closed environment. Therefore, self-calibration techniques which estimate camera parameters from information of the static image scenes without any special calibration object have been proposed. These methods take advantage of geometric invariants of some image features, such as straight lines, vanishing points and circles. The earliest research on the self-calibration is based on the so-called Kruppa equations, which links the camera intrinsic parameters with the epi-polar geometry of pairs of views taken by the camera [15]. Neither the camera motion information nor the structure information of the environment is used in this method. Some methods use sets of parallel line segments extracted from the views of environment to obtain two or more mutually orthogonal vanishing points [16]. Applying the orthogonality constraints imposed by these vanishing points, the calibration can be performed. These methods are effective and practical in special applications. For example, the vanishing points extracted from the features of the roads and traffic signs are used to calibrate a traffic camera [17]. These methods are flexible but yield less accurate results. Besides, self-calibration based on active vision system is also a hot research direction in this field. These methods require the camera undergoing a series of special motions, such as translational motions along arbitrary three dimensional directions, and calculate the camera parameters via the motion information and the feature points captured by the camera [18].

Concerning hand-eye calibration, various methods have been reported. In many literatures [19], [20], the hand-eye relation is determined by activating a sequence of robot motions and simultaneously measuring the induced camera

Manuscript received July 4, 2016; accepted January 16, 2017. This work was partly supported by the National Natural Science Foundation of China (61379097, 61401463, 61100098) and Youth Innovation Promotion Association CAS. Recommended by Associate Edlitor Jian Yang.

Citation: J. Sun, P. Wang, Z. K. Qin, and H. Qiao, "Effective self-calibration for camera parameters and hand-eye geometry based on two feature points motions," *IEEE/CAA Journal of Automatica Sinica*, vol. 4, no. 2, pp. 370−380, Apr. 2017.

J. Sun, P. Wang, Z. K. Qin and H. Qiao are with the Institute of Automation, Chinese Academy of Sciences (CASIA), Beijing 100190, China; J.Sun and Z. K. Qin are also with the University of Chinese Academy of Sciences, Beijing 100190, China (e-mail: jia.sun@ia.ac.cn; peng wang@ia.ac.cn; qinzhengke2012@ia.ac.cn; hong.qiao@ia.ac.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JAS.2017.7510556

TABLE I
COMPARISON OF METHODS FOR CALIBRATION

Method		Intrinsic parameters	Hand-eye geometry	Calibration object[a]	Motion times[b]	Feature points[c]
Traditional calibration	Zhang [8]	Yes	No	Yes	2	Yes
	Andreff *et al.* [20]	No	Yes	Yes	3	Yes
	Malm *et al.* [22]	Yes	Yes	Yes	4	Yes
Self-calibration	Ma [18]	Yes	Yes	No	7	No
	Wang *et al.* [25]	Yes	Yes	No	6	No
	Lei *et al.* [26]	Yes	Yes	No	6	No
Method proposed in this article		Yes	Yes	No	5	Yes

[a] Calibration object: If the method needs a calibration object to achieve the calibration;

[b] Motion times: How many robot motions or pictures does the method need to achieve the calibration;

[c] Feature points: If the method calibrates cameras using feature points directly.

motion by observing a calibration object. They adopt different motion representations with the similarly mathematical algorithms. In order to achieve hand-eye calibration in special environment, the methods without reference object have also emerged. A hand-eye calibration method using only image derivatives instead of point correspondences is proposed in [21]. In that work, the robot end-effector with camera is controlled to make two translational motions and two normal motions, and the motion of the camera is recovered using image derivatives. However, all these mentioned hand-eye calibration methods assumed that the intrinsic camera parameters have been calculated first, i.e., the camera parameters and hand-eye geometry are calculated respectively.

For a robot vision system, both camera calibration and hand-eye calibration are required, and approaches combining these two steps are researched. Malm *et al.* [22] fitted a hand-eye calibration algorithm and a plane-based intrinsic camera calibration into one framework to simplify the motions of the robot end-effector. Zhao *et al.* [23] proposed a joint algorithm that combines the two calibration processes together by moving the robot and observing a 2D reference object. The method proposed by Strobl *et al.* [24] relaxes the requirement of the accurate knowledge of the imprinted pattern dimensions, while still uses the planarity and regularity of the pattern. Ma [18] presented a self-calibration approach which performs two calibration steps simultaneously by applying the focus of expansion (FOE) and specially designed camera motions. Wang *et al.* [25] and Lei *et al.* [26] apply FOEs to calculate both the intrinsic parameters and hand-eye geometry, but require six robot motions. Specifically speaking, four times robot translations, of which no any three translations are co-planar, are needed to calculate the camera's intrinsic parameters and the camera's orientation matrix. Additionally, two more robot rotations are needed to calculate the position vector.

So far, the robot vision system calibration has been deeply researched. Table I shows the main properties of the aforementioned methods. There are many problems need to be solved. Firstly, the traditional methods [8], [20], [22] using calibration objects can obtain the parameters accurately. However, they are not available in some practical environments. Camera self-calibration based on Kruppa equations and vanishing points cannot satisfy high precision demand. Secondly, if the camera parameters and hand-eye geometry are calibrated respectively with different self-calibration methods, it will cause much large accumulative error. Similarly, methods applying FOEs to calibrate camera will increase accumulative error as well [18], [25], [26]. Thirdly, present combined approaches need plenty of camera motions and feature points, such as at least seven camera motions [18] and six camera motions [25], [26]. Therefore, an effective and practical combined self-calibration approach is urgently required for robot vision.

In this paper, a new self-calibration approach that can effectively estimate both the camera intrinsic parameters and the hand-eye transformation for robot vision is presented. Taking advantage of active vision system, only two arbitrary feature points of the environment and five robot motions are needed. Because the proposed method applies fewer feature points and robot motions, the efficiency of the calibration procedure is greatly improved.

New linear solution equations are deduced to calculate the calibration parameters. The proposed method has been tested by computer simulated data with different noise and disturbance. The results demonstrate the effectiveness and robustness of the proposed method.

The contributions of this paper can be summarized as following: Firstly, the procedures of camera calibration and hand-eye calibration are combined. The efficiency and practicality are greatly improved, and the accumulative errors are reduced simultaneously. Secondly, fewer feather points and robot motions are applied. Because the procedure of feature recognition and matching are simplified, the calibration efficiency is further enhanced. Thirdly, some new linear solution equations are deduced, and the calibration parameters are finally solved accurately and effectively.

The paper is organized as follows. Section II introduces some preliminaries such as the camera model and the robot vision system. In Section III, the proposed solution equations are derived to perform camera calibration and hand-eye self-calibration based on two feature points. The procedure of this calibration algorithm is also described here. The results of experiments are presented in Section IV and the paper is concluded in Section V.

II. PRELIMINARIES

As shown in Fig. 1, in a robot hand-eye system, the camera is rigidly mounted on the robot end-effector. The robot end-

effector can make rotational and translational motions and the motion parameters can be read from the controller.

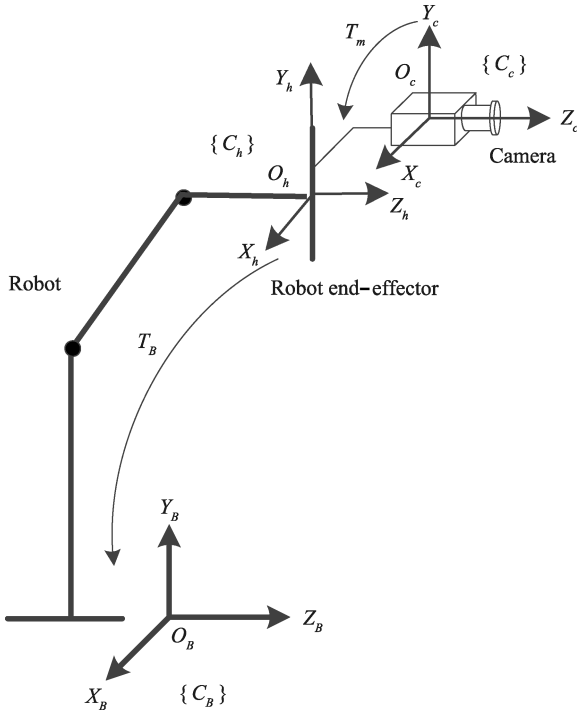

Fig. 1. Hand-eye relation in robot vision system.

There are three coordinate systems: the robot base coordinate system, the robot end-effector coordinate system and the camera coordinate system, which are denoted by $\{C_B\}$, $\{C_h\}$ and $\{C_c\}$, respectively (Fig. 1). The motion parameters read from controller are the transformation of $\{C_h\}$ in relation to $\{C_B\}$, which is described by the matrix T_B. The hand-eye relation can be denoted by a rotation matrix R_m and a translation vector p_m, where R_m represents the orientation of the camera, and p_m represents the position of the camera relative to $\{C_h\}$. Therefore, we can represent the hand-eye relation by the form

$$X_{ca} = R_m X_h + p_m \tag{1}$$

where $X_{\rm ca} = (x_c,\ y_c,\ z_c)^T$ and $X_h = (x_h,\ y_h,\ z_h)^T$ are the coordinates of a point described by $\{C_c\}$ and $\{C_h\}$, respectively.

Using the pinhole camera model, the transformation between the 2D coordinate of the image point $\boldsymbol{m} = (u,\ v,\ 1)^T$ and the 3D coordinate of the space point in $\{C_c\}$, $\boldsymbol{X}_{ca} = (x_c,\ y_c,\ z_c)^T$ can be described as

$$z_c \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{pmatrix} f_u & s & u_0 \\ 0 & f_v & v_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_c \\ y_c \\ z_c \end{pmatrix} = K \begin{pmatrix} x_c \\ y_c \\ z_c \end{pmatrix} \tag{2}$$

where $(u_0,\ v_0)$ is the principal point, f_u and f_v are scale factor of the image plane in the horizontal and vertical direction, s is skew factor of pixel, z_c is the depth value of the point $\boldsymbol{X}_{ca}$. The matrix K is called intrinsic camera matrix. Then, (2) can be written compactly as

$$z_c \boldsymbol{m} = K \boldsymbol{X}_{ca}. \tag{3}$$

In this paper, we propose a new calibration method which determines both the hand-eye relation (R_m and $\boldsymbol{p}_m$) and the camera intrinsic parameters u_0, v_0, f_u , f_v and s simultaneously.

III. Self-calibration Combining Camera and Hand-eye Calibration Based on Two Feature Points

Because of the limitation of calibration method based on reference objects in some practical environments, self-calibration appears particularly significant for hand-eye robot system. The known motion parameters of robot end-effector which are read from the controller provide important information for self-calibration in the active vision system. The method using these motion parameters can realize self-calibration effectively, and avoid the high-dimensional nonlinear estimation problem. Furthermore, in order to improve the efficiency, the calibration procedure should be implemented with fewer feather points and robot motions.

Considering all of the above, we present a novel self-calibration method combining camera and hand-eye calibration based on two points for robot vision in this paper. The proposed method is applied for camera and hand-eye calibration in one procedure, which can reduce the calibration errors and simplify the calibration procedure. To improve the efficiency, the proposed calibration procedure applies two arbitrary feature points of the environment, and three pure translational motions and two rotational motions of robot end-effector are required.

The calibration of camera intrinsic matrix K and rotation matrix R_m will be given in Section III-A. The calibration of translation vector $\boldsymbol{p}_m$ will be given in Section III-B. In addition, there is a requirement for the initial position of the camera in this method. Suppose P_1 and P_2 are two feature points in practical environment. Then, the image plane of the camera must be parallel to the line consisting of points P_1 and P_2, i.e., the depth values of the two points in $\{C_c\}$ are equal, as shown in Fig. 2.

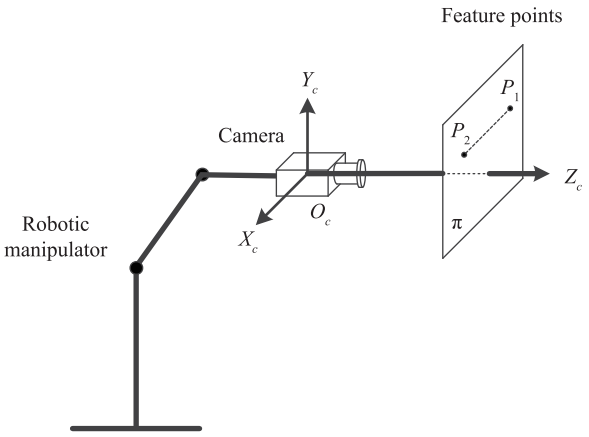

Fig. 2. The initial relative position of camera and feature points.

A. Calibration of Camera Intrinsic Matrix K and Rotation Matrix R_m

In this subsection, the camera intrinsic matrix K and the rotation matrix R_m through the three pure translational mo-

tions of robot end-effector will be calculated. Suppose $\boldsymbol{X}_{c01}$ and $\boldsymbol{X}_{h01}$ are the initial coordinates of point P_1 in $\{C_c\}$ and $\{C_h\}$, respectively. From (1) we have

$$\boldsymbol{X}_{c01} = R_m \boldsymbol{X}_{h01} + \boldsymbol{p}_m. \tag{4}$$

Let $\boldsymbol{X}_{c11}, \boldsymbol{X}_{h11}$ represent the coordinates of point P_1 in $\{C_c\}$ and $\{C_h\}$ after the motion, respectively. Then, we have

$$\boldsymbol{X}_{c11} = R_m \boldsymbol{X}_{h11} + \boldsymbol{p}_m. \tag{5}$$

Since the camera is attached rigidly to the robot end-effector, the camera motions are the same as the robot end-effector motions. The translation vector $\boldsymbol{p}_m$ in (4) and (5) will be eliminated when the motion of robot end-effector will be only a pure translational one. Therefore, we control the robot end-effector undergoing a pure translational motion described as $\boldsymbol{b}_1$, and the motion of the camera can be expressed as

$$\boldsymbol{X}_{c11} - \boldsymbol{X}_{c01} = R_m \boldsymbol{b}_1. \tag{6}$$

Let $\boldsymbol{m}_{01} = [u_{01},\ v_{01},\ 1]^T$ and $\boldsymbol{m}_{11} = [u_{11},\ v_{11},\ 1]^T$ be, respectively, the image coordinates of P_1 before and after the translational motion $\boldsymbol{b}_1$, and z_{c01}, z_{c11} denote the depth value of P_1 before and after the translational motion in $\{C_c\}$. From (3) we have

$$\begin{cases} z_{c01}\boldsymbol{m}_{01} = K\boldsymbol{X}_{c01} \\ z_{c11}\boldsymbol{m}_{11} = K\boldsymbol{X}_{c11}. \end{cases} \tag{7}$$

Substituting (7) into (6) yields

$$z_{c11}\boldsymbol{m}_{11} - z_{c01}\boldsymbol{m}_{01} = KR_m\boldsymbol{b}_1. \tag{8}$$

Let $M = KR_m$, and denote the ith row of the intermediate matrix M by $\boldsymbol{m}_i$. From (8) we have

$$\begin{cases} z_{c11}u_{11} - z_{c01}u_{01} = \boldsymbol{m}_1\boldsymbol{b}_1 \\ z_{c11}v_{11} - z_{c01}v_{01} = \boldsymbol{m}_2\boldsymbol{b}_1 \\ z_{c11} - z_{c01} = \boldsymbol{m}_3\boldsymbol{b}_1. \end{cases} \tag{9}$$

Eliminating the unknown depth value z_{c11} in (9), we can obtain two linear constraints on M:

$$\begin{cases} \boldsymbol{m}_1\boldsymbol{b}_1 - u_{11}\boldsymbol{m}_3\boldsymbol{b}_1 = z_{c01}(u_{11} - u_{01}) \\ \boldsymbol{m}_2\boldsymbol{b}_1 - v_{11}\boldsymbol{m}_3\boldsymbol{b}_1 = z_{c01}(v_{11} - v_{01}). \end{cases} \tag{10}$$

Similarly, we can gain the same form constraints for feature point P_2:

$$\begin{cases} \boldsymbol{m}_1\boldsymbol{b}_1 - u_{12}\boldsymbol{m}_3\boldsymbol{b}_1 = z_{c02}(u_{12} - u_{02}) \\ \boldsymbol{m}_2\boldsymbol{b}_1 - v_{12}\boldsymbol{m}_3\boldsymbol{b}_1 = z_{c02}(v_{12} - v_{02}). \end{cases} \tag{11}$$

As the assumption we proposed at the beginning of this section, in order to construct solvable linear equations, the depth value of the two points in $\{C_c\}$ should be equal. That means $z_{c01} = z_{c02}$. Combining (10) and (11), we have

$$\begin{cases} \dfrac{1}{z_{c01}}\boldsymbol{m}_1\boldsymbol{b}_1 - u_{11}\dfrac{1}{z_{c01}}\boldsymbol{m}_3\boldsymbol{b}_1 = (u_{11} - u_{01}) \\ \dfrac{1}{z_{c01}}\boldsymbol{m}_2\boldsymbol{b}_1 - v_{11}\dfrac{1}{z_{c01}}\boldsymbol{m}_3\boldsymbol{b}_1 = (v_{11} - v_{01}) \\ \dfrac{1}{z_{c01}}\boldsymbol{m}_1\boldsymbol{b}_1 - u_{12}\dfrac{1}{z_{c01}}\boldsymbol{m}_3\boldsymbol{b}_1 = (u_{12} - u_{02}) \\ \dfrac{1}{z_{c01}}\boldsymbol{m}_2\boldsymbol{b}_1 - v_{12}\dfrac{1}{z_{c01}}\boldsymbol{m}_3\boldsymbol{b}_1 = (v_{12} - v_{02}). \end{cases} \tag{12}$$

Let $M' = M/z_{c01}$ which is a 3×3 matrix. There are 9 unknown parameters in (12), and four linear equations are built from one pure translational motion. That is, N translational motions will provide $\mathrm{N}\times4$ equations, and the unknown parameters can be solved when $\mathrm{N}\times4 > 9$. Therefore, at least three pure translational motions are required. Hence, we control the robot end-effector to move along $\boldsymbol{b}_2$ and $\boldsymbol{b}_3$ which are also pure translational motions but different from the previous directions. Then, we will obtain twelve linear equations, denoted in matrix form as:

$$A\boldsymbol{X} = \boldsymbol{B} \tag{13}$$

where

$$A = \begin{bmatrix} \boldsymbol{b}_1^T & \boldsymbol{0} & -u_{11}\boldsymbol{b}_1^T \\ \boldsymbol{0} & \boldsymbol{b}_1^T & -v_{11}\boldsymbol{b}_1^T \\ \boldsymbol{b}_1^T & \boldsymbol{0} & -u_{12}\boldsymbol{b}_1^T \\ \boldsymbol{0} & \boldsymbol{b}_1^T & -v_{12}\boldsymbol{b}_1^T \\ \vdots & \vdots & \vdots \\ \boldsymbol{b}_3^T & \boldsymbol{0} & -u_{31}\boldsymbol{b}_3^T \\ \boldsymbol{0} & \boldsymbol{b}_3^T & -v_{31}\boldsymbol{b}_3^T \\ \boldsymbol{b}_3^T & \boldsymbol{0} & -u_{32}\boldsymbol{b}_3^T \\ \boldsymbol{0} & \boldsymbol{b}_3^T & -v_{32}\boldsymbol{b}_3^T \end{bmatrix}$$

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{m}_1' & \boldsymbol{m}_2' & \boldsymbol{m}_3' \end{bmatrix}^T$$

$$\boldsymbol{B} = \begin{bmatrix} u_{11} - u_{01} \\ v_{11} - v_{01} \\ u_{12} - u_{02} \\ v_{12} - v_{02} \\ \vdots \\ u_{31} - u_{01} \\ v_{31} - v_{01} \\ u_{32} - u_{02} \\ v_{32} - v_{02} \end{bmatrix}$$

and vector $\boldsymbol{0} = [0,\ 0,\ 0]^T$.

From the linear equations above, matrix M' can be solved by least square algorithm (LSA).

In order to obtain the unique solution of (13), there are some restrictions on the translational motions of the robot end-effector. The cases in which additional motions do not provide more constraints on the camera intrinsic parameters and hand-eye translation matrix are as following. 1) One of the translational motions of the robot end-effector is along any coordinate axis described in $\{C_c\}$. 2) Two of three translational motions are parallel with any camera coordinate planes at the same time. 3) These three translational motions are moved along the same direction. In practice, it is very easy to avoid the degenerate configurations mentioned above.

Once the intermediate matrix M' is obtained, we start to calculate the camera intrinsic matrix K and the rotation matrix R_m. From the equation $M = KR_m = z_{c01}M'$, we have

$$\begin{bmatrix} \boldsymbol{m}_1 \\ \boldsymbol{m}_2 \\ \boldsymbol{m}_3 \end{bmatrix} = \begin{bmatrix} f_u & s & u_0 \\ 0 & f_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{r}_1 \\ \boldsymbol{r}_2 \\ \boldsymbol{r}_3 \end{bmatrix} = z_{c01} \begin{bmatrix} \boldsymbol{m}_1' \\ \boldsymbol{m}_2' \\ \boldsymbol{m}_3' \end{bmatrix} \tag{14}$$

where $\boldsymbol{m}_i'$ and $\boldsymbol{r}_i$ are the ith row of the matrix M' and R. Expanding the (14) we have

$$\begin{cases} \boldsymbol{m}_1 = f_u\boldsymbol{r}_1 + s\boldsymbol{r}_2 + u_0\boldsymbol{r}_3 = z_{c01}\boldsymbol{m}_1' \\ \boldsymbol{m}_2 = f_v\boldsymbol{r}_2 + v_0\boldsymbol{r}_3 = z_{c01}\boldsymbol{m}_2' \\ \boldsymbol{m}_3 = \boldsymbol{r}_3 = z_{c01}\boldsymbol{m}_3'. \end{cases} \tag{15}$$

R_m is a unit orthogonal matrix. Therefore, its row vector $\boldsymbol{r}_3$ is a unit vector, whose 2-norm is 1. From the third equation of (15), we have

$$z_{c01} = \frac{1}{|\boldsymbol{m}_3'|}, \quad M = z_{c01}M'. \tag{16}$$

Since the rotation matrix R_m is a unit orthogonal matrix, we take the dot-product or cross-product of both sides of (15) with $\boldsymbol{r}_i$, the intrinsic camera parameters and the rotation matrix can be calculated as

$$\begin{aligned} &\boldsymbol{r}_3 = \boldsymbol{m}_3 \\ &v_0 = \boldsymbol{m}_2\boldsymbol{r}_3^T \\ &u_0 = \boldsymbol{m}_1\boldsymbol{r}_3^T \\ &f_v = |\boldsymbol{m}_2 \times \boldsymbol{r}_3| \\ &\boldsymbol{r}_1 = \frac{\boldsymbol{m}_2 \times \boldsymbol{r}_3}{|\boldsymbol{r}_2 \times \boldsymbol{r}_3|} \\ &\boldsymbol{r}_2 = \boldsymbol{r}_3 \times \boldsymbol{r}_1 \\ &f_u = \boldsymbol{m}_1\boldsymbol{r}_1^T \\ &s = \boldsymbol{m}_1\boldsymbol{r}_2^T. \end{aligned} \tag{17}$$

B. Calibration of Translation Vectors $\boldsymbol{p}_m$

After solving the camera intrinsic parameters and the orientation of the camera, in this section, we will calibrate the last unknown vector $\boldsymbol{p}_m$. From (6), the translation vector $\boldsymbol{p}_m$ will be eliminated when the robot moves with pure translational motion. This implies that $\boldsymbol{p}_m$ cannot be solved from pure translational motion, and the rotational motions of the robot end-effector should be involved to calculate the translation vector $\boldsymbol{p}_m$.

Suppose a pure rotational motion of robot end-effector is described by R_4. Let $\boldsymbol{X}_{c41}$, $\boldsymbol{X}_{h41}$ represent the coordinates of point P_1 described in $\{C_c\}$ and $\{C_h\}$ after the rotational motion, and then we have

$$\boldsymbol{X}_{h41} = R_4\boldsymbol{X}_{h01} \tag{18}$$

$$\boldsymbol{X}_{c41} = R_m\boldsymbol{X}_{h41} + \boldsymbol{p}_m. \tag{19}$$

Substituting (1), (3) and (18) into (19) yields

$$\begin{aligned} &K\left(R_mR_4R_m^{-1} - E\right)\boldsymbol{p}_m + z_{c41}\boldsymbol{m}_{41} \\ &\quad = z_{c01}KR_mR_4R_m^{-1}K^{-1}\boldsymbol{m}_{01}. \end{aligned} \tag{20}$$

Similarly, we gain the same form constraints for feature point P_2

$$\begin{aligned} &K\left(R_mR_4R_m^{-1} - E\right)\boldsymbol{p}_m + z_{c42}\boldsymbol{m}_{42} \\ &\quad = z_{c02}KR_mR_4R_m^{-1}K^{-1}\boldsymbol{m}_{02} \end{aligned} \tag{21}$$

where, E is unit matrix.

As the assumption $z_{c01} = z_{c02}$, we combine (20) and (21) into a matrix form

$$\begin{aligned} &\begin{bmatrix} K\left(R_mR_4R_m^{-1} - E\right) & \boldsymbol{m}_{41} & 0 \\ K\left(R_mR_4R_m^{-1} - E\right) & 0 & \boldsymbol{m}_{42} \end{bmatrix} \begin{bmatrix} \boldsymbol{p}_m \\ z_{c41} \\ z_{c42} \end{bmatrix} \\ &\quad = \begin{bmatrix} z_{c01}KR_mR_4R_m^{-1}K^{-1}\boldsymbol{m}_{01} \\ z_{c01}KR_mR_4R_m^{-1}K^{-1}\boldsymbol{m}_{02} \end{bmatrix}. \end{aligned} \tag{22}$$

For these five unknown parameters equations, we can solve them by twice pure rotational motions of robot end-effector.

Actually, general motions including rotational motions and translational motions of robot end-effector can obtain the vector $\boldsymbol{p}_m$ as well. Suppose the fourth motion is a general motion described by a pure rotational motion R_4 and a pure translational motion b_4, then (22) will be

$$\begin{aligned} &\begin{bmatrix} K\left(R_mR_4R_m^{-1} - E\right) & \boldsymbol{m}_{41} & 0 \\ K\left(R_mR_4R_m^{-1} - E\right) & 0 & \boldsymbol{m}_{42} \end{bmatrix} \begin{bmatrix} \boldsymbol{p}_m \\ z_{c41} \\ z_{c42} \end{bmatrix} \\ &\quad = \begin{bmatrix} z_{c01}KR_mR_4R_m^{-1}K^{-1}\boldsymbol{m}_{01} + KR_m\boldsymbol{b}_4 \\ z_{c01}KR_mR_4R_m^{-1}K^{-1}\boldsymbol{m}_{02} + KR_m\boldsymbol{b}_4 \end{bmatrix}. \end{aligned} \tag{23}$$

And its solution is the same as of (22).

C. Summary of the Proposed Calibration Procedure

Combining the two calibration sub-processes proposed in Section III-A and Section III-B, the procedure of the proposed self-calibration method can be summarized as follow:

Step 1: Control the robot end-effector attached with the camera to arrive the initial position as described in Section III. Take the first image. Extract the image coordinates of two feature points: $\boldsymbol{m}_{01}$ and $\boldsymbol{m}_{02}$.

Step 2: Take images at different positions of the robot end-effector, meanwhile read the robot motion parameters $\boldsymbol{b}_i$ (i = 1, 2, 3) and R_j ($j = i + 1, i + 2$). Extract the image coordinates of two feature points after corresponding robot motions: $\boldsymbol{m}_{i1}$, $\boldsymbol{m}_{i2}$, $\boldsymbol{m}_{j1}$ and $\boldsymbol{m}_{j2}$.

Step 3: Compute the intermediate matrix M (see (12) and (14)). Estimate the camera intrinsic matrix K and the rotation matrix R_m.

Step 4: From camera rotation motions R_j and the parameters solved before, compute the camera translation vector $\boldsymbol{p}_m$ (see (22)).

IV. Experiments

For a calibration method, the simplicity of calibration procedure and the accuracy of calibration results are two important aspects of evolution. In the aspect of simplicity, the proposed algorithm calibrates the camera and hand-eye parameters at the same time, and saves the process of making and placing the calibration object. Besides, we use only two arbitrary feature points and five motions of robot, which can improve the efficiency of feature matching and calibration procedure greatly. To evaluate the accuracy and robustness of the proposed method, experimental results on simulated data with different noise and disturbance are implemented in this section.

The spatial configuration in the simulations is shown in Fig. 2. The distance of two feature points P_1 and P_2 to the camera coordinate system is 500 mm. The simulated camera has the following parameters: f_u = 800, f_v = 600, s = 0.01, u_0 = 400, v_0 = 280. The image resolution is 1000×1000 (pixel). The simulated hand-eye relation is set as follow: the rotation matrix R_m represented by an angle vector is $R_m = [5, 80, 7]^T$ (degree) the translation vector $\boldsymbol{p}_m$ is $[20.21, 13.57, 17.83]^T$ (mm). The two feature points in camera coordinate system are $\boldsymbol{X}_{01} = [200, 200, 500]^T$ and $\boldsymbol{X}_{02} = [-200\ -200\ 500]^T$, respectively. Three translational motions and two rotational motions of robot end-effector are shown in Table II.

TABLE II
SIMULATED MOTIONS OF ROBOT END-EFFECTOR

Parameter		Value		
Translational motions (mm)	$\boldsymbol{b}_1$	20	40	10
	$\boldsymbol{b}_2$	13	20	14
	$\boldsymbol{b}_3$	35	16	28
Rotation motions (degree)	R_4	15	30	15
	R_5	30	15	25

We establish five group coordinates of feature points from forward setting parameters using MATLAB software. From the setting parameters R_m, $\boldsymbol{X}_{01}$, K, $\boldsymbol{b}_i$ and (6) and (7), we can derive $\boldsymbol{m}_{01}$, $\boldsymbol{m}_{02}$, $\boldsymbol{m}_{i1}$ and $\boldsymbol{m}_{i2}$. Similarly, from the setting parameters R_m, $\boldsymbol{X}_{01}$, K, R_j and (4), (18) and (19), we can derive $\boldsymbol{m}_{j1}$, $\boldsymbol{m}_{j2}$. Gaussian noises with different levels are added to the corresponding setting parameters. The calibration parameters using our calibration method are calculated. The proposed calibration procedure is described in Section III-C.

Usuallythere are mainly two factors to influence the accuracy of calibration results: the noise of feature points extracted from images, and the noise of the robot motion parameters read from the controller. In the proposed method, another influence factor needs to be considered, i.e., the relative depth difference of the two feature points at the initial position of camera. Therefore, through simulation experiments we will analyze the influence of these three aspects, respectively.

A. Influence of the Image Noise

Image noise exists inevitably in the image collection process. In this experiment, we will analyze the sensitivity to the image noise of our calibration method. Gaussian noise with 0 mean and σ standard deviation is added to the projected image points. We vary the noise level from 0.1 to 5 pixels. For each noise level, we perform 200 independent trials, and the results are averaged. Then, we compare the estimated camera and hand-eye parameters with the truth-value and calculate the absolute errors. Table III lists the results of camera intrinsic parameters under the influence of image noise, and Table IV lists the results of hand-eye parameters under the influence of image noise.

In order to display the results visually, we draw the error curves as shown in Fig. 3. The absolute and relative errors of camera intrinsic matrix K are shown in Fig. 3 (a), the rotation matrix R_m represented by three angles is shown in Fig. 3 (b), and the translation vector $\boldsymbol{p}_m$ are shown in Fig. 3 (c). As we can see from Fig. 3, for σ = 5 pixels (which is much larger than the normal noise in practical calibration), the errors of intrinsic parameters are less than 0.5 pixel, and the rotational and translational motion parameters of hand-eye relation are less than 0.004 degree and 0.35 mm, respectively. The relative errors of intrinsic parameters and rotation vector are less than 0.07 %. For σ = 1.5 pixels, the relative errors of f_u, f_v are about 0.7 %, and the absolute errors of u_0, v_0 are about 4 pixels in Zhang's paper [8]. From our results, we conclude that the proposed algorithm has strong disturbance rejection ability and great robustness against the image noise.

TABLE III
RESULTS OF CAMERA INTRINSIC PARAMETERS

σ	Δf_u	Δf_v	Δs	Δu_0	Δv_0
0.5	0.0038	0.0029	0.0003	0.0087	0.0087
1.0	0.0325	0.0248	0.0024	0.0069	0.0159
1.5	0.0522	0.0399	0.0039	0.0077	0.0292
2.0	0.1071	0.0817	0.0079	0.0055	0.0961
2.5	0.0173	0.0132	0.0013	0.0682	0.1090
3.0	0.1011	0.0772	0.0074	0.0012	0.0735
3.5	0.1749	0.1335	0.0129	0.1280	0.0438
4.0	0.1045	0.0798	0.0077	0.0199	0.0407
4.5	0.3886	0.2967	0.0287	0.1532	0.0854
5.0	0.4392	0.3353	0.0324	0.1700	0.1029

TABLE IV
RESULTS OF HAND-EYE PARAMETERS

σ	$\Delta\varphi(^\circ)$	$\Delta\theta(^\circ)$	$\Delta\Psi(^\circ)$	ΔX (mm)	ΔY (mm)	ΔZ (mm)
0.5	0.0000	0.0000	0.0000	0.0101	0.0068	0.0137
1.0	0.0002	0.0001	0.0001	0.0077	0.0007	0.0139
1.5	0.0004	0.0001	0.0002	0.0168	0.0009	0.0179
2.0	0.0008	0.0002	0.0004	0.0677	0.0269	0.0040
2.5	0.0001	0.0000	0.0001	0.1060	0.0650	0.1063
3.0	0.0008	0.0002	0.0005	0.0523	0.0119	0.0162
3.5	0.0014	0.0004	0.0008	0.0898	0.0870	0.2221
4.0	0.0008	0.0002	0.0005	0.0222	0.0122	0.0547
4.5	0.0030	0.0008	0.0017	0.0111	0.0766	0.2833
5.0	0.0033	0.0009	0.0019	0.0069	0.0842	0.3155

B. Influence of the Robot Motion Noise

In practice, the robot brings motion errors as mentioned before. This experiment examines the influence of the motion errors from robot. With the development of the techniques of robots, most of the precise industrial robots have the repeated accuracy of positioning from 0.02 mm to 0.1 mm. Therefore, we add Gaussian noise with mean 0 and standard deviation from 0 to 0.1 mm to the robot motion parameters and repeat this process. The average errors are calculated and the results are shown in Fig. 4. The absolute and relative errors of camera intrinsic matrix K are shown in Fig. 4 (a), the rotation matrix R_m represented by three angles is shown in Fig. 4 (b), and the translation vector $\boldsymbol{p}_m$ is shown in Fig. 4 (c).

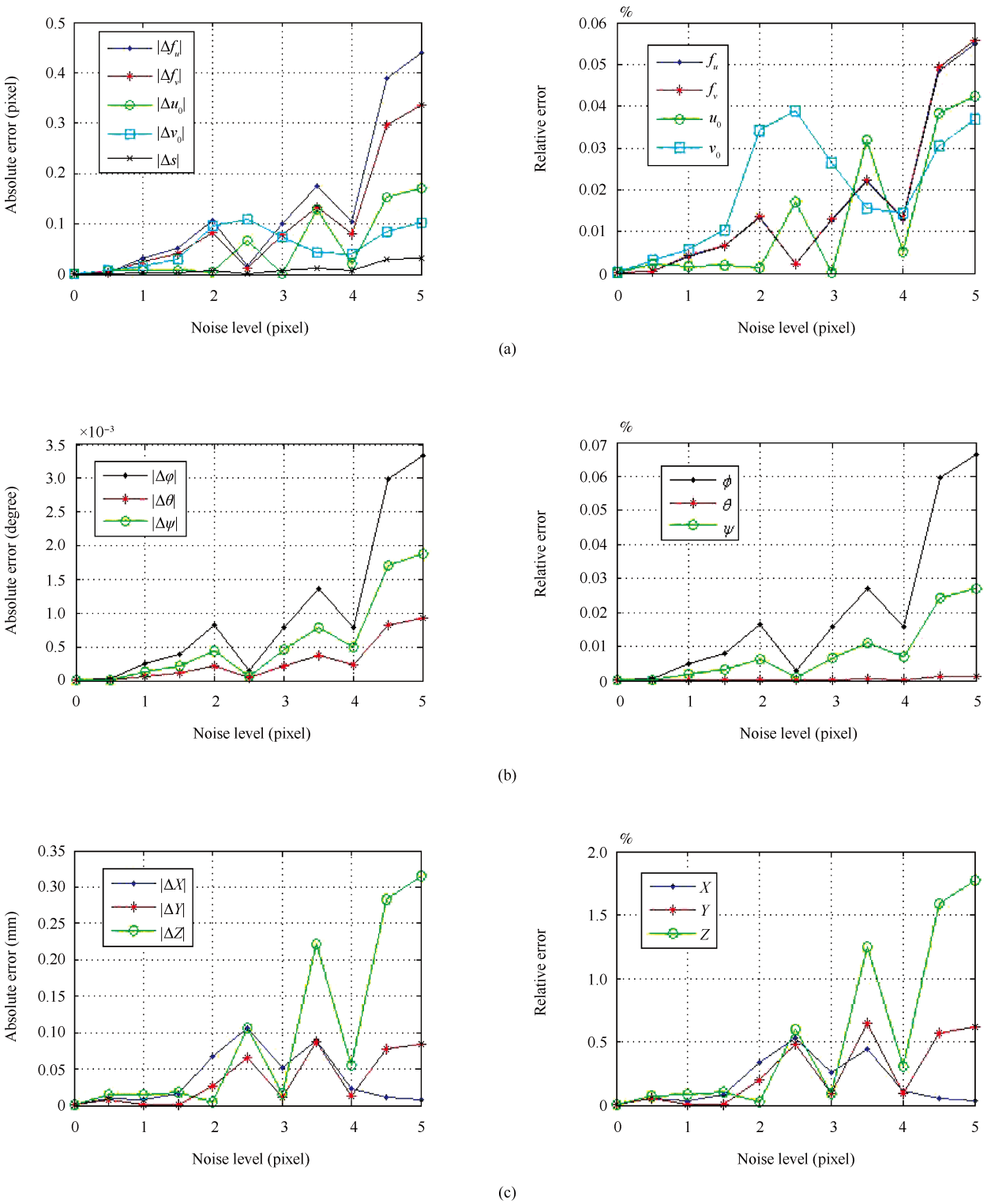

Fig. 3. Errors vs. the noise level of the image points. (a) The absolute and relative errors of camera intrinsic matrix. (b) The absolute and relative errors of the rotation vector. (c) The absolute and relative errors of the translation vector.

As shown in Fig. 4, when the noise level of the robot motions is 0.08mm, the absolute errors of intrinsic parameters are less than 0.6 pixel, and the rotational errors and translational errors of hand-eye relation are less than 0.02 degree and 0.4 mm, respectively. For $\sigma = 0.1$ mm, the relative errors of intrinsic parameters are less than 0.14 %. Compared with intrinsic parameters, the rotation and translation vectors are more sensitive to the robot motion noise, but still can meet the requirements of practical application. Because of the uncertainty of the Gaussian noise, the errors are fluctuating but still increase with the noise level generally. Many robots have quite high position repeatability, so the proposed method can achieve the calibration even more accurately in practical applications.

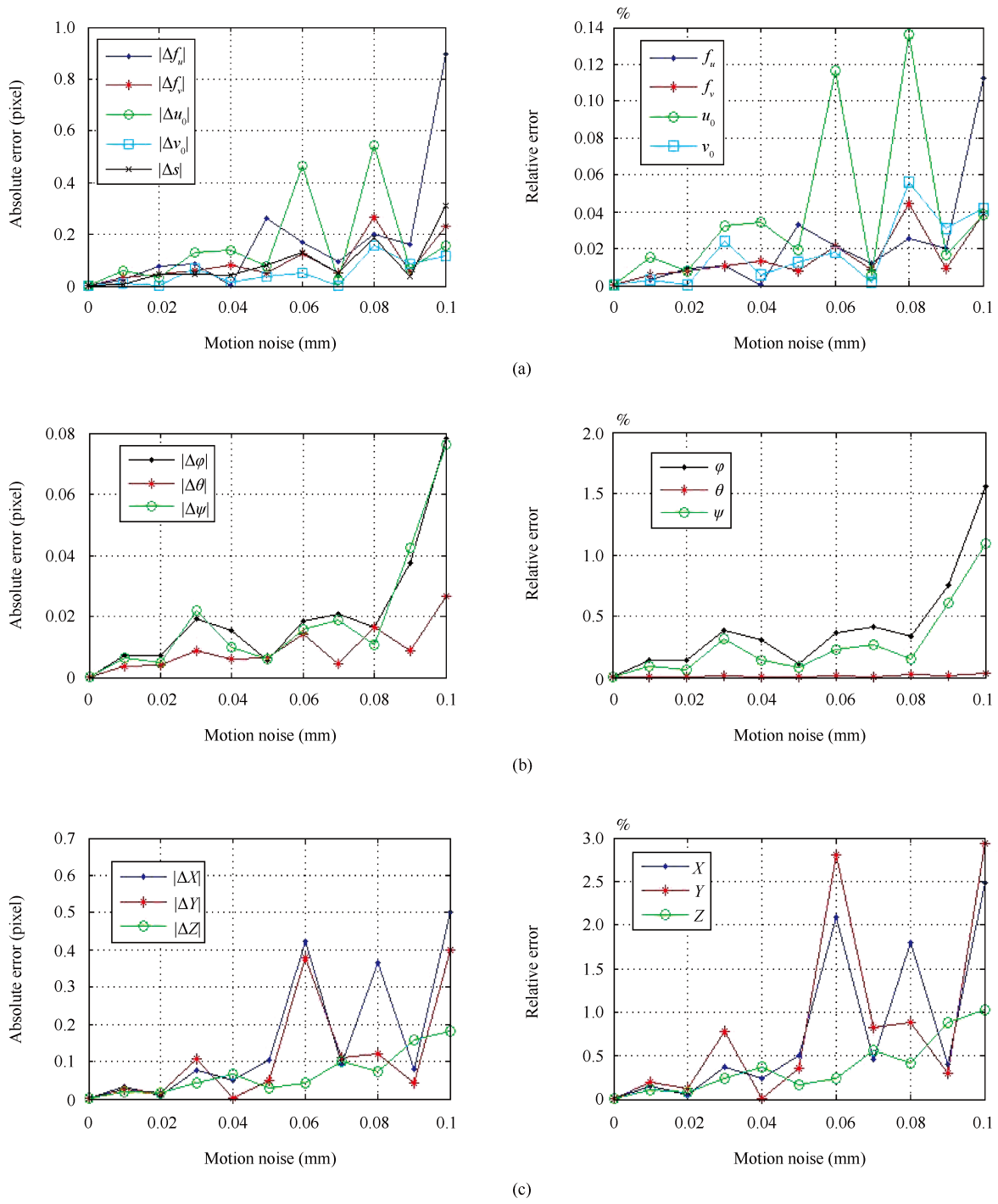

Fig. 4. Errors vs. the noise level of robot motion. (a) The absolute and relative errors of camera intrinsic matrix K. (b) The absolute and relative errors of the rotation vector. (c) The absolute and relative errors of the translation vector.

C. Influence of the Initial Position of Camera

In Section III, we have an assumption about initial position of the camera: the depth values of the two feature points P_1 and P_2 in the camera coordinate system are equal. However, it is not practical to promise this condition absolutely. This experiment investigates the sensitivity of the proposed calibration method with respect to the initial position of camera. Suppose the relative depth difference of the feature points is z_c, the sensitivity of self-calibration with respect to z_c is examined in this experiment. Gaussian noise with 0 mean and standard deviation σ is added to the depth value of feature point P_1. We vary the noise level from 0.1 to 1 mm. 200 independent trials are performed for each noise level, and the average errors are computed. The results are shown in Fig. 5. The absolute and relative errors of camera intrinsic matrix K are shown in Fig. 5 (a), the rotation matrix R_m represented by three angles are shown in Fig. 5 (b), and the translation vector $\boldsymbol{p}_m$ are shown in Fig. 5 (c).

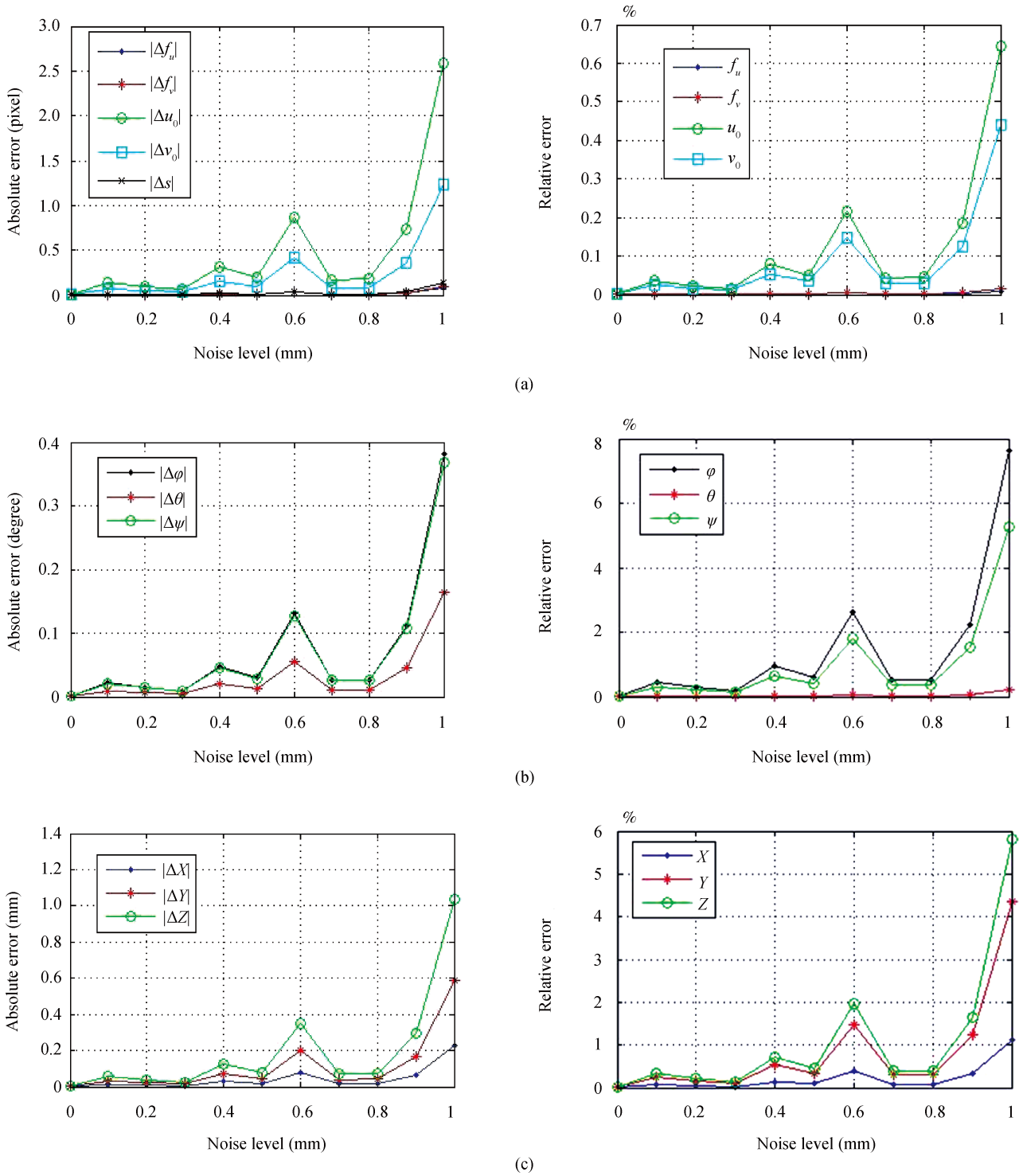

Fig. 5. Errors vs. the noise level of $\Delta_z c$. (a) The absolute and relative errors of camera intrinsic matrix K. (b) The absolute and relative errors of the rotation vector. (c) The absolute and relative errors of the translation vector.

In Fig. 5 (a), it can be seen that the absolute and relative errors of intrinsic parameters are less than 3 pixels and 0.7 %, respectively, for σ = 1 mm. The focal length yields better results than principlalpoint, i.e., the parameters u_0 and v_0 are more sensitive to the influence of z_c. The relative errors of the rotation and translation vector parameters are less than 8 % and 6 %. The errors of the translation vector parameters are less than 2 mm when the noise level is in the range of 0 to 1 mm. Except the focal length, θ and $\boldsymbol{X}$, the other parameters have a relative large variation after the 0.5 mm noise level. Therefore, the assumption of the camera initial position has an influence on the calibration results. But in practice, many methods can be taken to reduce z_c. For example, in certain robot assembly system, laser distance sensors (LDSs) are used to measure the angles and distance between the object and the robot end-effector [27], [28]. Because the camera and LDSs

are both rigidly mounted on the robot end-effector, we can use these LDSs to adjust the initial position of camera.

The above experimental results show that the presented self-calibration is simple and practical from the aspect of calibration procedure and accurate in calibration results. It has strong anti-disturbance ability and great robustness against the image noise and the robot motion error. The initial position of camera can be controlled in a small range by other auxiliary sensors, such as laser range-finders, and with the feedback of laser range-finders the robot end-effector can be adjusted to the correct initial location in a very short time.

From the experiments, there are three factors which influence the accuracy of calibration results: the image noise, the robot motion noise and the initial position of feature points. The noise of robot motion can be reduced by choosing high precision robot system. For the other two factors, we can control their influence by choosing better feature points. From the results of simulation experiments, two important principles can be concluded. Firstly, the chosen points should be easy to extract, such as centers of circles and cross corners. Second, in order to reduce the relative depth difference of the two feature points, they should be located in the same plane. Therefore, we can use LDSs to adjust the initial position of camera. Considering these two principles above, in practical environments, we can choose centers of regular circles in the same plane, or cross corners of square ceiling, or intersections of object's edges in some assembly tasks.

V. Conclusions and Future Work

A novel self-calibration approach that can effectively estimate both the camera intrinsic parameters and the hand-eye transformation for robot vision are presented in this paper. The proposed method performs calibration without any calibration objects, and combines the camera calibration and hand-eye calibration in one procedure. Based on two arbitrary feature points of the environment, the proposed calibration procedure requires only three pure translational motions and two rotational motions of robot end-effector. Because it needs only fewer feature points and robot motions, the efficiency of the calibration procedure is greatly improved. Some new linear solution equations are deduced and experiments on computer simulated data have been performed to analyze how different factors influence the calibration results, such as image noise, robot motion noise and the difference of the depth value of feature points. The results demonstrate the proposed method is effective and practical.

For future work, we will focus on two points: First, the improvement on the methods to relax the requirement of the initial camera position. For example, we can use certain compensation methods to solve this problem in principle, or use the LDSs mentioned in experiment to solve the problem from the aspect of hardware. Second, the application and evaluation of the proposed method in a practical robot vision system for some special tasks in industrial environments, such as grasping, handling and assembly.

References

[1] D. Fontanelli, F. Moro, T. Rizano, and L. Palopoli, "Vision-Based Robust Path Reconstruction for Robot Control," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 4, pp. 826−837, Apr. 2014.

[2] J. Y. Lu, D. Xu, Z. K. Qin, P. Wang, and C. Ren, "An automatic alignment strategy of large diameter components with a multi-sensor system," *Acta Automat. Sinica*, vol. 41, no. 10, pp. 1711−1722, Oct. 2015.

[3] A. A. Shafie, A. B. M. Ibrahim, and M. M. Rashid, "Smart objects identification system for robotic surveillance," *Int. J. Automat. Comput.*, vol. 11, no.1, pp. 59−71, Feb. 2014.

[4] J. Sun, P. Wang, Z. K. Qin, and H. Qiao, "Overview of camera calibration for computer vision," in *Proc. 11th World Congr. Intelligent Control and Automation (WCICA)*, Shenyang, China, 2014, pp. 86−92.

[5] J. Hieronymus, "Comparison of methods for geometric camera calibration," in *Proc. Int. Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XXXIX-B5, 2012, XXII ISPRS Congress*, Melbourne, Australia, 2012, pp. 595−599.

[6] L. Wang, F. Q. Duan, and K. Lv, "Camera calibration with one-dimensional objects based on the heteroscedastic error-in-variables model," *Acta Automat. Sinica*, vol. 40, no. 4, pp. 643−652, Apr. 2014.

[7] R. Y. Tsai, "An efficient and accurate camera calibration technique for 3D machine vision," in *Proc. 1986 IEEE Conf. Computer Vision and Pattern Recognition*, Miami, FL, USA, 1986, 364−374.

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

[9] M. Große, M. Schaffer, B. Harendt, and R. Kowarschik, "Camera calibration using time-coded planar patterns," *Opt. Eng.*, vol. 51, no. 8, pp. Article ID 083604, Aug. 2012.

[10] Z. J. Zhao, Y. C. Liu, and Z. Y. Zhang, "Camera calibration with three noncollinear points under special motions," *IEEE Trans. Image Process.*, vol. 17, no. 12, pp. 2393−2402, Dec. 2008.

[11] X. C. Cao and H. Foroosh, "Camera calibration using symmetric objects," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3614−3619, Nov. 2006.

[12] P. Hammarstedt, P. Sturm, and A. Heyden, "Degenerate cases and closed-form solutions for camera calibration with one-dimensional objects," in *Proc. 10th IEEE Int. Conf. Computer Vision*, Beijing, China, 2005, pp. 317−324.

[13] Z. Y. Zhang, "Camera calibration with one-dimensional objects," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 7, pp. 892−899, Jul. 2004.

[14] I. Miyagawa, H. Arai, and H. Koike, "Simple camera calibration from a single image using five points on two orthogonal 1-D objects," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1528−1538, Jun. 2010.

[15] S. J. Maybank and O. D. Faugeras, "A theory of self-calibration of a moving camera," *Int. J. Comput. Vis.*, vol. 8, no. 2, pp. 123−151, Aug. 1992.

[16] H. Wildenauer and A. Hanbury, "Robust camera self-calibration from monocular images of Manhattan worlds," in *Proc. 2012 IEEE Conf. Computer Vision and Pattern Recognition*, Providence, RI, USA, 2012, pp. 2831−2838.

[17] D. N. Dawson and S. T. Birchfield, "An energy minimization approach to automatic traffic camera calibration," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 3, pp. 1095−1108, Sep. 2013.

[18] S. D. Ma, "A self-calibration technique for active vision systems," *IEEE Trans. Robot. Automat.*, vol. 12, no. 1, pp. 114−120, Feb. 1996.

[19] C. Almeida Santos, C. Oliveira Costa, and J. Batista, "A vision-based system for measuring the displacements of large structures: Simultane-

ous adaptive calibration and full motion estimation," *Mech. Syst. Signal Process.*, vol. 72−73, pp. 678−694, May 2016.

[20] N. Andreff, R. Horaud, and B. Espiau, "Robot hand-eye calibration using structure-from-motion," *Int. J. Robot. Res.*, vol. 20, no. 3, pp. 228−248, Mar. 2001.

[21] H. Malm and A. Heyden, "Hand-eye calibration from image derivatives," in *Proc. 6th European Conf. Computer Vision*, Dublin, Ireland, 2000, pp. 493−507.

[22] H. Malm and A. Heyden, "Simplified intrinsic camera calibration and hand-eye calibration for robot vision," in *Proc. 2003 IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, Las Vegas, NV, USA, vol. 1, pp. 1037−1043, 2003.

[23] Z. J. Zhao and Y. Weng, "A flexible method combining camera calibration and hand-eye calibration," *Robotica*, vol. 31, no. 5, pp. 747−756, Aug. 2013.

[24] K. H. Strobl and G. Hirzinger, "More accurate camera and hand-eye calibrations with unknown grid pattern dimensions," in *Proc. 2008 IEEE Int. Conf. Robotics and Automation*, Pasadena, CA, USA, 2008, pp. 1398−1405.

[25] H. X. Wang, X. H. Fan, and X. Lu, "Application of a hand-eye self-calibration technique in robot vision," in *Proc. 25th Chinese Control and Decision Conf.*, Guiyang, China, 2013, pp. 3765−3769.

[26] C. Lei, F. C. Wu, and Z. Y. Hu, "A new camera self-calibration method based on active vision system," *Chin. J. Comput.*, vol. 23, no. 11, pp. 1130−1139, Nov. 2000.

[27] Y. Kim, Y. K. Kim, K. S. Kim, S. Kim, B. M. Kwak, I. G. Jang, Y. S. Jung, and E. H. Kim, "Structure optimization of 1-D laser sensors assembly for robust 6-DOF measurement," in *Proc. 2012 IEEE Int. Systems Conf.*, Vancouver, BC, Canada, 2012, pp. 1−4.

[28] Z. K. Qin, P. Wang, J. Sun, J. Y. Lu, and H. Qiao, "Precise robotic assembly for large-scale objects based on automatic guidance and alignment," *IEEE Trans. Instrum. Meas.*, vol. 65, no. 6, pp. 1398−1411, Jun. 2016.

Jia Sun received the B.S. degree from the North University of China, Taiyuan, China, in 2009, and the M.S. degree from the Beijing Institute of Technology, Beijing, China, in 2012. She is currently pursuing the Ph.D. degree with the University of Chinese Academy of Sciences, Beijing. Her current research interests include robot vision, precise assembly, and camera calibration.

Peng Wang (M'11) received the B.Eng. degree in electrical engineering and automation from Harbin Engineering University, Harbin, China, in 2004, the M.Eng. degree in automation science and engineering from Harbin Institute of Technology, Harbin, China, in 2007, and the Ph.D. degree from Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2010. He is currently an associate professor with the Institute of Automation, Chinese Academy of Sciences, Beijing, China. His current research interests include intelligent robot, industrial robot, robotic assembly, robotic vision, image processing and visual perception model. Corresponding author of this paper.

Zhengke Qin received the B.S. degree from the Huazhong University of Science and Technology, Wuhan, China, in 2012. He is currently pursuing the Ph.D. degree with the University of Chinese Academy of Sciences, Beijing. His current research interests include computer vision and robotic assembly.

Hong Qiao (SM'06) received the B.Eng. degree in hydraulics and control, the M.Eng. degree in robotics from Xi'an Jiaotong University, Xi'an, China, the M.Phil. degree in robotics control from the Industrial Control Center, University of Strathclyde, Strathclyde, U.K., and the Ph.D. degree in robotics and artificial intelligence from De Montfort University, Leicester, U.K., in 1995. She was a research assistant professor from 1997 to 2000 and an assistant professor from 2000 to 2002 with the Department of Manufacturing Engineering and Engineering Management, City University of Hong Kong, Kowloon, Hong Kong. In 2002, she joined as a lecturer with the School of Informatics, University of Manchester, Manchester, U.K. She is currently a professor with the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China.