# SIMULTANEOUS CAMERA AND HAND-EYE CALIBRATION FOR EYE-ON-HAND SYSTEMS 

Nguyen Huu Hung ${ }^{1}$, Nguyen Quang Thi ${ }^{1}$, Tran Cong Manh ${ }^{1}$


#### Abstract

Determination of the location and orientation of objects in the robot workspace is a fundamental function of manufacturing automation. This problem is solved by using a robot vision system with a camera mounted on the robot end-effector with a known hand-eye transformation. In this case, a viable solution to deal with the complexity of calibration is of necessity, including the calibration of internal and external parameters associated with the camera as well as the calibration of the hand-eye parameters. To this end, the paper presents a simple and efficient method of calibration for a camera-on-hand system, where the internal and external parameters of a 2D camera as well as the hand-eye parameters are simultaneously calibrated. The method is based on the 3D-to-2D projections of a calibrationblock with minimum two pure translations and two pure rotations of hand motions. Being evaluated on simulation data and real robot vision system indicate that our method can work stably with different noise and number of stations.


## Index terms

Camera Calibration, Hand-Eye Calibration.

## 1. Introduction

The robot vision system consists of one or more cameras and a robot or robots in industry field for various applications such as bin picking, modeling [1], robotic grasping [2] and medical procedures [3]. The measurement accuracy of the robot vision system relies on each component in the system containing camera parameters, handeye parameters and the hand's repeatability accuracy, obviously. Usually, the calibration process for the robot-vision system is done separately in turn, camera calibration first and then hand-eye calibration later. Especially, the images used for camera calibration are not reused for hand-eye calibration process.

Camera calibration is a necessary process in 3D computer vision in order to solve the unknown parameters of the camera model. It is performed by observing a calibration object whose geometry in 3D space is known with the high precision. The calibration object usually contains one, two or three planes perpendicular to each other. Much work has been done, for example [4], [5], [6]. The approaches using one plane, named planar chessboard, usually require an expensive apparatus and an elaborate setup.

[^0]Usually, the hand-eye calibration problem is formulated as solving homogeneous transformation equations of the forms $A X=X B[7](*)$, where $X$ is the 4 homogeneous transformation from the robot hand coordinate frame to the sensor coordinate frame, $A$ and $B$ are the measurable 4 homogeneous transformations of the robot hand and camera from its first to second position. Several closed-form solutions were proposed to solve for $X$ such as [7] and [8]. The unknown hand-eye transformation also can be estimated by solving $A X=Z C$ where $A$ is the known homogeneous transformation from hand pose measurements, $C$ is computed using the calibrated manipulator internal-link forward kinematics, $X$ is the unknown transformation from the robot hand frame to sensor frame, and $Z$ is the unknown transformation from the world frame to the robot-base frame. Such problem has been solved in [9] and [10]. This hand-eye calibration process is independent to camera calibration. Hand-eye calibration by teaching pendant moving the robot hand to a sequence of location repeatedly also has been used for several decades. This teaching robot moving to pick chessboard corner was known as to be dangerous for operators and time consuming. Any incorrect operation could cause severe injury people close to the robot. Note that, for above-mentioned approaches, to successfully obtain the unknown transformation $X$ accurately, it is necessary to solve a non-linear system obtaining from multiple hand motions and images captured by camera. Additionally, the camera should be calibrated previously and separately. Recently, simultaneous hand-eye calibration was done by using chessboard corners at multiple hand positions [11]. This methodology requires multiple hand motions so it shares the same manner with teaching pendant method.


Fig. 1. The proposed simultaneous hand-eye calibration.

In this paper, we focus on reducing the elaborating time by introducing an approach which simultaneously calibrates camera and camera-hand using 3D calibration-block by minimum two pure hand translations and two pure hand rotations, meaning four hand motions in total. Firstly, the camera was calibrated successfully at each robot stations by taking advantage of 3D calibration-block which contains two orthogonal planes.

Secondly, hand-eye rotation is estimated by minimum two pure hand translations and then the hand-eye translation was obtained by minimum two pure hand rotations with the pre-estimated hand-eye rotation. Fig. 1 summarizes the step-by-step procedure of the proposed approach.

The remaining of this paper is organized as follows. In Section II, background of camera model and traditional hand-eye calibration are summarized. In Section III, a process of camera calibration is estimated by finding projection matrices and decomposing into internal and external parameters. Hand-eye calibration based on pure hand translation and rotation motions is described in Section IV. Experimental results on simulated and real data are shown in Section V.

## 2. Preliminaries

### 2.1. Camera model

In pinhole camera system, the relation between a 3D world point ${ }^{W} X=[x, y, z, 1]^{T}$ and a 2D point in image plane is known as the full perspective camera model [12].

$$
{ }^{C}\left(\begin{array}{l}
u  \tag{1}\\
v \\
1
\end{array}\right)={ }^{C}\left[\begin{array}{ccc}
f_{u} & s & c_{u} \\
0 & f_{v} & c_{v} \\
0 & 0 & 1
\end{array}\right]_{W}^{C}\left[\begin{array}{llll}
R_{11} & R_{12} & R_{13} & t_{x} \\
R_{21} & R_{22} & R_{23} & t_{y} \\
R_{31} & R_{32} & R_{33} & t_{z}
\end{array}\right]^{W}\left(\begin{array}{c}
x \\
y \\
z \\
1
\end{array}\right)
$$

This includes a homogeneous transformation matrix which transforms the 3D world point to camera coordinate and a projection matrix from camera coordinate to image co-ordinate in pixel. Equation (1) can be rewritten in a compact way

$$
\begin{equation*}
{ }^{C} p={ }^{C} K\left[{ }_{W}^{C} R{ }_{W}^{C} t\right]{ }^{W} P \tag{2}
\end{equation*}
$$

or be compacted as

$$
\begin{equation*}
{ }^{C} p={ }_{W}^{C} M{ }^{W} P \tag{3}
\end{equation*}
$$

where matrix $3 \times 3{ }^{C} K$ is intrinsic matrix or camera matrix contains the internal parameters of camera, rotation matrix $3 \times 3{ }^{C} R$ and translation matrix $3 \times 1{ }_{W}^{C} t$ are external matrix represent the transformation from world coordinate to camera coordinate. Matrix $3 \times 4{ }_{W}^{C} M$ is called as projection matrix. In this case, the pair 2D point ${ }^{C} p$ and 3D point ${ }^{W} P$ is called a $2 D-3 D$ correspondence. The calibration process is to estimate intrinsic matrix ${ }^{C} K$ from $2 D-3 D$ correspondences.


Fig. 2. Camera Calibration Process using Calibration-Block (left) Captured Image (middle) $7 \times 7$ line detection, (right) $7 \times 7$ line intersections with sub-pixel refinement.

### 2.2. Hand-eye calibration

In a vision system with camera mounted on hand, the well-known hand-eye calibration equation is expressed as $A X=X B$ where A is hand motion, B is camera motion and $X$ is transformation from the camera to the end-effector that is necessary to estimate correctly. From equation (*), two following constraints should be satisfied the

$$
\begin{align*}
R_{A} R_{X} & =R_{X} R_{B}  \tag{4}\\
\left(I-R_{A}\right) t_{X} & =t_{A}-R_{X} t_{B} \tag{5}
\end{align*}
$$

Several approaches have been proposed for the estimation of $R_{X}$ from equation (4), for instance, using the rotation axis and angle [13], quaternions [14] and canonical matrix representation [9]. After that, the estimation of translation is done by solving the Pseudo-Inverse and then can be refined by nonlinear optimization [8]. For these approaches, the hand motions include both translation and rotation components.

## 3. Camera Calibration

### 3.1. Calibration-block and corner detection

Calibration-block contains two orthogonal planar chessboards $8 \times 8$ with 20 mm square size. That means there are $7 \times 7=49$ inner corners in each surface and 98 corners in total. To detect these corners, consecutively, we performed Hough Line Transform method to extract lines, then the initial location of inner corners are determined by line intersections, and finally the sub-pixel process was performed to accurately detect location inner corners. These main steps to detect chessboard corners are summarized in Fig. 2.

### 3.2. Camera calibration

At each robot station, the camera was calibrated; $3 \times 4$ projection matrix was estimated from solving linear equations then was decomposed into intrinsic and extrinsic
parameter. Specifically, equation (3) can be written detailed as

$$
\begin{gather*}
{ }^{C}\left(\begin{array}{l}
u \\
v \\
1
\end{array}\right)={ }^{C}\left[\begin{array}{ccc}
f_{u} & s & c_{u} \\
0 & f_{v} & c_{v} \\
0 & 0 & 1
\end{array}\right]_{W}^{C}\left[\begin{array}{llll}
R_{11} & R_{12} & R_{13} & t_{x} \\
R_{21} & R_{22} & R_{23} & t_{y} \\
R_{31} & R_{32} & R_{33} & t_{z}
\end{array}\right]^{W}\left(\begin{array}{l}
x \\
y \\
z \\
1
\end{array}\right)  \tag{6}\\
u_{i}=\frac{m_{11} x_{i}+m_{12} y_{i}+m_{13} z_{i}+m_{14}}{m_{31} x_{i}+m_{32} y_{i}+m_{33} z_{i}+m_{34}}  \tag{7}\\
v_{i}=\frac{m_{21} x_{i}+m_{22} y_{i}+m_{23} z_{i}+m_{24}}{m_{31} x_{i}+m_{32} y_{i}+m_{33} z_{i}+m_{34}} \tag{8}
\end{gather*}
$$

Equations (7) and (8) are written in compact way as

$$
\left[\begin{array}{ccc}
{ }^{W} P_{i}^{T} & 0 & u_{i}{ }^{W} P_{i}^{T}  \tag{9}\\
0 & { }^{W} P_{i}^{T} & v_{i}{ }^{W} P_{i}^{T}
\end{array}\right] m=0
$$

where $m=\left[m_{11} m_{12} m_{13} m_{14} m_{21} m_{22} m_{23} m_{24} m_{31} m_{32} m_{33} m_{34}\right]$. To find each element of matrix $M$, it is necessary to solve the linear equation $A m=0$. The simple way to solve (9) is to find the minimum eigenvector by minimizing the following objective function

$$
\begin{equation*}
\min \|A \bar{m}\| \tag{10}
\end{equation*}
$$

with a constraint $\|\bar{m}\|=1$. The solution of $m$ vector is normalized, however matrix

$$
{ }^{C} K_{W}^{C} R={ }_{W}^{C}\left[\begin{array}{lll}
\bar{m}_{11} & \bar{m}_{12} & \bar{m}_{13}  \tag{11}\\
\bar{m}_{21} & \bar{m}_{22} & \bar{m}_{23} \\
\bar{m}_{31} & \bar{m}_{32} & \bar{m}_{33}
\end{array}\right]
$$

where $\bar{m}_{31}^{2}+\bar{m}_{32}^{2}+\bar{m}_{33}^{2}=1$. For that reason, we have

$$
{ }^{C} K_{W}^{C} R=\frac{\left[\begin{array}{ccc}
\bar{m}_{11} & \bar{m}_{12} & \bar{m}_{13}  \tag{12}\\
\bar{m}_{21} & \bar{m}_{22} & \bar{m}_{23} \\
\bar{m}_{31} & \bar{m}_{32} & \bar{m}_{33}
\end{array}\right]}{\operatorname{sqrt}\left(\bar{m}_{31}^{2}+\bar{m}_{32}^{2}+\bar{m}_{33}^{2}\right)}
$$

And then intrinsic parameters and rotation matrix from world to camera are obtained by QR decomposition. Finally, the translation from world to camera is calculated by solving

$$
{ }^{C} K_{W}^{C} t=\left(\begin{array}{l}
m_{14}  \tag{13}\\
m_{24} \\
m_{34}
\end{array}\right)
$$



Fig. 3. Robot vision system with a camera mounted on hand.

## 4. Hand-eye Calibration

In this section, we propose a simplifying hand-eye calibration using at least two pure hand translations to estimate the hand-eye rotation and two pure hand rotations to obtain hand-eye translation with known hand-eye rotation. Fig. 3 shows two positions of the robot vision system. Hand position is so called a station. For other approaches, the hand motions including both translation and rotation components. However, the proposed approach takes advantage using special motions: pure translation and pure rotation.

The hand can be controlled by human so hand motion is obtained by

$$
\begin{equation*}
{ }_{H_{2}}^{H_{1}} T={ }_{B}^{H_{1}} T_{B}^{H_{2}} T^{-1} \tag{14}
\end{equation*}
$$

When camera is calibrated to world coordinate so camera motion also easily obtained

$$
\begin{equation*}
{ }_{C_{2}}^{C_{1}} T={ }_{W}^{C_{1}} T_{W}^{C_{2}} T^{-1} \tag{15}
\end{equation*}
$$

The hand motion and camera motion share a constraint $A X=X B$. This constraint is rewritten as

$$
\begin{equation*}
{ }_{C_{2}}^{H_{1}} T={ }_{H_{2}}^{H_{1}} T{ }_{C_{2}}^{H_{2}} T={ }_{C_{1}}^{H_{1}} T_{C_{2}}^{C_{1}} T \tag{16}
\end{equation*}
$$

or in detail as

$$
\left[{ }_{H_{2}}^{H_{1}} R_{H_{2}}^{H_{1}} t\right]\left[\begin{array}{c}
{ }_{C_{2}}  \tag{17}\\
H_{2}
\end{array}{ }_{C_{2}}^{H_{2}} t\right]=\left[\begin{array}{c}
{ }_{C_{1}}{ }_{C_{1}}
\end{array} R_{C_{1}}^{H_{1}} t\right]\left[\begin{array}{l}
{ }_{C}^{2}
\end{array} R_{C_{2}}^{C_{1}} t\right]
$$

The rotation constraint is

$$
\begin{gather*}
{ }_{H_{2}}^{H_{1}} R_{C_{2}}^{H_{2}} R={ }_{C_{1}}^{H_{1}} R_{C_{2}}^{C_{1}} R  \tag{18}\\
{ }^{H} R{ }_{C}^{H} R={ }_{C}^{H} R^{C} R \tag{19}
\end{gather*}
$$

and, the translation constraint is

$$
\begin{equation*}
{ }_{H_{2}}^{H_{1}} t+{ }_{H_{2}}^{H_{1}} R{ }_{C_{2}}^{H_{2}} t={ }_{C_{1}}^{H_{1}} t+{ }_{C_{1}}^{H_{1}} R_{C_{2}}^{C_{1}} t \tag{20}
\end{equation*}
$$

Finally, the translation component is described by hand motion and camera motion as

$$
\begin{equation*}
{ }^{H} t+{ }^{H} R{ }_{C}^{H} t={ }_{C}^{H} t+{ }_{C}^{H} R^{C} t \tag{21}
\end{equation*}
$$

In case of pure translation motion, there is no change in rotation that means ${ }^{H} R=I$ and then equation (21) becomes ${ }^{H} t={ }_{C}^{H} R^{C} t(* *)$. Many authors noticed that at least three pairs of $\left({ }^{H} t,{ }^{C} t\right)$ are necessary in order to uniquely determine ${ }_{H}^{C} R$. However, rotation ${ }_{H}^{C} R$ is orthogonal matrix so cross vector two hand motions and two camera motion has relationship

$$
\begin{equation*}
{ }^{H} t_{1} \times{ }^{H} t_{2}={ }_{C}^{H} R\left({ }^{H} t_{1} \times{ }^{H} t_{2}\right) \tag{22}
\end{equation*}
$$

For that reason, we need at least two pairs of hand motion to obtain the hand-eye rotation. Assume there are $N$ pair of hand motion and camera motion. Matrix $\Lambda$ is collection of hand motions and matrix $\Psi$ is collection of camera motions. From equation (22), relationship between $\Lambda$ and $\Psi$ is written in a closed-form as

$$
\begin{equation*}
\Lambda_{3 \times N}={ }_{C}^{H} R \Psi_{3 \times N} \tag{23}
\end{equation*}
$$

In case $N=2$, the third column of matrix $\Lambda$ and $\Psi$ is replaced by cross product of two hand motions and camera motions. Otherwise, in case number of stations is larger than 2 , rotation ${ }_{C}^{H} R$ is easily obtain by using $S V D$

$$
\begin{gather*}
{[U, S, V]=S V D\left(\Psi_{3 \times N} \Lambda_{3 \times N}^{T}\right)}  \tag{24}\\
{ }_{C}^{H} R=V U^{T} \tag{25}
\end{gather*}
$$

Until now, the hand-eye rotation is estimated, it is necessary to estimate the hand-eye translation. With known hand-eye rotation we can rewrite equation (21) as

$$
\begin{equation*}
{ }^{H} t-{ }_{C}^{H} R^{C} t=\left(I-{ }^{H} R\right){ }_{C}^{H} t \tag{26}
\end{equation*}
$$

For each hand rotation, we can have equation (26) with three linear equation, but matrix $\left(I-{ }^{H} R\right)$ has rank 2 so we need at least two hand motion in pure rotation to estimate translation ${ }_{C}^{H} t$.

## 5. Experimental Results

To verify the proposed method, we do experiment on both simulation data and real data which are captured by a robot vision system including a Computar camera with focal length 12 mm resolution $1280 \times 960$ mounted on Schunk LWA3 robot. Simulation data contained 20 pairs of hand-eye motion including 10 motion in pure translation and 10 motions in pure rotation. The hand motion is in a limitation 100 mm for translation direction, 20 degrees for rotation components. Noise level for rotation and translation were set to 0.05 degree and 0.05 mm , respectively. To verify our approach in real situation, we choose randomly a small number of motions from a dataset containing 60 pure translations and 50 pure rotations captured from above-mentioned system.


Fig. 4. Configuration of our robot vision system.

The configuration of our robot vision system is shown in Fig 4.
Firstly, we considered the hand error. There are two errors related to hand: physical error and repeatability error. The repeatability error can be roughly estimated by moving the hand from fixed positions to different positions. The arm controller can provide the joint information and the hand location in Base coordinate. By controlling the hand from the default position to a chosen position 60 times and measuring the standard deviation of transformation from Hand $(H)$ to Base $(B){ }_{H}^{B} T$, the repeatability is around 0.2 mm for translation and 0.05 degree for rotation. It is summarized in Table 1.

Table 1. Statistic Hand Repeatability Error

|  | $\mathbf{R}_{\mathbf{x}}(\mathbf{d e g})$ | $\mathbf{R}_{\mathbf{y}}(\mathbf{d e g})$ | $\mathbf{R}_{\mathbf{z}}(\mathbf{d e g})$ | $\mathbf{T}_{\mathbf{x}}(\mathbf{m m})$ | $\mathbf{T}_{\mathbf{y}}(\mathbf{m m})$ | $\mathbf{T}_{\mathbf{z}}(\mathbf{m m})$ |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $\mathbf{B}-\mathbf{H}$ | 0.047 | 0.02 | 0.037 | 0.14 | 0.1 | 0.12 |

However, it is not easy to measure the physical error directly. To measure this error roughly, we evaluated the hand-eye system by evaluating the relative translations of pure hand translation and corresponding camera motion instead. In ideal cases, the relative motions of hand and camera are identical. Due to the mechanic error, two motions are different with a very small change. This difference indicates the error of hand-eye system. To this end, we measured the hand motion and camera motion by moving hand 60 times in pure translation and then created a histogram of difference between two motions is shown in Fig. 5. It shows the mean and standard deviation of this difference to be $0.94 \pm 0.65$.

Fig. 5 indicates that the physical error of robot hand is around 0.9 mm . The system was evaluated on accuracy and precision for camera calibration and hand eye calibration. To make the evaluation understandable, we introduced the evaluation methodology which was used for evaluating camera calibration and hand-eye calibration, consecutively.


Fig. 5. Difference of hand and camera relative translation histogram.

### 5.1. Evaluation Methodology

This section introduces how the system was evaluated and the metric used to evaluate performance. For camera calibration, we measured precision of intrinsic and extrinsic parameters by measuring distribution of 20 calibration times at the same position. For hand-eye calibration, the performance was evaluated in both simulation and the real data. For accuracy evaluation, it is necessary to have metrics. We took advantage of metric of rotation and translation proposed in [7]. Assume there are two measurements of same transformation ${ }_{H_{1}}^{H_{0}} \tilde{T}$ and ${ }_{H_{1}}^{H_{0}} \hat{T}$ where ${ }_{H_{1}}^{H_{0}} \tilde{T}$ is calculated from two robot hand location, and ${ }_{H_{1}}^{H_{0}} \hat{T}$ estimated from camera motions. Residual error rotations such as $R_{e}={ }_{H_{1}}^{H_{0}} \tilde{R}_{H_{0}}^{H_{1}} \hat{R}$. The rotation error can be expressed as

$$
\begin{equation*}
O_{r o t}= \pm \arccos \left(\frac{\left.\operatorname{trace}\left(R_{e}\right)-1\right)}{2}\right) \tag{27}
\end{equation*}
$$

And the metric for translation error is expressed as

$$
\begin{equation*}
O_{\text {tranl }}=\frac{\left\|{ }_{H_{1}}^{H_{0}} \tilde{t}-{ }_{H_{1}}^{H_{0}} \hat{t}\right\|+\left\|{ }_{H_{0}}^{H_{1}} \tilde{t}-{ }_{H_{0}}^{H_{1}} \hat{t}\right\|}{2} \tag{28}
\end{equation*}
$$

From multiple pair of hand-cam motion, the error is measured as the variance

$$
\begin{align*}
\sigma_{\text {rot }}^{2} & =\frac{1}{N} \sum_{i=1}^{N} O_{r o t}^{2}  \tag{29}\\
\sigma_{\text {transl }}^{2} & =\frac{1}{N} \sum_{i=1}^{N} O_{\text {transl }}^{2} \tag{30}
\end{align*}
$$

Precision of our system is evaluated on both simulated and real data. For simulated data, there is a hand-eye transformation ground-truth that is obviously used for evaluating. There is no ground-truth for real data, so it is necessary to measure the distribution of estimation.

### 5.2. Camera Calibration Evaluation

At a fixed hand position, we captured calibration-block images 20 times, did calibration, and then measured the distribution the intrinsic and extrinsic parameters. In addition, we measured the distribution of re-projection error which was the distance between observation and projected points. The result is summarized in Table 2.

Table 2. The accuracy of camera intrinsic parameters

|  | Camera Calibration (pixel) |
| :--- | :---: |
| Focal length $f_{u}$ | $2544.05 \pm 2.07$ |
| Focal length $f_{v}$ | $2554.3 \pm 2.14$ |
| Skew | $-4.7 \pm 0.13$ |
| Principal $C_{u}$ | $599.5 \pm 1.9$ |
| Principal $C_{v}$ | $500.3 \pm 1.6$ |
| Re-projection error | $0.26 \pm 0.04$ |

Each row of in Table 2 includes the mean and variance. The first two rows indicate that the variance of intrinsic parameters were less than 2 pixels. Finally, we measured the extrinsic parameters from calibration-block to camera ${ }_{W}^{C} T$ by measuring their distribution.

Table 3. The accuracy of camera extrinsic parameters

|  | $\mathbf{R}_{\mathbf{x}}$ (deg) | $\mathbf{R}_{\mathbf{y}}$ (deg) | $\left.\mathbf{R}_{\mathbf{z}} \mathbf{( d e g}\right)$ | $\mathbf{T}_{\mathbf{x}}(\mathbf{m m})$ | $\mathbf{T}_{\mathbf{y}}(\mathbf{m m})$ | $\mathbf{T}_{\mathbf{z}}(\mathbf{m m})$ |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $\mathbf{C - W}$ | $103.7 \pm 0.07$ | $-50.9 \pm 0.03$ | $168.3 \pm 0.07$ | $27.5 \pm 0.61$ | $-44.7 \pm 0.53$ | $810.4 \pm 0.71$ |

Additionaly, we make a comparison our approach to the one in [4] which used the planar chessboard instead of using calibration-block as we used by measuring the reprojection errors. For method in [4], we measured the reprojection with different number of images. For our method we use only one image. The results shown in Fig. 6 indicate that the reprojection error of chessboard's method reduces gradually and converges around 10-20 images. With only one image, the reprojection error of our method around 0.26 pixel is equivalent to the reprojection error of chessboard's method at around 6 and 7 images.


Fig. 6. Reprojection errors of two methods: 1) The conventional method using multiple planar chessboard. The proposed method using 3D calibration-block

### 5.3. Hand-Eye Calibration Evaluation

Hand-eye calibration results were compared to other methods on simulation and the real data. Firstly, we made comparison the proposed algorithm to the traditional methods listed as follows: 1) using dual-quaternions and Kronecker product proposed in [10], 2) quaternion presentation proposed by Dormaika [8], 3) Kronecker product proposed by Shah [15] and 4) solving homogeneous transformation equations [16] .These methods focused on solving simultaneous hand-eye/robot-world calibration in general case that means the camera/hand motions are randomly chosen. Because of that our proposed method took advantage of pure motions in rotation and translation to simplify the handeye calibration process. For that reason, the popular hand-eye dataset with random hand motions is not suitable for evaluating our method. To fairly compare our proposed algorithm to four methods mentioned above, the simulation dataset and real dataset were generated with only pure rotation and translation motions. Note that the inputs of our method are hand motion and camera motions while that of other methods are camera motion and transformation from the end-effector to the robot base. The hand motions were easily converted to that type of transformation by integrating with transformation of a fixed end-effector position in the robot base coordinate. As mentioned in [17], methods in [10] and [15] show good performance. We measured rotation and translation error and the difference of estimated value to the ground-truth. Fig. 7 shows the results on simulation data. Horizontal axis represents the number of hand motions, each robot hand position called a station. For our method, the number of pure translation and pure rotation motion is similar that means if there are total $N_{\text {sim }}$ motions then the number of pure translation is $N_{s i m} / 2+1$ and number of pure rotation is $N_{s i m} / 2$. This assignation is applied for all evaluations in experiment.


Fig. 7. Comparison rotation (left) and translation (right) error of the proposed algorithm to the popular approaches on simulation data

Compare to other methods, the proposed approach has stable rotation error even though the number of station is smaller. The errors of other methods are reduced since number of stations increases. However, the translation of proposed method is smallest compare to the others at almost every station.
Fig. 8 proves that the 6DOF estimated by proposed method is closest to the ground-


Fig. 8. Comparison 6DOF estimation of the proposed algorithm, conventional approaches to the ground-truth on simulation


Fig. 9. Comparison rotation (left) and translation (right) error of proposed algorithm to popular approaches on real environment
truth while other methods are more sensitive to the noise.
For the real data, we follow two statistics to evaluate the proposed method: 1) Compare the accuracy to other methods by a small set of data, and 2) Evaluate detail the accuracy and precision on a large data set including 60 pure translations and 50 pure rotations.

In order to compare the accuracy of proposed hand-eye calibration with other methods, we used a real data set including 11 pure translations and 10 pure rotations to evaluate the two erroneous metric of rotation and translation described in (29) and (30). The horizontal is the number of stations. If $N_{\text {real }}$ is the station number then there is are $N_{\text {real }} / 2+1$ pure translations and $N_{\text {real }} / 2$ pure rotation. Fig. 9 shows the comparison results, which indicate that our method has higher rotation error than others but has less translation error than others. Especially, the proposed method could provide stable result even though the number of stations is smaller. This small error is obtained due to the benefit in using pure rotation and pure translation hand motions.

Note that, in hand-eye calibration process time for controlling robot is far longer than that for running algorithm. Instead measuring the algorithm running time, we can use the number of camera/hand motion to evaluate the efficiency of methods. The translation and rotation error results in Fig. 7 and Fig. 9 indicate our errors with 45 stations are equivalent to other method errors at 812 stations. Means that our proposed method is more efficient to others


Fig. 10. Our system precision, the distribution of rotation and translation components

To evaluate the accuracy and precision of our system in detail, we captured a data containing 60 pure translations and 50 pure rotations. Next, a set with $n T$ pure translations and $n R$ pure rotations was selected randomly and repeated 100 times. We measured the distribution of estimated parameters. Firstly, the precision of our hand-eye system with $n T=8, n R=8$ is shown, for instance, in Fig. 10. And Table 4 shows standard deviation of hand eye transformation with different number of pure translations $n T$ and pure rotations $n R$.

Additionally, we also measured accuracy by measuring the distribution of rotation and translation error with different number of translations and rotations. Table 4 shows the statistic the rotation and translation of hand eye calibration with different number of translations and rotations repeated 100 times. It is visualized by Fig. 10, and indicates that the error of our system is around 0.8 degree in rotation and 4 mm in translation. This accuracy is sufficient requirement for several applications such as object recognition and pose estimation or point cloud registration.

Both translation and rotation errors reduce since the number of stations increases, however, it takes more time to control arm and do calibration. Because of that reason we need to trade-off between required accuracy and the number of hand motion.

Table 4. Accuracy of our system. Standard deviation of hand-eye transformation

| $n T$ | $n R$ | Total | $R_{x}$ | $R_{y}$ | $R_{z}$ | $T_{x}$ | $T_{y}$ | $T_{z}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 2 | 2 | 4 | 0.86 | 0.48 | 0.92 | 9.60 | 12.10 | 10.71 |
| 3 | 2 | 5 | 0.85 | 0.47 | 0.89 | 9.27 | 11.78 | 10.31 |
| 3 | 3 | 6 | 0.86 | 0.45 | 1 | 6.94 | 8.11 | 7.01 |
| 4 | 3 | 7 | 0.83 | 0.41 | 0.9 | 8.13 | 7.26 | 6.71 |
| 4 | 4 | 8 | 0.78 | 0.39 | 0.79 | 7.3 | 6.66 | 6.2 |
| 5 | 4 | 9 | 0.72 | 0.39 | 0.69 | 8.12 | 6.93 | 5.36 |
| 5 | 5 | 10 | 0.68 | 0.41 | 0.85 | 6.94 | 6.39 | 5.34 |
| 6 | 5 | 11 | 0.58 | 0.36 | 0.64 | 6.48 | 6.01 | 5.13 |
| 6 | 6 | 12 | 0.64 | 0.34 | 0.76 | 6.25 | 5.54 | 4.86 |
| 7 | 6 | 13 | 0.68 | 0.35 | 0.68 | 5.41 | 5.72 | 4.81 |
| 7 | 7 | 14 | 0.61 | 0.33 | 0.58 | 5.13 | 5.78 | 4.68 |
| 8 | 7 | 15 | 0.51 | 0.27 | 0.6 | 4.89 | 4.96 | 3.66 |
| 8 | 8 | 16 | 0.53 | 0.3 | 0.59 | 5.53 | 3.97 | 3.94 |
| 9 | 8 | 17 | 0.47 | 0.28 | 0.55 | 5.43 | 4.43 | 3.55 |
| 9 | 9 | 18 | 0.48 | 0.24 | 0.54 | 4.58 | 4.56 | 4.29 |
| 10 | 9 | 19 | 0.37 | 0.2 | 0.42 | 4.76 | 4.68 | 3.94 |
| 11 | 10 | 20 | 0.35 | 0.23 | 0.4 | 4.66 | 4.36 | 3.52 |
| 11 | 10 | 21 | 0.39 | 0.22 | 0.39 | 4.21 | 3.94 | 3.17 |



Fig. 11. The rotation (left) and translation (right) errors along the number of stations

## 6. Conclusions

We proposed a calibration solution for a robot vision system with a camera mounted on robot hand. Evaluation on simulation indicates that the estimation of hand-eye transformation of the proposed method is closest to the ground-truth than others even though with higher rotation error and smaller translation error. The results on real system with robot hand 1.0 mm repeatability indicates that the accuracy of our solution is 0.8 degree in rotation and 4.0 mm . Experiment on simulation and real data also indicates that our proposed algorithm can work with small number of stations.

## References

[1] J. Kim, H. H. Nguyen, Y. Lee, and S. Lee, "Structured light camera base 3d visual perception and tracking application system with robot grasping task," in 2013 IEEE International Symposium on Assembly and Manufacturing (ISAM). IEEE, 2013, pp. 187-192.
[2] S. Levine, P. Pastor, A. Krizhevsky, J. Ibarz, and D. Quillen, "Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection," The International Journal of Robotics Research, vol. 37, no. 4-5, pp. 421-436, 2018.
[3] K. Pachtrachai, M. Allan, V. Pawar, S. Hailes, and D. Stoyanov, "Hand-eye calibration for robotic assisted minimally invasive surgery without a calibration object," in 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2016, pp. 2485-2491.
[4] Z. Zhang, "Flexible camera calibration by viewing a plane from unknown orientations," in Proceedings of the seventh ieee international conference on computer vision, vol. 1. Ieee, 1999, pp. 666-673.
[5] J. Heikkila and O. Silven, "A four-step camera calibration procedure with implicit image correction," in Proceedings of ieee computer society conference on computer vision and pattern recognition. IEEE, 1997, pp. 1106-1112.
[6] J. Weng, P. Cohen, M. Herniou et al., "Camera calibration with distortion models and accuracy evaluation," IEEE Transactions on pattern analysis and machine intelligence, vol. 14, no. 10, pp. 965-980, 1992.
[7] K. H. Strobl and G. Hirzinger, "Optimal hand-eye calibration," in 2006 IEEE/RSJ international conference on intelligent robots and systems. IEEE, 2006, pp. 4647-4653.
[8] F. Dornaika and R. Horaud, "Simultaneous robot-world and hand-eye calibration," IEEE transactions on Robotics and Automation, vol. 14, no. 4, pp. 617-622, 1998.
[9] M. Li and D. Betsis, "Head-eye calibration," in Proceedings of IEEE International Conference on Computer Vision. IEEE, 1995, pp. 40-45.
[10] A. Li, L. Wang, and D. Wu, "Simultaneous robot-world and hand-eye calibration using dual-quaternions and kronecker product," International Journal of Physical Sciences, vol. 5, no. 10, pp. 1530-1536, 2010.
[11] H. Sung, S. Lee et al., "A robot-camera hand/eye self-calibration system using a planar target," in IEEE ISR 2013. IEEE, 2013, pp. 1-4.
[12] A. M. Andrew, "Multiple view geometry in computer vision," Kybernetes, 2001.
[13] R. Y. Tsai, R. K. Lenz et al., "A new technique for fully autonomous and efficient 3 d robotics hand/eye calibration," IEEE Transactions on robotics and automation, vol. 5, no. 3, pp. 345-358, 1989.
[14] J. C. Chou and M. Kamel, "Finding the position and orientation of a sensor on a robot manipulator using quaternions," The international journal of robotics research, vol. 10, no. 3, pp. 240-254, 1991.
[15] M. Shah, "Solving the robot-world/hand-eye calibration problem using the kronecker product," Journal of Mechanisms and Robotics, vol. 5, no. 3, 2013.
[16] H. Zhuang, Z. S. Roth, and R. Sudhakar, "Simultaneous robot/world and tool/flange calibration by solving homogeneous transformation equations of the form $\mathrm{ax}=\mathrm{yb}$," IEEE Transactions on Robotics and Automation, vol. 10, no. 4, pp. 549-554, 1994.
[17] I. Ali, O. Suominen, A. Gotchev, and E. R. Morales, "Methods for simultaneous robot-world-hand-eye calibration: A comparative study," Sensors, vol. 19, no. 12, p. 2837, 2019.

Manuscript received 20-2-2020; Accepted 14-5-2020.

Nguyen Huu Hung received his Ph.D. degree at Sungkyunkwan University, South Korea in
 computer vision, in 2020. He is currently researcher at Institude of System Integration, Le Quy Don Technical University. His research interests include computer vision, simultaneous localization and mapping (SLAM), 3D point cloud processing, deep learning and AI.


Nguyen Quang Thi received his Ph.D. degree at Changchun University of Science and Technology, China in Communication and Information System in 2014. He is currently lecturer/researcher at Institute of System Integration, Le Quy Don Technical University, Vietnam. His research interests include computer vision, blind deconvolution, image processing and pattern recognition.


Tran Cong Manh got his master-degree in computer science from Le Quy Don Technical University of Vietnam in 2007. In 2017 Manh got his PhD degree from Department of Computer Science, National Defense Academy, Japan. His current research interests include network security, intelligent computing, and data analysis. Currently, Dr. Manh works as a researcher in Le Quy Don Technical University, Hanoi, Vietnam.

## PHƯƠNG PHÁP ĐỒNG THỜI HIỆU CHỈNH THÔNG SỐ CỦA CAMERA VÀ TAY-MẮT

## Tóm tắt

Xác định vị trí và hướng của các vật thể trong không gian làm việc của robot là một chức năng quan trọng của các hệ thống robot tự động hóa. Vấn đề này được giải quyết bằng cách sử dụng camera gắn trên cánh tay robot. Trong trường hợp này, một giải pháp khả thi để hiệu chỉnh hệ thống là cần thiết, bao gồm hiệu chuẩn các thông số camera cuñg như hiệu chỉnh các tham số ma trận dịch chuyển của hệ thống robot-camera. Trong bài báo này, chúng tôi trình bày một phương pháp hiệu chỉnh hệ̀ thống camera-on-hand một cách đơn giản và hiệu quả, trong đó các thông số của máy ảnh cũng như các thông số của ma trận chuyển vị được hiệu chỉnh đồng thời. Phương pháp này dựa trên việc kết hợp các phép chiếu 3D sang 2 D của khối hiệu chuẩn với tối thiểu hai phép tịnh tiến và hai phép quay thuần túy của chuyển động cánh tay. Phương pháp đề xuất dược đánh giá trên dữ liệu mô phỏng và hệ thống camera-on-hand thật cho thấy rẳng phương pháp của chúng tôi có thể hoạt động ổn định với nhiều mức độ nhiễu và số lượng vị trí robot khác nhau.


[^0]:    ${ }^{1}$ Institute of System Integration, Le Quy Don Technical University

