

Eye to hand calibration using ANFIS for stereo vision-based object manipulation system

Taryudi¹ · Ming-Shyan Wang¹

Received: 17 January 2017 / Accepted: 1 February 2017 / Published online: 8 February 2017
© Springer-Verlag Berlin Heidelberg 2017

Abstract Standard industrial robot manipulator has possibility to manipulate the objects accurately in arbitrary position by using additional vision sensor. This applies in eye-to-hand camera configuration, where the stereo camera can cover the robot arm workspace and the object can be kept accurately for three dimension position estimation. The goal of this system is to control a robot arm to pick up and place the targeted object autonomously within its workspace. In order to accomplish this goal, a stereo vision technique including eye to hand calibration is applied. First, a stereo camera is calibrated to obtain the intrinsic and extrinsic camera parameters. Then, the specific color features of desired object are extracted using image processing algorithms. After that, based on centroid of desired object, 3D (3 dimension) position of the desired object is calculated. Then, camera and robot arm coordinate transformation was calibrated using adaptive artificial network based fuzzy interference system (ANFIS) method. Finally, the 3D coordinates of desired object are sent to robot arm controller as the inputs to move the end-effector equipped with gripper to the object location and pick up the object then place it to the desired location. The experiment was carried out using 6 DOF industrial robot arm and the results demonstrated the good performance of the system with acceptable errors.

List of symbols

f	Focal length of the camera, pixel
α, β	Scale factors in image u and v axes, pixel
γ	Skewness of the image axis
u_0, v_0	Coordinate of the principal point, pixel
b	Distance between two cameras, mm
d	Disparity
P	A world point
p	An image plane point
A	A matrix
k_u	Column-wise density, pixel/mm
k_v	Row-wise density, pixel/mm
u, v	Image plane coordinate
O_i	Projection center
A_i, B_i	Linguistic variables
p_i, q_i, r_i	Linear parameter set of ANFIS
x, y	ANFIS input nodes
z	ANFIS output node
ω_i	Firing strength
ϖ_i	Normalized firing strength
c_i, s_i	Gaussian membership function parameter set
${}^B\xi_E$	End-effector coordinate frame with respect to robot base frame
${}^E\xi_G$	Gripper coordinate frame with respect to end-effector frame
${}^C\xi_T$	Targeted object coordinate frame with respect to camera frame
${}^B\xi_T$	Targeted object coordinate frame with respect to robot base frame
${}^B\xi_C$	Camera coordinate frame with respect to robot base frame
(X, Y, Z)	3D world Cartesian coordinate
(X_r, Y_r, Z_r)	3D robot arm Cartesian coordinate
(X_c, Y_c, Z_c)	3D camera Cartesian coordinate
(x, y, z)	3D object Cartesian coordinate

✉ Ming-Shyan Wang
mswang@stust.edu.tw

Taryudi
taryud@yahoo.com

¹ Department of Electrical Engineering, Southern Taiwan University of Science and Technology, Tainan, Taiwan

R	An orthonormal rotation matrix (3×3)
t	Translation matrix (3×1), mm

1 Introduction

In the future, most of the operations in factory will be done by autonomous robots that need visual feedback to work within the workspace (Pérez et al. 2016). However, a standard industrial robot manipulator has not yet equipped with the visual sensor in term of the capability improvement to manipulate objects in arbitrary location accurately. Many types of vision techniques which can be applied in the related system were introduced in (Pérez et al. 2016; Wilson 2016), but stereo vision is the commonly used due to safety, wide range and accuracy among other vision techniques. Stereo vision is an imaging technique for recovering depth from camera images by comparing two views of the same scene (Borangiu and Dumitrache 2010; Corke 2011; Fevery et al. 2010). It has been applied in industrial automation and three-dimension (3D) machine vision applications to perform tasks such as bin picking, volume measurement, and 3D object location and identification (Nof 2009; Pérez et al. 2016; Point_Grey 2015). This technique makes robots be more flexible of position control capability in tracking and grasping the unknown object in arbitrary location (Borangiu and Dumitrache 2010). One of important advantages of stereo vision is non-invasive technique with respect to the surrounding environment due to it does not require additional light sources (Borangiu and Dumitrache 2010). Another advantage of stereo vision over another methods is the extracting the relative depth information among the passive methods (Kheng et al. 2010; Pérez et al. 2016).

In order to apply the stereo vision system in desired control application, it often requires coordinate transformation between stereo camera and robot arm, namely eye to hand calibration. The eye to hand calibration is the task of computing the relative 3D position and orientation between the camera and the robot arm such as in eye to hand camera configuration (Miseikis et al. 2016). However, this calibration is difficult to be determined due to the non-linear relationship (Daniilidis 1999; Dornaika and Horaud 1998; Tao 2015; Wu et al. 2014). Previous researchers used a mathematical approach for calibration such as series of generic geometric (Tsai and Lenz 1989) and closed form methods (Dornaika and Horaud 1998; Tsai and Lenz 1989). However, those techniques are limited by long computation time. In the current work, many researchers have utilized artificial intelligence (AI) approach which adopts human brain behavior such as fuzzy logic and neural network in order to reduce long computation.

Utilization of neural networks and fuzzy logic for solving the camera to robot arm coordinate transformation are reported in (Jafari and Jarvis 2004; Juang et al. 2015; Wu et al. 2014). Neural network does not require prior knowledge, however it needs more sufficient data training and several learning algorithms. On the contrary, a fuzzy logic requires linguistic rules (If–Then rules) instead of learning examples as prior knowledge and this system is not capable to learn (Abe 2012). There exist numerous possibilities for the fusion of neural networks and fuzzy logic technique so that both of them can overcome their individual drawbacks as well as get benefits from each other's merits.

This paper presents a method of integrating the measuring functions of a 3D binocular stereo vision system into an industrial robot system to manipulate the targeted object. We consider the ANFIS structure with first order Sugeno model that contains 343 rules. Gaussian membership functions with product inference rule are used at the fuzzification level. Further, to adjust the parameters of membership functions, we used hybrid learning algorithm that combines least square and gradient descent methods (Jang 1993). The ANFIS controller developed consists of three inputs of 3D object position coordinate (X_c, Y_c, Z_c) in camera coordinate frame which is obtained by stereo vision system and three outputs of 3D object position coordinate (X_r, Y_r, Z_r) in robot frame.

The study deals with the design techniques and procedures related to vision system setup, stereo camera calibration, camera to robot coordination calibration, and system performance analysis. The rest of this paper is structured as follows. Section 2 is the summary of the stereo vision system including stereo camera calibration, object feature extraction and pose estimation. Coordinate transformation of camera to robot arm and training data of ANFIS are also described. Section 3 explains the experiment results and discussion. Finally, a brief conclusion is presented in Sect. 4.

2 Eye to hand calibration for stereo vision-based object manipulation system

In this study, the stereo vision based object manipulation system with eye to hand calibration using ANFIS is shown in Fig. 1. Personal computer (PC) is connected to Robot arm controller via RS232 serial communication interface to deliver commands and receive the responses of control robot arm movement process. While the gripper and the stereo camera is connected to PC through USB serial interface. PC is also used for graphic user interface development to control robot arm, gripper and stereo camera. Stereo camera is made of two identical cameras, Logitech C310, aligned in y axis and distance apart in x axis. Stereo camera

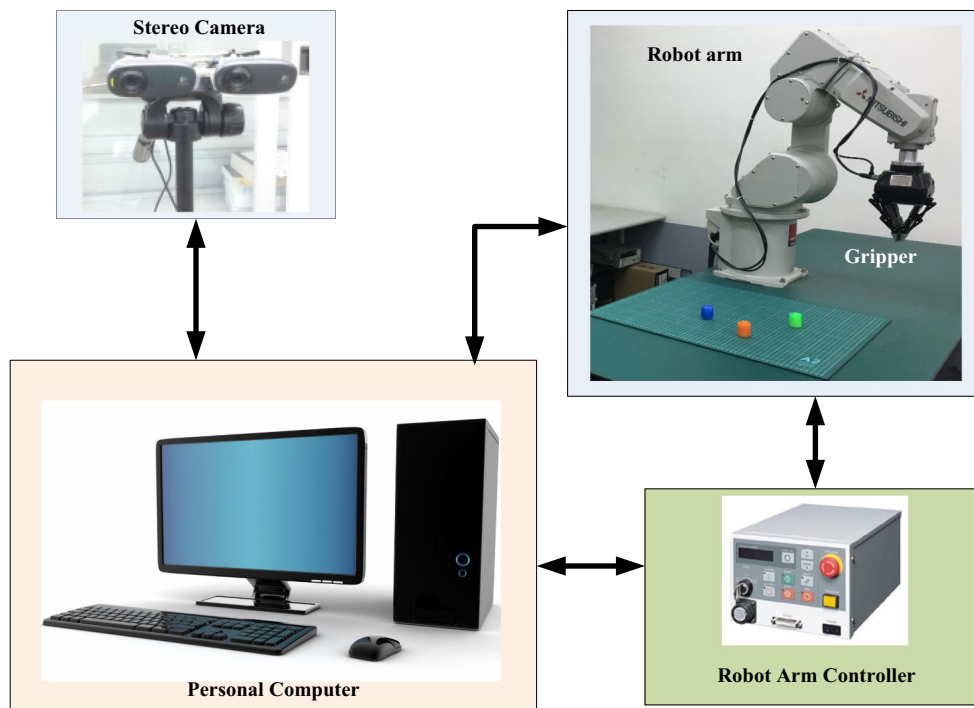


Fig. 1 Stereo vision based object manipulation system architecture

is used as a vision sensor to capture the object in 3D world coordinate and then the feature extraction of the object is obtained to calculate the pose estimation using image processing algorithm. The robot arm controller is programmed to receive the commands from PC and drive the robot arm to move according to the desired position, and also send the position data of the end-effector of robot arm to the PC as feedbacks.

6-DOF robot arm is driven by robot controller to move to the estimated object position, and then robot controller reads the position of end effector of robot arm by requesting command from PC. Therefore, when the position is achieved, the gripper grasps the object by command from PC and continues the tasks. In this paper, the stereo vision system and Eye to Hand calibration using ANFIS with the detailed description of each part of the system is described in following sections.

2.1 Stereo vision system

Stereo vision attempts to compute the 3D data in a way similar to the human brain. A 3D binocular stereo vision system uses two cameras which capture images of the same scene from different positions, and then calculates the 3D coordinates for each pixel by comparing the parallax shifts between the two images (Borangiu and Dumitrache 2010; Corke 2011; Jiadi et al. 2014; Point_Grey 2015).

In order to use the cameras in stereo vision system, the knowledge of camera model and its parameters are important. The projection of an object with respect to the pinhole camera system is described in Fig. 2. In Fig. 2, the coordinate vector of a 3D point $P = [X, Y, Z]^T$ is projected in 2D camera image plane coordinates as $p = [x, y]^T$ and from the comparison of similar triangles can be calculated through Eq. (1). The parameter f represents the focal length of the camera.

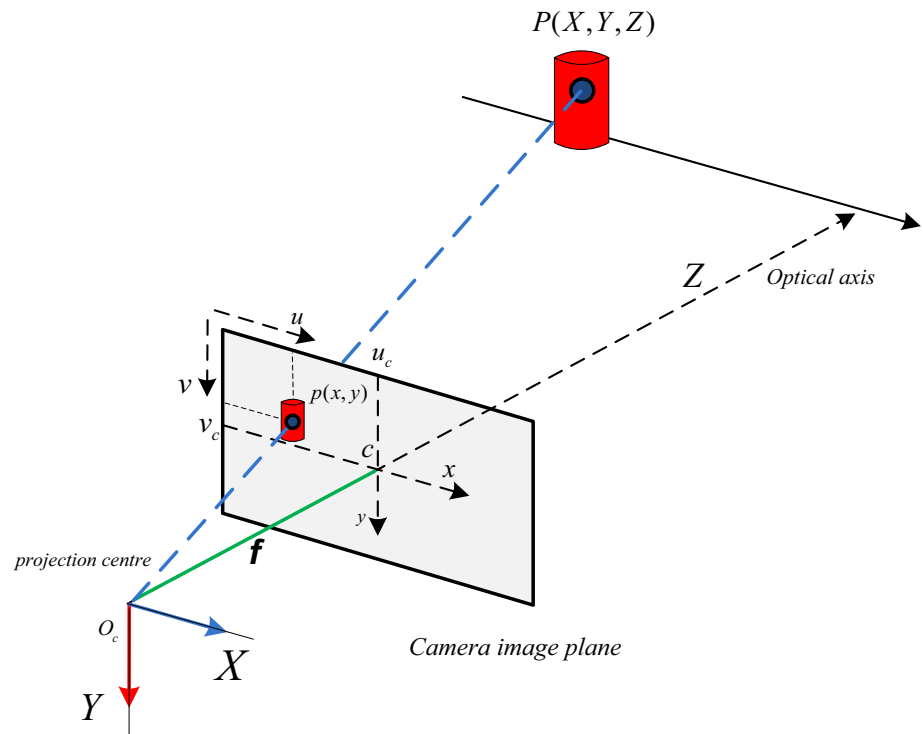
$$x = f \frac{X}{Z}; \quad y = f \frac{Y}{Z}. \tag{1}$$

Since, the image coordinates are measured in pixels, while the spatial coordinates are in millimeters. This means that equations must be obtained to convert units between these two measurement systems.

$$u = k_u(x + x_0) = k_u f \frac{X}{Z} + k_u x_0,$$

$$v = k_v(y + y_0) = k_v f \frac{Y}{Z} + k_v y_0,$$

where u and v are the number of pixels, k_u and k_v are column-wise and row-wise density of pixels, respectively, measured as number of pixels per millimeter. The relationship between the world reference frame and image frame of projection 3D point P' can be formulated as:

Fig. 2 Pinhole camera model

$$p' = A[R|t]P', \quad p'^T = [p'^T \ |s], \quad t^T = [t_1 \ t_2 \ t_3], \quad P'^T = [P'^T \ |1], \quad (2)$$

or

$$\begin{bmatrix} x \\ y \\ s \end{bmatrix} = \begin{bmatrix} \alpha & \gamma & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}. \quad (3)$$

In Eq. (3), where $\alpha = f/k_u$ and $\beta = f/k_v$ are the focal lengths in horizontal and vertical pixels, respectively [f is the focal length in millimeter, and (k_u, k_v) are the pixel size in millimeter], (u_0, v_0) are the coordinates of the principle point, and γ is the skew factor that models non-orthogonal $u-v$ axes. Since (x, y, s) is homogeneous, the pixel coordinates x and y can be retrieved by dividing the scale factor s .

The intrinsic and extrinsic camera parameters are obtained by camera calibration that is principally important in stereo vision systems and has the crucial role in many computer vision tasks. The accuracy of the estimation of actual object distances is determined by the precision of the camera calibration. Camera calibration is the process of determining the camera's intrinsic parameters and the extrinsic parameters with respect to the world coordinate system as mentioned in (Bouguet 2015; Corke 2011; Nguyen et al. 2015). Intrinsic parameters are the characteristics of the camera, i.e., $(\alpha, \beta, \gamma, u_0, v_0)$ where (u_0, v_0) is the coordinate of the principal point, α and β are the scale factors in image u and v axes, and γ is the parameter describing the skewness of the two image axes. Extrinsic parameters are the orientation and location

of the camera, i.e., (R, t) where R is the rotation and t is the translation of right camera with respect to left camera, respectively (Zhang et al. 2011).

In this study, an object feature extraction based on HSV color space is used to detect the targeted object on the workspace and the pose estimation, described in algorithm on Matlab platform as shown in Fig. 3. First, initialize the stereo camera model and object color threshold adjustment based on HSV space value. Second, capture the image pair with the both cameras at the same time. Next, extract the object from the image based on color using HSV space threshold. Median filtering is applied for noise removing and morphological operation acts like opening and closing to locate the boundary of the object. Next, locate boundary object. Then, the centroid of the located object in both left and right image is calculated. Finally, the 3D location of the object is obtained by applying the triangulation.

The configuration scheme of the stereo vision used in the approach task in this paper is shown in Fig. 4, which is fixed in the work space with two cameras in parallel. The distance of two cameras optical center is b , and they have the same focal length f . Given a reference point $P(X_p, Y_p, Z_p)$, the projection is $p_1(x_1, y_1)$ in image plan 1 and $p_2(x_2, y_2)$ in the image plan 2. Then by perspective projection, we have the image coordinate of P in two image planes and we can see in Fig. 5 to simplify the calculation

$$\frac{X}{x_1} = \frac{Z}{f}; \quad \frac{Y}{y_1} = \frac{Z}{f}; \quad \frac{b-X}{x_2} = \frac{Z}{f}. \quad (4)$$

Fig. 3 Flow chart diagram of object feature extraction and pose estimation

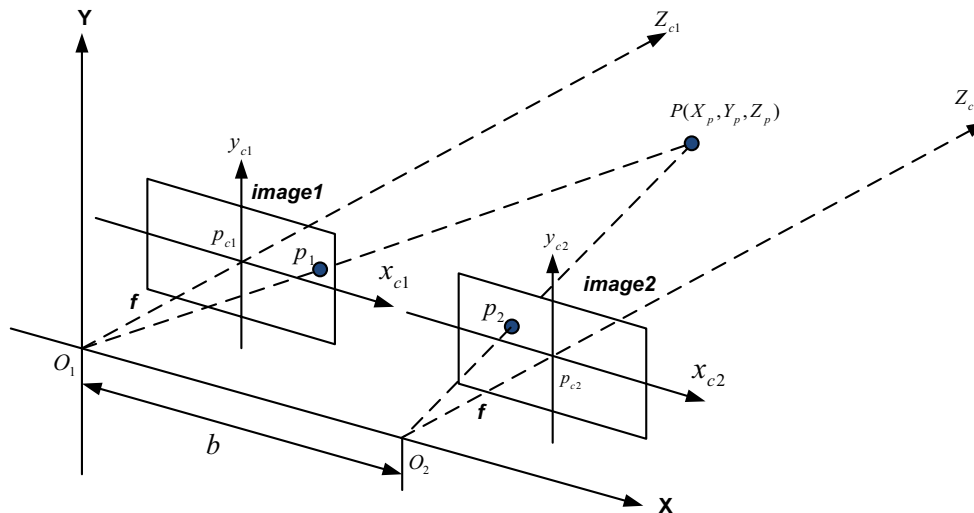
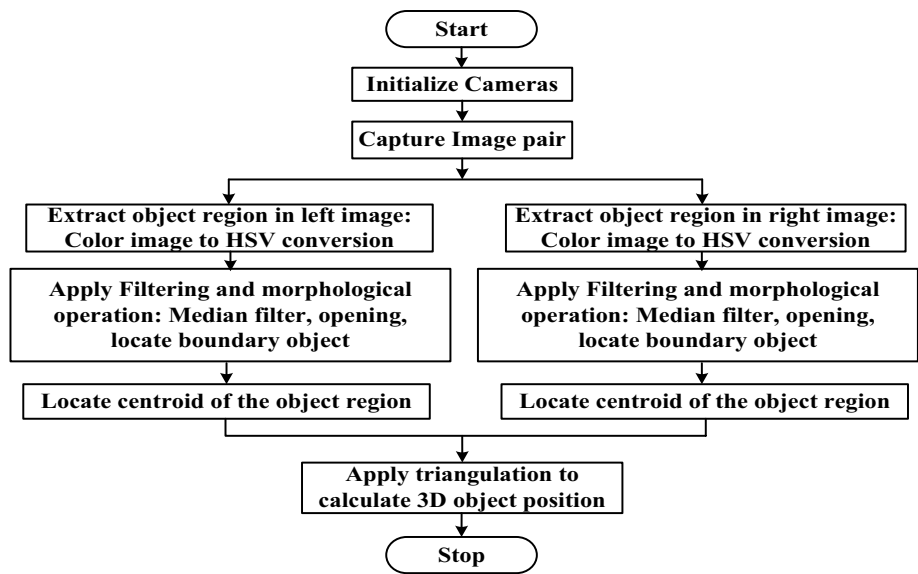


Fig. 4 Configuration scheme of stereo vision

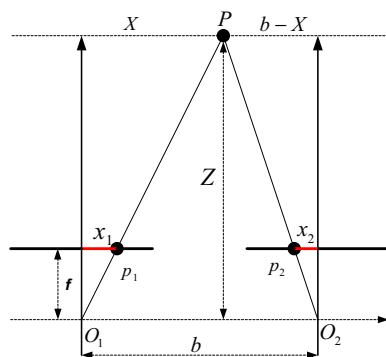


Fig. 5 Triangulation scheme of stereo vision

In Fig. 4, we assume that two cameras have the identify camera parameters, which are obtained by stereo camera calibration in Matlab (Bouquet 2015). Its images on the two cameras are p_1 and p_2 , $d = x_1 + x_2$ is the parallax and Y-axis is perpendicular to the page (Liu and Chen 2009). According to the principle of similar triangles, we can get the Eq. (4). From Fig. 5, we also can see that b can be written in Eq. (5) and Z is the depth of P point obtained by Eq. (6)

$$b = \frac{Z}{f}x_1 + \frac{Z}{f}x_2, \tag{5}$$

$$Z = \frac{b \times f}{x_1 + x_2}, \quad (6)$$

The disparity d is the difference x coordinate in image 1 and image 2 as written in Eq. (7)

$$d = x_1 + x_2. \quad (7)$$

Substituting (7) to Eq. (6), the depth (Z) of point P can be seen in Eq. (8). After Z is obtained, then we can obtain the X and Y coordinates of P point using Eqs. (9) and (10), respectively.

$$Z = \frac{b \times f}{d}, \quad (8)$$

$$X = \frac{Z \times x_1}{f} \quad (9)$$

$$Y = \frac{Z \times y_1}{f}, \quad (10)$$

where x_1 and x_2 are the pixel locations on the 2-D image, and X , Y and Z are the actual positions on 3D image.

2.2 ANFIS based eye to hand calibration

Due to the difference views of the camera and robot arm to the targeted object, it is necessary to know the relative position and orientation between gripper and end-effector, the end-effector and robot base, object and the robot base, between the camera and the robot base, and between the object and the camera. Those coordinate transformation relationships are

shown in Fig. 6. ${}^B\xi_E$ is the coordinate relationship between end-effector to robot base that can be found by using forward kinematic equation of 6 DOF robot arm with considering its DH parameters (Kucuk and Bingul 2006). ${}^E\xi_G$ is the coordinate relationship between end-effector and gripper which is only the translation of z axis from end effector. ${}^C\xi_T$ is the targeted object coordinate with respect to camera coordinate that can be obtained by stereo vision system, ${}^B\xi_T$ is the targeted object coordinate with respect to robot base that can be obtained by using teaching box of robot arm controller to get the object coordinate by pointing the end-effector of robot arm to the desired object position. ${}^B\xi_C$ is the camera coordinate with respect to robot base that will be obtained using ANFIS method in following section.

2.2.1 ANFIS architecture

The ANFIS can perform the mapping relation between the input and output through a learning algorithm to optimize the parameters of a given FIS. The ANFIS architecture consists of fuzzy layer, product layer, normalized layer, defuzzy layer, and summation layer. To give a simple explanation, Fig. 7 shows the structure of a two-input type-3 ANFIS with 4 rules considered, in which a circle indicates a fixed node, whereas a square indicates an adjustable node. For example, we consider two inputs x , y and one output z in the FIS. The ANFIS used in this paper implements a first-order Sugeno FIS. Among many fuzzy systems, the Sugeno fuzzy model is the most widely applied, because of its high interpretability and computational efficiency, and built-in optimal and adaptive techniques.

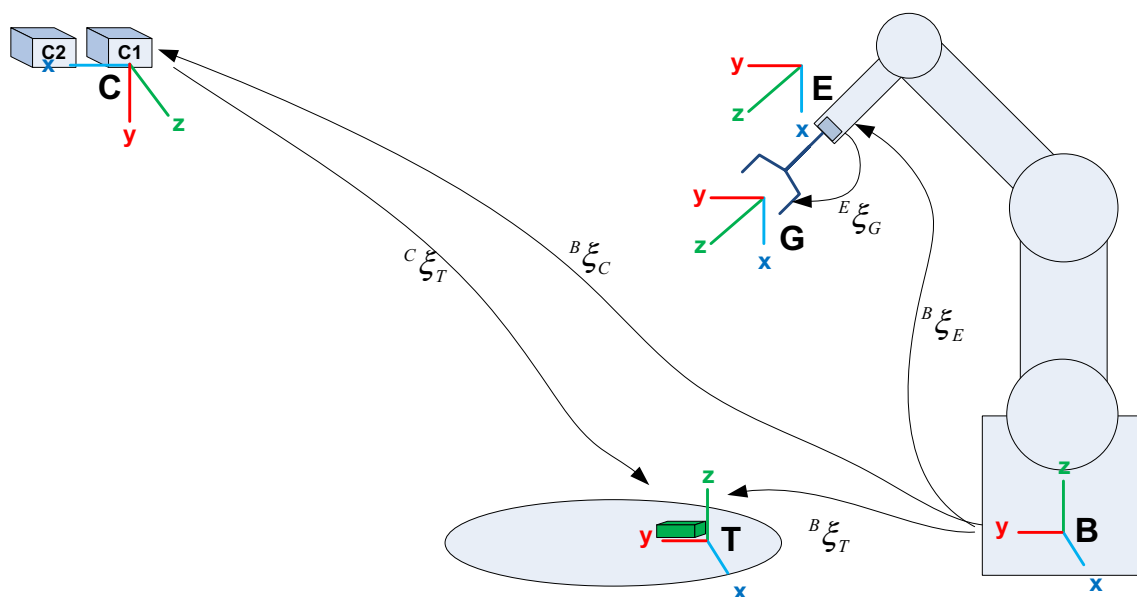


Fig. 6 Coordinate transformation relationship

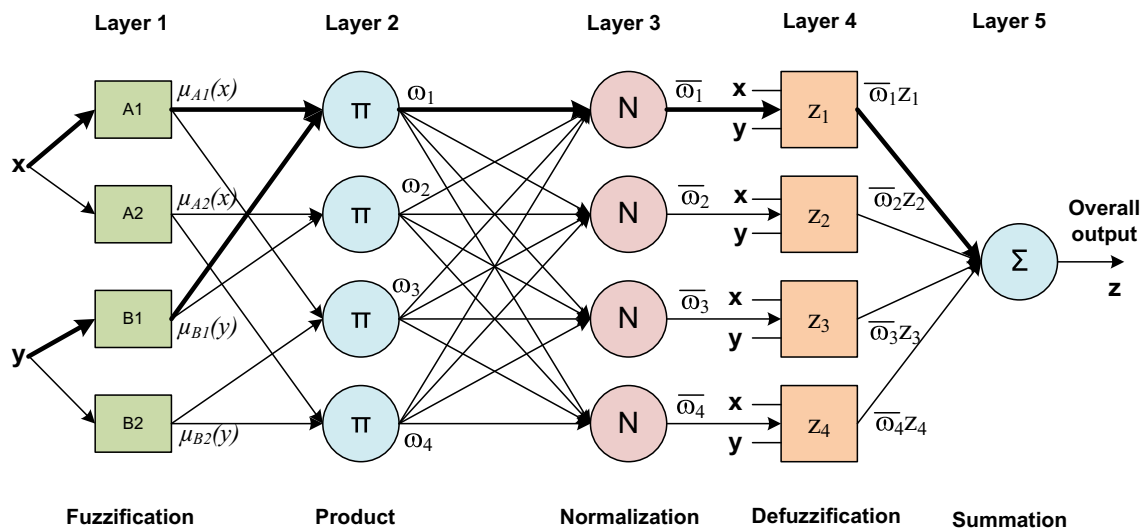


Fig. 7 ANFIS structure

According to Jang (1993), type-3 ANFIS uses Takagi–Sugeno if–then rules of the following form:

Rule 1 : IF x is A_1 and y is B_1 THEN $z_1 = p_1x + q_1y + r_1$,

Rule 2 : IF x is A_2 and y is B_1 THEN $z_2 = p_2x + q_2y + r_2$,

Rule 3 : IF x is A_1 and y is B_2 THEN $z_3 = p_3x + q_3y + r_3$,

Rule 4 : IF x is A_2 and y is B_2 THEN $z_4 = p_4x + q_4y + r_4$ (11)

where x and y are two input variables, $z_i(x,y)$ ($i = 1:4$) is one output variable, A_i and B_i ($i = 1,2$) are linguistic variables that cover the input variable universe of discourse and p_i, q_i and r_i ($i = 1:4$) are linear consequent parameters. The output of each rule is linear combination of input variables and a constant. The typical ANFIS consists of 5 layer structure. The layers and their functions can be described as follows:

Layer 1: fuzzification layer

In this layer, each node denotes the membership function (MF) of fuzzy sets A_i and B_i , ($i = 1, 2$), $\mu_{A_i}(x)$ and $\mu_{B_i}(y)$. The membership value directly depends on the membership function. Since the membership functions have adjustable parameters, this layer can be called adaptive layer. The Gaussian membership function as shown in (12) is used in this study.

$$gaussmf(x, c, s) = e^{-\frac{(x-c_i)^2}{2s_i^2}}, \tag{12}$$

where x is the input and (c_i, s_i) is the parameter set that changes the shape of the MF. The parameters of this layer are termed the premise parameters.

Layer 2: product layer

In this layer, the T-norm operation is used to calculate the firing strength of a rule via multiplication:

$$\omega_i = \mu_{A_i}(x)\mu_{B_i}(y). \tag{13}$$

Layer 3: normalization layer

In this layer the ratio of a rule’s firing strength to the total of all firing strengths is calculated:

$$\bar{\omega}_i = \frac{\omega_i}{\sum_{i=1}^4 \omega_i} = \frac{\omega_i}{\omega_1 + \omega_2 + \omega_3 + \omega_4}. \tag{14}$$

Layer 4: defuzzification layer

In the fourth layer, the linear compound is obtained from the inputs of the system as THEN part of fuzzy rules as:

$$\bar{\omega}_i z_i(x, y) = \bar{\omega}_i(p_i x + q_i y + r_i), \tag{15}$$

where $\bar{\omega}_i$ is the output of layer 3 and $\{p_i + q_i + r_i\}$ is the consequent parameter set.

Layer 5: summation layer

A node in the fifth layer is fixed node that calculates the overall output as the summation of all incoming inputs:

$$z = \sum_{i=1}^4 \bar{\omega}_i z_i(x, y) = \frac{\omega_1 z_1 + \omega_2 z_2 + \omega_3 z_3 + \omega_4 z_4}{\omega_1 + \omega_2 + \omega_3 + \omega_4}. \tag{16}$$

2.2.2 ANFIS structure for computing camera to robot arm calibration

The ANFIS structure for computing the solution of camera to robot arm calibration, is shown in Fig. 8. It consists of three ANFIS networks with first-order Sugeno fuzzy system for three axis (x, y, z) in camera to robot arm 3D coordinate transformation. 3, 5 and 7 Gaussian MFs with

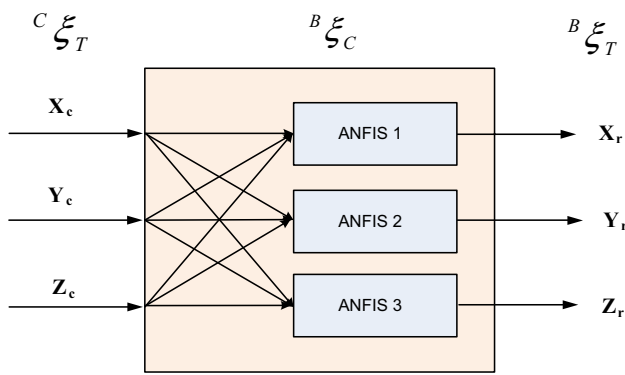


Fig. 8 Proposed ANFIS architecture for computing eye to hand coordinate transformation

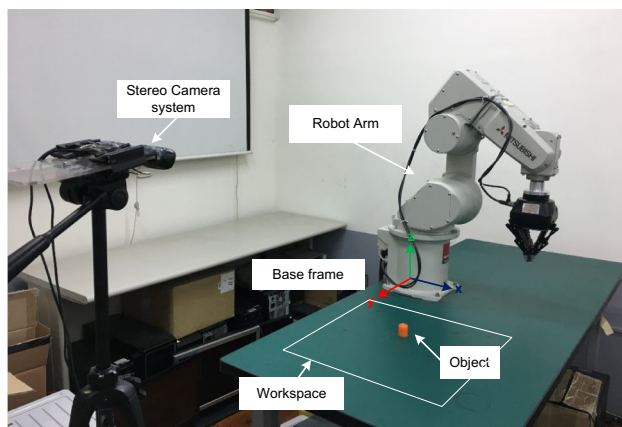


Fig. 9 Experimental setup

product inference rule are used at the fuzzification layer, and hybrid learning algorithm is used to adjust the premise and consequent parameters. In Fig. 8, ${}^C \xi_T$ or (X_c, Y_c, Z_c) is the targeted object coordinate with respect to camera coordinate which is obtained by stereo vision system, ${}^B \xi_T$ or (X_r, Y_r, Z_r) is the targeted object coordinate with respect to robot base which is obtained by positioning the end effector

to the desired object position using teaching box of robot arm controller. ${}^B \xi_C$ is the camera coordinate with respect to robot base that will be obtained by training the ANFIS.

The training data are very important in camera to robot arm position calibration process to get the accurate value of the object 3D position. The training data generation consists of the following steps: First, the object was placed on one of the position which is used in calibration. Second, the object 3D position measurement was obtained by object feature extraction and position estimation process. Third, the 3D position of object was recorded in the file. Fourth, repetition was done from first to third step to get the comparisons of the object position, so as one position will have twice measurement results. After that, repetition process was done from the first to fourth step for the next 3D position object calibration. In this research there were 234 position data calibrated with double for every point.

3 Experimental results

In this part, several experiments are performed to evaluate ANFIS method for eye to hand calibration. Stereo camera calibration is applied prior image processing to get both intrinsic and extrinsic parameters. Then, eye to hand calibration is performed based on ANFIS. Figure 9 depicts the experimental setup. We choose a red cylinder bottle cup as the target object which is placed into a workspace. A pair of Logitech C310 cameras is placed toward the robot arm to build the stereo vision system, each camera works at 640×480 pixels.

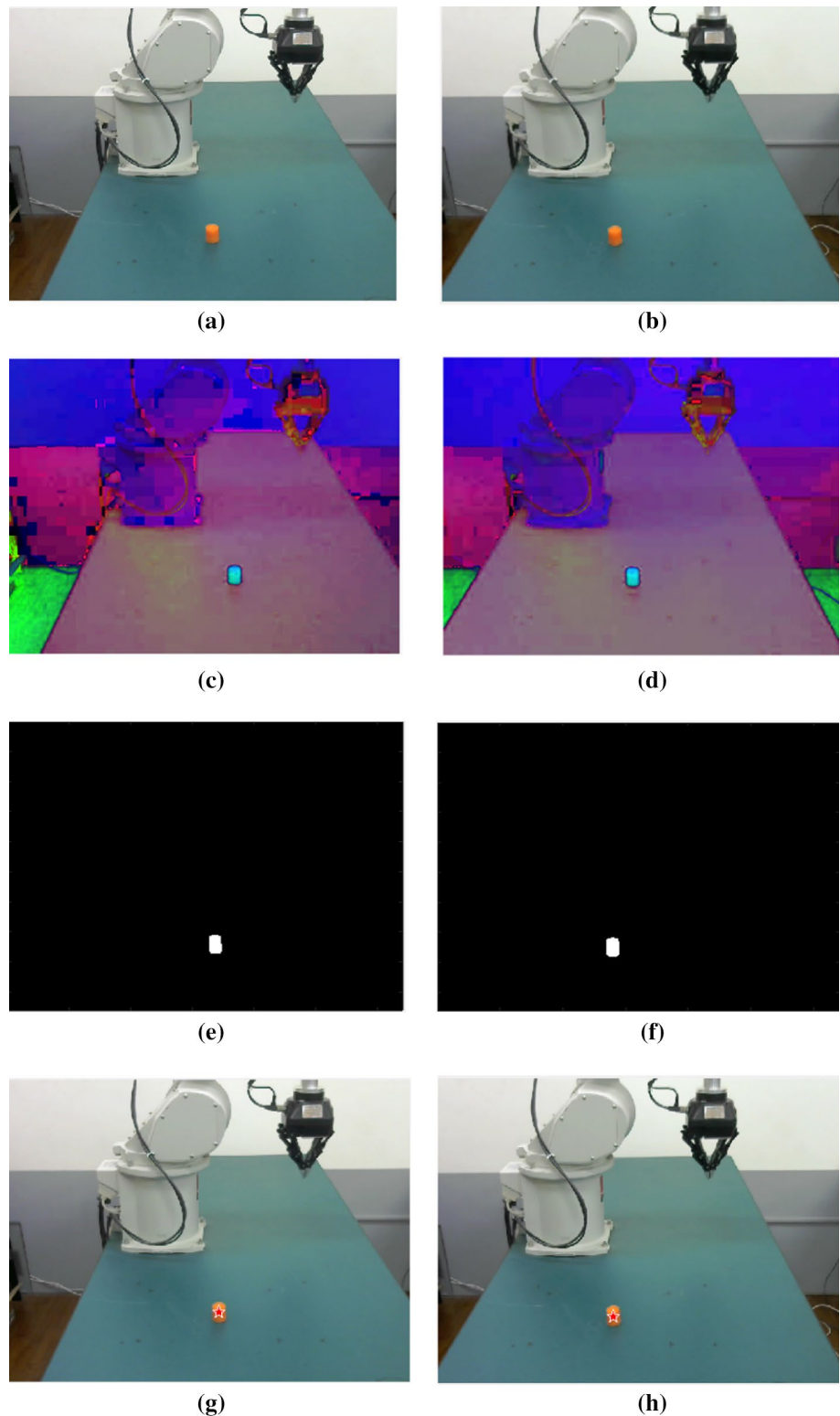
3.1 Experiment 1: stereo camera calibration

Before eye to hand calibration, we performed stereo camera calibration to get the intrinsic and extrinsic camera parameters. In this paper, we utilize a method proposed by (Bouguet 2015) with a classical black-white chessboard to calibrate the cameras. A stereo camera system was set up for baseline of 92 mm between the two cameras, we therefore do calibration. Our chessboard has 63 square-blocks

Table 1 The intrinsic and extrinsic parameters of stereo camera

Intrinsic parameter	Left camera	Right camera
Focal length (pixel)	$\alpha = 791.5897$	$\alpha = 793.8894$
	$\beta = 905.1776$	$\beta = 908.0247$
Principal point (pixel)	$u_0 = 318.6567$	$u_0 = 311.2690$
	$v_0 = 349.4288$	$v_0 = 351.5866$
Skew	$\gamma = -4.6270$	$\gamma = -0.9511$
Extrinsic parameters (position of right camera with respect to left camera)	Rotation	$R = \begin{bmatrix} 0.9999 & 0.0004 & -0.0109 \\ -0.0004 & 1.0000 & -0.0003 \\ 0.0109 & 0.0003 & 0.9999 \end{bmatrix}$
	Translation (mm)	$t = [-91.9891 \quad -1.8890 \quad 2.3283]^T$

Fig. 10 Object feature extraction and pose estimation process. **a** Color image captured by left camera. **b** Color image captured by right camera. **c** HSV color image from left camera. **d** HSV color image from right camera. **e** Filtered image of red color object on left camera. **f** Filtered image of red color object on right camera. **g** Red color object pose estimation. **h** Red color object pose estimation



in a 9×7 pattern, which has a size of 40 mm x 40 mm for each square. Accurate chessboard size is required to give an accurate estimation of object targeted features. For the calibration procedure, the 18-different positions and

orientations of 640×480 pixel images of chessboard are captured by each camera simultaneously and then loaded into Matlab. The corner of chessboard square is detected with subpixel accuracy as the input of the calibration

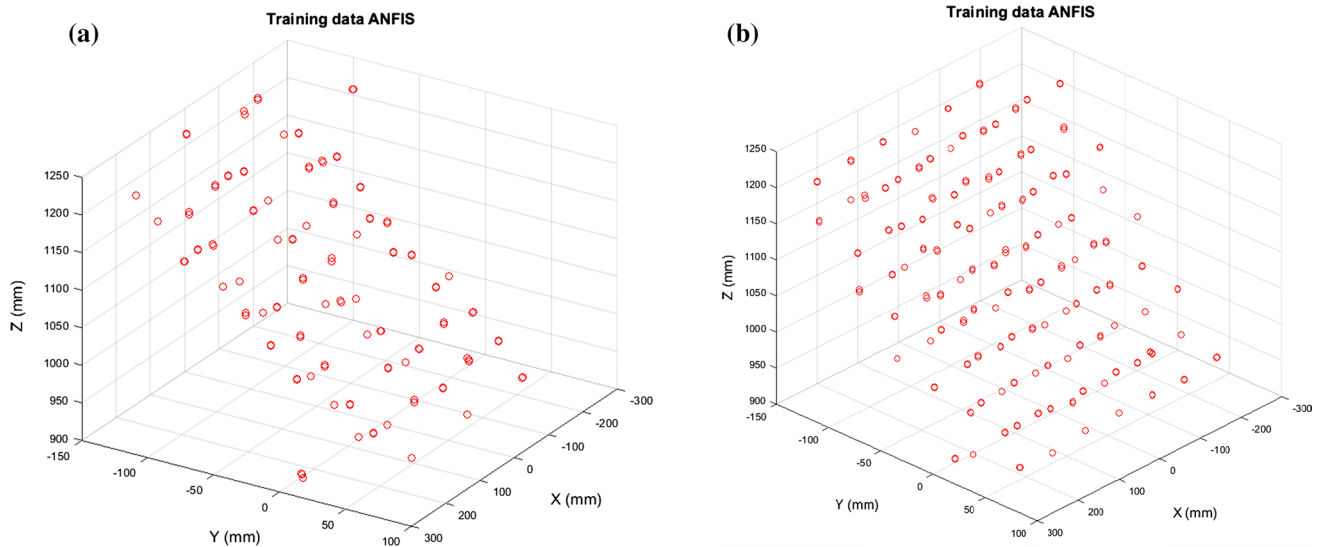


Fig. 11 Input data of ANFIS Training. **a** 138 data recorded. **b** 234 data recorded

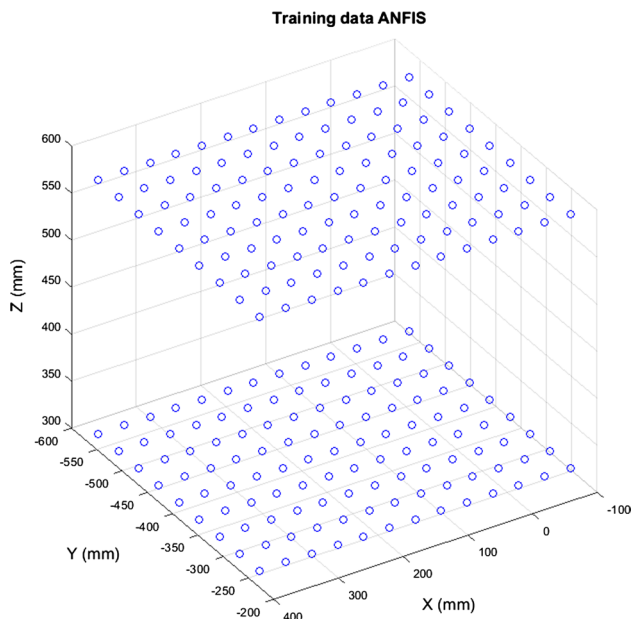


Fig. 12 Output data of training ANFIS

method; and the output includes intrinsic, distortion and extrinsic matrices of the two cameras and the perspective transformation matrix. All those outputs are needed in re-projecting depth information to the real-world coordinates. The results of stereo camera calibration are shown in Table 1.

From Table 1, the intrinsic parameters are slightly similar between right and left camera, including focal length and principal point. All those parameters will be used for triangulation process in the next step. For the extrinsic

parameters, the two-camera rotation matrix R is approximately the unit matrix, which means that there is no rotation, but translation only between two cameras. From the translation vector t , we know that the translation of y-axis and z-axis is small and the distance between two cameras baseline vision system is 92 mm. According to the experiment results, the calibration of stereo camera has been successful with the intrinsic and extrinsic parameters results that can be used in the triangulation process. But the resulted camera parameters are not precisely accurate because of the hand-made stereo camera with the low resolution.

3.2 Experiment 2: object feature extraction and pose estimation

After obtaining the stereo camera calibration with both intrinsic and extrinsic camera parameters, image processing system performed the object feature extraction and pose estimation as mentioned in Sect. 2. Figure 10 shows the object feature extraction and pose estimation results in every step of the two cameras. First, capture the image pair with both cameras at the same time. Next, extract the object from the image based on color using HSV space threshold. Median filtering is applied for noise removing and morphological operation acts like opening and closing to locate the boundary of the object. Next, locate boundary object and the centroid of the located object in both left and right images. Finally, the position of object will be determined from centroid object estimation.

In regard to the stereo camera coordinate, the 3D object position coordinate is obtained using triangulation method as discussed in Sect. 2. The coordinate of 3D object

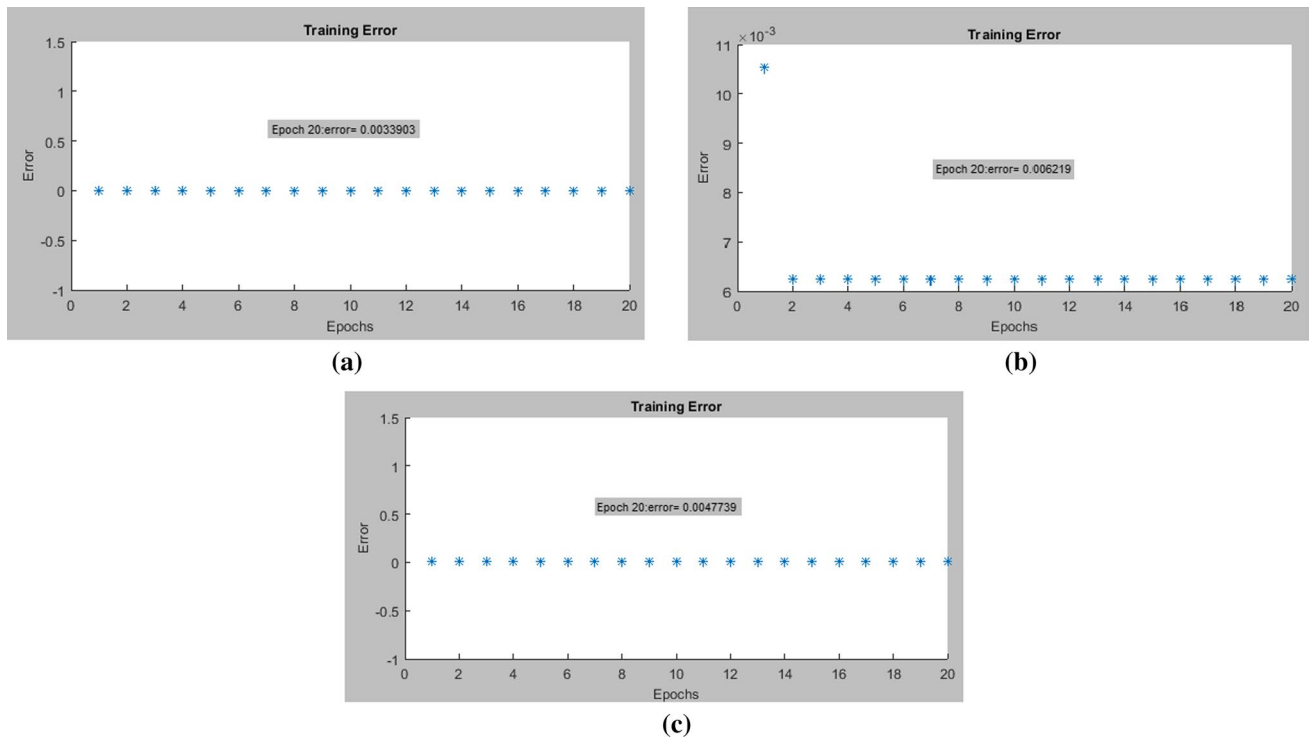


Fig. 13 The smallest training error of ANFIS in different number of membership functions. **a** X-axis training error. **b** Y-axis training error. **c** Z-axis training error

Table 2 ANFIS training error

Number of data	Number of MFs	Training error (mm)			Error average (mm)
		X	Y	Z	
138	3	0.003799	0.017577	0.010973	0.010783
	5	0.014152	0.023203	0.025356	0.020904
	7	0.016763	0.042518	0.032789	0.030690
234	3	0.375590	1.163600	0.161240	0.566810
	5	0.003390	0.006219	0.006021	0.005210
	7	0.010098	0.018891	0.018434	0.015808

position is (−78.03, −50.28, 1022.62) for (x, y, z), respectively, and saved into a file for ANFIS training. According to the results, the color based object detection using HSV color space thresholding succeeds to distinguish the object from the background with robustness from the light changes.

3.3 Experiment 3: ANFIS based eye to hand calibration

Three ANFIS structure with first-order Sugeno fuzzy system is trained to calibrate the position of stereo camera with respect to base frame of robot arm using 3, 5, and 7 Gaussian membership functions (MFs) with twice data collection. The product inference rules are used at the fuzzification layer and hybrid learning algorithm to adjust the

premise and consequent parameters. All the procedure of training followed the steps described in Sect. 2. The 138 and 234 points of 3D object are resulted from twice data collection, for input ANFIS training that captured by calibrated stereo vision system as shows in Fig. 11a, b, respectively. Figure 12 shows the 3D object position toward base frame of robot arm used as the output training data ANFIS is resulted from positioning of end-effector using teaching box control.

We further trained the proposed ANFIS structures to obtain minimal error and/or accepted error using different MFs (3, 5, and 7 MFs) from twice data collection. We found that 5 MFs gave the smallest error. Figure 13 depict the result of training error for (X, Y, Z) axis using 5 MFs. At the end of training process, the ANFIS network would

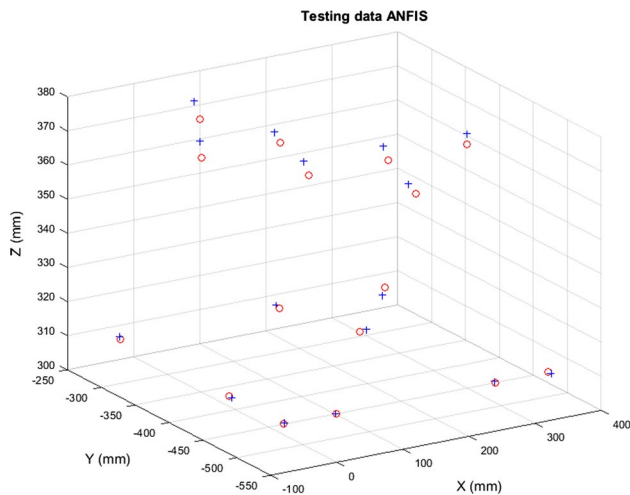


Fig. 14 Testing data for ANFIS

have learned the input–output mapping and it is tested with the testing data.

The details of training results using different MFs with twice data collection are shown in Table 2. We compared the points of 3D object position with 138 and 234 point data collection for input ANFIS training using stereo vision system. Table 2 summarizes the comparison of training error results between 138 and 234 points of double captured data collection using 3, 5, and 7 numbers of membership functions. We found that 5 membership functions with 234 points of data collection generated the smallest training error results (shown in *italics*) compared to 3 and 7 number of membership functions. The 138 training data could not reach the smallest error, even though using different number of MFs.

After training ANFIS and the smallest training error was obtained, then testing to the ANFIS is conducted by positioning the 16 points to test the performance of the hand to eye calibration system as shown in Fig. 14. Based on the experimental results, the error results of ANFIS testing were (0.44, 2.01, 1.53) in mm for x, y and z axis, respectively. It showed a success of designed systems because the object targeted still can be reached using the gripper.

In this experiment, there are three processes to achieve the object 3D pose estimation as follows: (1) image capture using stereo camera with elapsed time is 0.001546 s. (2) Object feature extraction and 3D pose estimation with elapsed time is 0.313688 s. In this process, we process two images which captured from two cameras to detect the object centroid using HSV color thresholding, and then estimate the 3D pose using triangulation process. (3) 3D object position estimation using ANFIS based eye to hand calibration with elapsed time is 0.187859 s. So, the total elapsed time is 0.503093 s, which means the sampling rate of 1 Hz. If we increase the efficiency of the process 2 or 3 a

little, the sampling rate can be increased to be 2 Hz. It will be applicable to the middle-case applications.

4 Conclusions

In this work, the calibrated stereo camera was successfully developed and implemented in a stereo vision-based object manipulation system with eye to hand calibration using the ANFIS method. Based on the experimental results, it is concluded that eye-to-hand calibration with the ANFIS method can achieve a good performance and thus can be implemented for different applications such as object tracking and grasping.

Acknowledgements The authors would like to express their appreciation to Ministry of Science and Technology, Taiwan under contract Nos. MOST 105-2221-E-218-017- and 105-2218-E-218-002- for financial supporting.

References

- Abe S (2012) Neural networks and fuzzy systems: theory and applications. Springer Science & Business Media, New York
- Borangiu T, Dumitrache A (2010) Robot arms with 3D vision capabilities. In: Hall E (ed) Advance in Robot Manipulator. INTECH Open Access Publisher. <http://www.intechopen.com/books/advances-in-robot-manipulators/robot-arms-with-3d-vision-capabilities>. Accessed 21 Nov 2015
- Bouguet J-Y (2015) Matlab camera calibration toolbox http://www.vision.caltech.edu/bouguetj/calib_doc
- Corke P (2011) Robotics, vision and control: fundamental algorithms in MATLAB, vol 73. Springer Science & Business Media, Berlin
- Daniilidis K (1999) Hand-eye calibration using dual quaternions. Int J Robot Res 18:286–298
- Dornaika F, Horaud R (1998) Simultaneous robot-world and hand-eye calibration. IEEE Trans Robot Autom 14:617–622. doi:10.1109/70.704233
- Fevry B, Wyns B, Boullart L, Llata García JR, Torre Ferrero C (2010) Industrial robot manipulator guarding using artificial vision. In: Ude A (ed) Robot Vision. pp 429–454. <http://www.intechopen.com/books/robot-vision/industrial-robotmanipulator-guarding-using-artificial-vision>. Accessed 12 Dec 2015
- Jafari S, Jarvis R Robotic hand-eye coordination: from observation to manipulation. In: Hybrid intelligent systems, 2004. HIS '04. Fourth international conference on Robotics and biomimetics (ROBIO). 5–8 December 2004. pp 20–25. doi:10.1109/ICHIS.2004.82
- Jang J-SR (1993) ANFIS: adaptive-network-based fuzzy inference system Systems. IEEE Trans Man Cybern 23:665–685
- Jiadi Q, Fuhai Z, Yili F, Shuxiang G (2014) Approach movement of redundant manipulator using stereo vision. In: 2014 IEEE international conference on Robotics and biomimetics (ROBIO). 5–10 December 2014. pp 2489–2494. doi:10.1109/ROBIO.2014.7090714
- Juang J-G, Tsai Y-J, Fan Y-W (2015) Visual recognition and its application to robot arm control. Appl Sci 5:851
- Kheng ES, Hassan AHA, Ranjbaran (2010) A stereo vision with 3D coordinates for robot arm application guide. In: Sustainable utilization and development in engineering and technology (STUDENT), 2010 IEEE conference on, 2010. IEEE, pp 102–105

- Kucuk S, Bingul Z (2006) Robot kinematics: forward and inverse kinematics. INTECH Open Access Publisher. http://www.intechopen.com/books/industrial_robotics_theory_modelling_and_control/robot_kinematics_forward_and_inverse_kinematics. Accessed 23 July 2015
- Liu Z, Chen T (2009) Distance measurement system based on binocular stereo vision. In: Artificial intelligence, 2009. JCAI '09. International joint conference on, 25–26 April 2009. pp 456–459. doi:10.1109/JCAI.2009.77
- Miseikis J, Glette K, Elle OJ, Torresen J (2016) Automatic calibration of a robot manipulator and multi 3D Camera System. (arXiv:160.101566)
- Nguyen T, Slaughter D, Max N, Maloof J, Sinha N (2015) Structured light-based 3D reconstruction. Syst Plants Sens 15:18587
- Nof SY (2009) Springer handbook of automation. Springer Science & Business Media, Berlin
- Pérez L, Rodríguez Í, Rodríguez N, Usamentiaga R, García DF (2016) Robot guidance using machine vision techniques in industrial environments: a comparative review (Basel, Switzerland). Sensors 16:335. doi:10.3390/s16030335
- Point_Grey (2015) Stereo vision introduction and applications technical application note (TAN2008005) Revised November 3
- Tao Z (2015) Visual servoing for a non-reactive industrial manipulator. Thesis, Vanderbilt University
- Tsai RY, Lenz RK (1989) A new technique for fully autonomous and efficient 3D robotics hand/eye calibration. IEEE Trans Robot Autom 5:345–358. doi:10.1109/70.34770
- Wilson A (2016) 3D expands the dimensions of vision systems Vis Syst January 2016 <http://www.vision-systems.com/articles/print/volume-21/issue-1/features/3d-expands-the-dimensions-of-vision-systems.html>. Accessed 15 June 2016
- Wu H, Tizzano W, Andersen TT, Andersen NA, Ravn O (2014) Hand-eye calibration and inverse kinematics of robot arm using neural network. In: Proceedings of the 2013 Robot intelligence technology and applications. Springer, pp 581–591. doi:10.1007/978-3-319-05582-4_50
- Zhang Z, Matsushita Y, Ma Y (2011) Camera calibration with lens distortion from low-rank textures. In: Computer vision and pattern recognition (CVPR), 2011 IEEE conference on, 2011. IEEE, pp 2321–2328