Vision-Based 3-D Object Pick-and-Place Tasks of Industrial Manipulator

Guor-Yieh Luo
Department of Electrical Engineering
National Cheng Kung University, Tainan 701, Taiwan

Ming-Yang Cheng*
Department of Electrical Engineering
National Cheng Kung University, Tainan 701, Taiwan

Chia-Ling Chiang
Department of Electrical Engineering
National Cheng Kung University, Tainan 701, Taiwan

***Abstract*—When dealing with complicated tasks such as object pick-and-place, it is harder for robotic arms alone to complete them. One of the possible solutions to overcoming the aforementioned difficulties is to introduce machine vision into the robotic arm system. In this paper, the eye-to-hand camera configuration is adopted in the development of the vision-based automatic pick-and-place systems for 3-D objects. The vision-based automatic pick-and-place system developed in this paper consists of three main sections – calibration of machine vision system, object recognition, and transformations of object coordinates. Experimental results indicate that the vision-based automatic pick-and-place system developed in this paper is able to perform an automatic pick-and-place task for 3-D objects.**

Keywords—Machine vision, Calibration, Objection recognition, Stereo vision

I. INTRODUCTION

In an industrial automation system, robotic arms are the most flexible devices that can perform many different kinds of tasks. Since more and more industries are following the trend of small-volume large-variety production, robotic arms have become more and more popular in industrial automation systems. Nevertheless, without the assistance of other sensors such as vision, there are tasks that a robotic arm cannot handle by itself. For example, in object pick-and-place tasks, one commonly seen scenario is that the target objects are scattered around on a conveyor belt. It is very difficult to plan a desired path for a robotic arm to automatically pick and place those target objects if their 3D poses (positions and orientations) in the Cartesian frame are unknown. In view of this, machine vision systems are introduced in order to solve this problem. The machine vision system is responsible for recognizing the target objects and estimating their poses. Once the pose of an object is estimated, the robotic arm can then know where to pick up the target object. Based on the reasons mentioned above, this paper mainly focuses on machine vision systems of industrial manipulators. For simplicity, this paper only discusses the object pick-and-place application, which is one of the most common applications in industry.

The authors would like to thank the Ministry of Science and Technology, Taiwan, under Grant Nos. MOST 103-2221-E-006-185-MY2 and Catcher technology CO., LTD. for their financial support of this research.

Typically, there are two types of camera configurations used in vision-based robotic applications – eye-in-hand and eye-to-hand [1]. In eye-in-hand configuration, one or more cameras are attached to the tool center point (TCP) of a robotic arm and the camera(s) can be moved freely. In eye-to-hand configuration, one or more cameras are fixed in the workspace. This paper adopts the robotic system with eye-to-hand configuration which implements two cameras (a stereo camera) to perform 3D object recognition so as to execute object pick-and-place tasks.

To obtain the 3D information of an object, additional devices or algorithms are required [2]. The depth of the target object can be estimated by the stereo camera. However, its disadvantage is that the eye-to-hand stereo camera often cannot be moved freely and has a fixed field of view. The stereo camera is usually placed far away from the target objects in order to increase its field of view and to prevent collision with the robotic arm. Unfortunately, this dramatically decreases the accuracy of object recognition [3].

In this paper, vision-based automatic pick-and-place systems can be divided into three main sections – calibration of the machine vision system, object recognition, and the transformations of object coordinates [4,5]. Calibration of the machine vision system includes three parts: camera calibration [6-8], stereo calibration and hand-eye calibration [9,10]. In particular, this paper uses the technique proposed by Zhang [6] to perform camera calibration and stereo calibration. For hand-eye calibration, this paper uses transformation matrices to calculate the transformation relationship from the robot base coordinate system (hand) to the camera coordinate system (eye). The main purpose of object recognition [11] is to recognize objects in an image and also calculate their poses. In general, 3D object recognition [12] consists of four steps – interest point detection, interest point descriptor, match-point error elimination, and recognition of model. 3D object recognition is more suitable for feature-based methods, such as SIFT [13] and SURF [14]. After performing object recognition, the pose of a target object in the camera coordinate system (i.e. camera frame) is obtained. However, in order for the robotic

arm to pick and place the object, the object's pose needs to be represented in the robotic arm frame instead. Thus, the coordinate of the object's pose is transformed from the camera frame to the robotic arm frame by using homogenous transformation matrices [15,16]. Several experiments are performed to assess the performance of the vision-based automatic pick-and-place system developed in this paper.

The remainder of the paper is organized as follows. Section 2 addresses the calibration of the machine vision system. Object recognition and object coordinate transformation are detailed in Section 3. Experimental setups and results are provided in Section 4. Conclusions are given in Section 5.

II. Calibration of Machine Vision System

Before performing an object pick-and-place task, the robotic system with eye-to-hand configuration is required to perform camera calibration, stereo calibration, and hand-eye calibration. These calibrations provide the parameter information required in performing 3D object recognition and transformations of object coordinates. Since the parameter values obtained from calibrations are fixed, these calibrations only need to be performed once. In this paper, all the calibrations are performed by using EmguCV/OpenCV libraries [6,8,17].

A. Camera Calibration

The purpose of camera calibration is to obtain the intrinsic matrix, extrinsic matrix, and distortion coefficients of a camera. The intrinsic matrix describes the relationship between the image plane and the camera frame, while the extrinsic matrix describes the relationship between the camera frame and the chessboard frame. The distortion coefficients are used for correcting image distortion. In addition, the EmguCV/OpenCV calibration method also implements the Levenberg-Marquardt optimization algorithm [18] to minimize the reprojection error (i.e. geometric error corresponding to the image distance between a projected point and its re-projected point).

B. Stereo Calibration

The purpose of stereo calibration is to obtain the geometric relationship between the left camera and the right camera. This geometric relationship is the transformation matrix ${}^{CR}H_{CL} = [{}^{CR}R_{CL}\ {}^{CR}t_{CL}]$, as shown in Fig.1.

There are two purposes for calculating the transformation matrix ${}^{CR}H_{CL}$. The first purpose is to use ${}^{CR}H_{CL}$ to perform image rectification. The image rectification requires ${}^{CR}H_{CL}$ to rectify the left and the right image plane so that they can reach the ideal binocular stereo vision structure. The second purpose is to use ${}^{CR}H_{CL}$ to calculate the baseline b. With the ideal binocular stereo vision structure and the baseline b, the depth estimation based on disparity can be performed to estimate the depth information ${}^{C}Z$ of an object.

For stereo calibration, the chessboard plate is moved around while both the left and the right cameras capture the images of the chessboard plate at the same time. In order to ensure that the quality of stereo calibration is consistent, this paper uses 35 pairs of chessboard images taken from both the left camera and the right camera to perform stereo calibration. Below are the steps for performing stereo calibration employed in this paper:

- Take 35 images of the chessboard plate from the left camera.
- Perform the left camera calibration.
- Take 35 images of the chessboard plate from the right camera.
- Perform the right camera calibration.
- Take 35 pairs of images of the chessboard plate from both the left camera and the right camera at the same time.
- Perform stereo calibration with the chessboard plate model, the detected chessboard corners, and the intrinsic matrices of the right and the left cameras as input parameters. The output of the stereo calibration is the transformation matrix ${}^{CR}H_{CL}$.

C. Hand-Eye Calibration

The purpose of hand-eye calibration is to establish the transformation relationship between the cameras and the robotic arm. Once the transformation relationship has been established, an object's position in the camera frame can be transformed into the object's position in the robot base frame. Then, the robotic arm can know where to pick and place the object. The chain of transformations for a robotic system with eye-to-hand configuration is shown in Fig. 2. The goal of an eye-to-hand calibration [10] is to obtain the transformation matrix between the robot base and the camera (${}^{base}H_{cam}$).

To obtain ${}^{base}H_{cam}$, an equation is derived from the chain in Fig. 2:

$$ {}^{cam}H_{cal} = {}^{cam}H_{base} \cdot ({}^{base}H_{tcp} \cdot {}^{tcp}H_{cal}) = {}^{cam}H_{base} \cdot {}^{base}H_{cal} \quad (1) $$

In (1), ${}^{cam}H_{base}$ is the transformation matrix that needs to be calculated in the eye-to-hand calibration. ${}^{cam}H_{cal}$ is the transformation matrix from the calibration rig frame to the camera frame, which can be calculated by using either camera calibration (extrinsic matrix) or 3D recognition of the calibration rig. ${}^{base}H_{cal}$ is the transformation matrix from the calibration rig frame to the robot base frame, which can be calculated by using forward kinematics.

To solve for ${}^{cam}H_{base}$, (1) is rearranged to (2):

$$ \begin{aligned} {}^{cam}H_{base} &= {}^{cam}H_{cal} \cdot ({}^{base}H_{cal})^{-1} \\ {}^{cam}H_{base} &= {}^{cam}H_{cal} \cdot {}^{cal}H_{base} \end{aligned} \quad (2) $$

Lastly, ${}^{base}H_{cam}$ is obtained by taking an inverse of ${}^{cam}H_{base}$ in (2):

$$ {}^{base}H_{cam} = ({}^{cam}H_{base})^{-1} \quad (3) $$

III. Object Recognition and Transformations of Object Coordinates

A. Object Recognition Results

The object recognition used in the eye-to-hand configuration is classified as 3D object recognition. This paper uses a stereo camera to estimate the depth information ${}^{C}Z$ of objects based on their disparities. However, in order to calculate the depth information, additional steps are required in

the procedure of object recognition. The training process for the target objects used in the robotic system with eye-to-hand configuration is divided into two parts. In the first part, the object images captured from the stereo camera are processed by SIFT and SURF. The SIFT and SURF algorithms not only detect the interest points of the target object, but also calculate the gradients of these interest points. If an interest point in the left image and another interest point in the right image have a similar gradient value, they are matched. However, there might be a few falsely matched points. A mismatch occurs when there are more than two interest points having the same gradient value. Thus, this paper uses the disparity vector and the depth information to eliminate the falsely matched points. The final match points are used for estimating the depth of the object. The average depth of the target object is saved and used as a feature to identify and classify the object.

In the second part of the training process, an image that contains the target objects (a square, a cuboid, and a cylinder) are processed by Canny edge detection [19]. Canny edge detection extracts the edge of each target object and the edge is used for calculating the contour vector [20]. Since each object class has unique contour vector data, the contour vector can also be used as a feature to identify and classify an object. Therefore, each target object has two types of unique features – depth information and contour vector data.

In the classification process, the target objects are randomly placed on the experimental platform and are captured by the stereo camera. Since the SIFT and SURF algorithms take a longer time to compute the interest points and the interest point descriptors, these two algorithms are not used in the classification process. Instead, only the Canny edge detection algorithm is used to find the edge of each object; this reduces the computation time by about 1sec. The edge of each target object is then used for calculating the object's contour vector and gripping point. The object's gripping point is further used for estimating the object's depth. The contour vector information is used for classifying the object class (cube class, cuboid class, and cylinder class), and the depth information is used for determining the size of the object (e.g., cubes with sides equal to 2cm, 3cm, or 4.5cm). Once the target objects are identified and their poses are estimated, the robotic arm can perform the pick-and-place task for those target objects.

B. Transformations of Object Coordinates

The main goal of the transformations of object coordinates is to transform the coordinates of target objects from the image plane coordinates to the robot base coordinates, so that the robotic arm knows where to pick and place the target objects. The chain of coordinate transformations for a robotic system with eye-to-hand configuration is constructed by combining the pinhole camera model and the chain of transformations for the robotic system with eye-to-hand configuration (Fig. 2). Totally, there are three different frames – image plane, stereo camera frame, and robot base frame. Thus, there are two coordinate transformations in total. The first operation sequence is the coordinate transformation from the image plane to the stereo camera frame. The equation for the first operation sequence is expressed as follows:

$$ {}^{cam}P_{obj} = {}^{C}Z \cdot K^{-1} \cdot p_{obj} \qquad (4) $$

where: ${}^{cam}P_{obj}$: The object point $[{}^{C}X, {}^{C}Y, {}^{C}Z]^T$ in the camera frame; p_{obj}:The object point $[u, v, 1]^T$ on the image plane.

The intrinsic matrix K in (4) is obtained from the camera calibration of the eye-to-hand stereo camera; and the depth ${}^{C}Z$ is estimated by using depth estimation based on disparity.

The second operation sequence is the coordinate transformation from the stereo camera frame to the robot base frame. This is done by using the transformation matrix ${}^{base}H_{cam}$, which is calculated from the eye-to-hand calibration. Equation (5) is derived by combining the first and the second operation sequence:

$$ {}^{base}P_{obj} = {}^{base}H_{cam} \cdot {}^{cam}P_{obj} \qquad (5) $$

where: ${}^{base}P_{obj}$:The object point in the robot base frame; ${}^{base}H_{cam}$: The transformation matrix from the stereo camera frame to the robot base frame

Lastly, the overall transformation equation for the entire system (i.e. (6)) is formed by combining (4) and (5). Equation (6) represents the chain of coordinate transformations, from the image plane to the robot base frame, for a robotic system with eye-to-hand configuration.

$$ \begin{aligned} {}^{base}P_{obj} &= {}^{base}H_{cam} \cdot {}^{cam}P_{obj} \\ {}^{base}P_{obj} &= {}^{base}H_{cam} \cdot ({}^{C}Z \cdot K^{-1} \cdot p_{obj}) \end{aligned} \qquad (6) $$

IV. Experimental Setups and Results

A. Experimental Setups

In the experiment, to avoid collisions between the robotic arm and the stereo camera, the stereo camera is placed high above the experimental platform (around 817.5mm) as shown in Fig.3. However, the drawback is that the accuracy of object recognition is greatly reduced. In addition, the gripper is attached on the TCP for picking/placing target objects such as the ones shown in Fig. 4.

The schematic diagram of the robotic system with eye-to-hand configuration for object pick-and-place is shown in Fig. 5.

B. Camera Calibration Results

For the camera calibrations of the stereo camera (the left camera and the right camera) in the eye-to-hand configuration, this paper uses a 6×9 square chessboard with each square's side equal to 2cm. The size of the chessboard depends on the FOV of the camera. In each camera calibration, 35 chessboard images are taken.

Below are the results of camera calibrations for the two cameras, where K is the intrinsic matrix, (k_1, k_2, k_3, k_4, k_5, k_6, p_1, p_2) are the distortion coefficients, and $err_{reproj.}$ is the reprojection error:

- The left camera:

$$ K_L = \begin{bmatrix} 4175.81 & 0 & 639.53 \\ 0 & 4178.11 & 479.48 \\ 0 & 0 & 1 \end{bmatrix} \qquad (7) $$

$$k_{L1} = 0.05206 \quad k_{L5} = 337.99680$$
$$k_{L2} = -9.40974 \quad k_{L6} = -0.05206$$
$$k_{L3} = 0.00046 \quad p_{L1} = 9.40974$$
$$k_{L4} = -0.00077 \quad p_{L2} = -337.99680$$
$$err_{L\,reproj.} = 0.58273$$

- The right camera:

$$K_R = \begin{bmatrix} 4170.54 & 0 & 639.47 \\ 0 & 4171.98 & 479.48 \\ 0 & 0 & 1 \end{bmatrix} \quad (8)$$

$$k_{R1} = -0.03036 \quad k_{R5} = -39.67438$$
$$k_{R2} = 0.80114 \quad k_{R6} = 0.03036$$
$$k_{R3} = 0.00179 \quad p_{R1} = -0.80114$$
$$k_{R4} = -0.00022 \quad p_{R2} = 39.67438$$
$$err_{R\,reproj.} = 0.50470$$

Because the two cameras and lenses used in this paper are identical, their calibration results are very similar to each other. The average focal length estimated from the camera calibrations is 4174.11 pixels. Since the width and height of a pixel in the image sensor are 3.75μm, the average estimated focal length is 15.65mm, which is close to the focal length of the lens 15mm. Furthermore, the average image center coordinate estimated from camera calibrations is (639.50, 479.48), which is also close to the image center coordinate of the camera from datasheet (640, 480). Therefore, the results of camera calibrations for the two cameras are acceptable. In addition, the reprojection errors ($err_{reproj.}$) of the two cameras are within the range between 0.0 and 1.0. This indicates that the results of camera calibrations are accurate.

C. *Stereo Calibration Results*

This paper uses a 6×9 square chessboard with each square's side equal to 2cm to perform the stereo calibration. As with camera calibration, 35 pairs of chessboard images are captured by the left and the right cameras. However, the difference is that both the left camera and the right camera need to take the chessboard images at the same time. After the stereo calibration, the transformation matrix ${}^{CR}H_{CL}$ is obtained, as shown in (9). The baseline b is the norm of the translation vector ${}^{CR}t_{CL}$ in ${}^{CR}H_{CL}$, as described by (10). The estimated baseline b from the stereo calibration is 50.0435mm, which is almost the same as the measured distance (5cm) between the left camera and the right camera. This indicates that the stereo calibration result is accurate.

$${}^{CR}H_{CL} = \begin{bmatrix} 1.0000 & 0.0076 & 0.0048 & -49.9936 \\ -0.0075 & 1.0000 & -0.0057 & -0.9582 \\ -0.0049 & 0.0057 & 1.0000 & -2.0195 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (9)$$

$$b = \left\| {}^{CR}t_{CL} \right\| = 50.0435 \text{ mm} \quad (10)$$

After the image rectification, the left and the right image planes will be shifted and rotated. Because the right and the left image planes have been modified, the right and the left camera intrinsic matrices will also be modified. The new left intrinsic matrix K_L' and the new right intrinsic matrix K_R' are described by (11).

$$K_L' = \begin{bmatrix} 4171.79 & 0 & 457.54 \\ 0 & 4171.79 & 481.61 \\ 0 & 0 & 1 \end{bmatrix} \quad (11)$$

$$K_R' = \begin{bmatrix} 4171.79 & 0 & 457.54 \\ 0 & 4171.79 & 481.61 \\ 0 & 0 & 1 \end{bmatrix}$$

In order to see more clearly the difference between the original images and the rectified images, several ~~red~~ parallel lines are drawn on these images, as illustrated in Fig. 6 and Fig. 7. Originally, before image rectification (Fig. 6), the center point of the wooden cube in the left image plane and in the right image plane do not lie on the same horizontal line. However, after image rectification (Fig. 7), they lie on the same horizontal line. Thus, in order to obtain a more accurate depth information ${}^{C}Z$ of an object, it is necessary to use the rectified images and the new intrinsic matrices instead of the original images and the old intrinsic matrices.

D. *Hand-Eye Calibration Results*

For eye-to-hand configuration, the transformation matrix from the camera frame to the robot base frame ${}^{base}H_{cam}$ is computed. Based on (2), the eye-to-hand calibration requires the transformation matrices ${}^{base}H_{cal}$ and ${}^{cam}H_{cal}$ to compute ${}^{base}H_{cam}$. The transformation matrix ${}^{base}H_{cal}$ can be obtained easily by using the forward kinematics of the six-axis robotic arm. For the transformation matrix ${}^{cam}H_{cal}$, this paper uses camera calibration to compute it. This is because the extrinsic matrix obtained from camera calibration is in fact equivalent to the transformation matrix ${}^{cam}H_{cal}$. The results of eye-to-hand calibration are shown in (12).

$${}^{base}H_{cal} = \begin{bmatrix} 0.9999 & 0.0676 & 0.0000 & 648.9 \\ 0.0671 & -0.9977 & 0.0000 & 1016.5 \\ 0.0000 & 0.0000 & -1.0000 & 580.5 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$${}^{cam}H_{cal} = \begin{bmatrix} 0.9999 & 0.0001 & -0.0136 & -70.3 \\ 0.0002 & 0.9997 & 0.0234 & -34.6 \\ 0.0136 & -0.0234 & 0.9996 & 817.6 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$${}^{base}H_{cam} = \begin{bmatrix} 0.9977 & 0.0676 & 0.0120 & 711.4 \\ 0.0670 & -0.9974 & 0.0243 & 966.7 \\ 0.0136 & -0.0243 & -0.9996 & 1398.0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (12)$$

E. *Object Pick-and-Place Results*

After all the calibrations and the training process are

complete, the robotic system with eye-to-hand configuration is ready to perform the automatic object pick-and-place task. Fig.8 is the sequence of the object pick-and-place process for the robotic system with eye-to-hand configuration. The order of the sequence is from left to right and top to bottom.

V. CONCLUSIONS

The main purpose of this paper is to develop machine vision systems for object pick-and-place tasks of industrial manipulators. The machine vision system developed in this paper consists of three main sections – calibration of machine vision system, object recognition, and transformations of object coordinates. In this paper, the EmguCV/OpenCV library is used to perform camera calibration, stereo calibration, and hand-eye calibration. In addition, 3D object recognition is performed and the machine vision system uses the transformation matrices to define the transformation relationship between the camera and the robotic arm. Experimental results indicate that a 6DOF industrial manipulator equipped with the machine vision system developed in this paper can successfully perform the 3D object pick-and-place task.

REFERENCES

[1] P. J. Sanz, A. P. del Pobil, J. M. Iñesta, and G. Recatalá, "Vision-guided grasping of unknown objects for service robots," in *Proc. of the IEEE Int. Conf. on Robotics & Automation*, 1998, vol. 4, pp. 3018-3025.

[2] A. Schrott, "Feature-based camera-guided grasping by an eye-in-hand robot," in *Proc. of the IEEE Int. Conf. on Robotics & Automation*, 1992, vol. 2, pp. 1832-1837.

[3] G. D. Haǵer, W. C. Change, and A. S. Morse, "Robot hand-eye coordination based on stereo vision," *IEEE Control Systems*, vol. 15, pp. 30-39, 1995.

[4] H. J. Tsai, "Application of homography matrix based 3D reconstruction algorithm on six-axis articulated robot," M.S. thesis, Dept. Elect. Eng., National Cheng Kung Univ., Tainan, Taiwan, June 26, 2014.

[5] C. C. Lin, "Study on vision based object grasping of industrial manipulator," M.S. thesis, Dept. Elect. Eng., National Cheng Kung Univ., Tainan, Taiwan, July 13, 2015.

[6] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, pp. 1330-1334, 2000.

[7] R. Y. Tsai. "A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses," *IEEE Journal of Robotics and Automation*, vol. 3, pp. 323-344, 1987.

[8] Itseez. (2016). *OpenCV | OpenCV* [Online]. Available: http://opencv.org/

[9] *HALCON Application Note: Machine Vision in World Coordinates*, MVTec Software GmbH, München, Germany, 2003.

[10] T. C. Chiang, "Study on Hand-Eye Calibration of Six-Axis Articulated Robot," M.S. thesis, Dept. Elect. Eng., National Cheng Kung Univ. Taiwan, Aug. 2014.

[11] Wikipedia. (2016, Apr. 15). *Outline of Object Recognition* [Online]. Available: https://en.wikipedia.org/wiki/Outline_of_object_recognition

[12] K. Alhamzi, M. Elmogy, and S. Barakat, "3D object recognition based on image features: a survey," *International Journal of Computer and Information Technology*, vol. 3, pp. 651-660, May 2014.

[13] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision,* vol. 60, pp. 91-110, 2004.

[14] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool, "SURF: Speeded up robust features," *Computer Vision and Image Understanding*, vol. 110, pp. 346-359, 2008.

[15] K. S. Fu, R. C. Gonzalez, and C. S. G. Lee, *Robotics: Control Sensing Vision and Intelligence*, Vis: Tata McGraw-Hill Education, 1987.

[16] M. W. Spong, S. Hutchinson, and M. Vidyasagar, *Robot Modeling and Control*, Wiley, 2006.

[17] Wikipedia. (2016, May 1). *EmguCV: OpenCV in .NET (C#, VB, C++ and more)* [Online]. Available: http://www.emgu.com/wiki/index.php/Main_Page

[18] J. More, "The Levenberg-Marquardt Algorithm, Implementation, and Theory," *Numerical Analysis*, G.A. Watson, ed., Springer-Verlag, 1977.

[19] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 8, pp. 679-698, Nov. 1986.

[20] P. Torgashov. (2014, Jun. 8). *Contour Analysis for Image Recognition in C#* [Online]. Available: http://www.codeproject.com/Articles/196168/Contour-Analysis-for-Image-Recognition-in-C

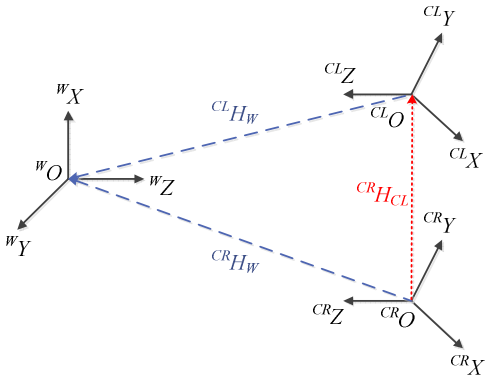

Figure 1. Geometric relationship between the left camera and the right camera; Where ^{W}O: The origin of the world (chessboard) frame; ^{CL}O: The origin of the left camera frame; ^{CR}O:The origin of the right camera frame; $^{CL}H_W$: The left camera extrinsic matrix; $^{CR}H_W$: The right camera extrinsic matrix; $^{CR}H_{CL}$: The geometric relationship between the left and the right camera

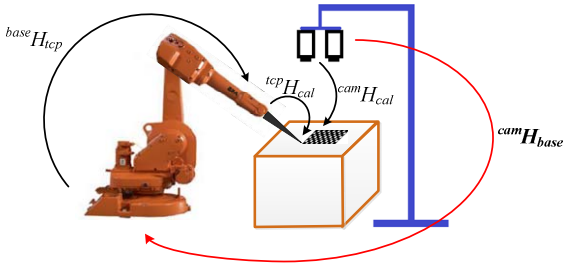

Figure 2. Chain of transformations for a robotic system with eye-to-hand configuration

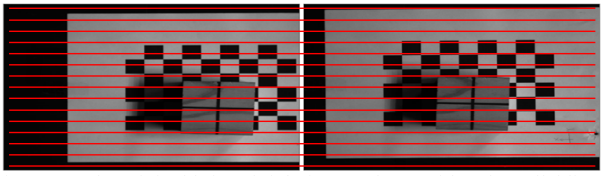

Figure 6. The original left and right image plane (with ~~red~~ parallel lines)

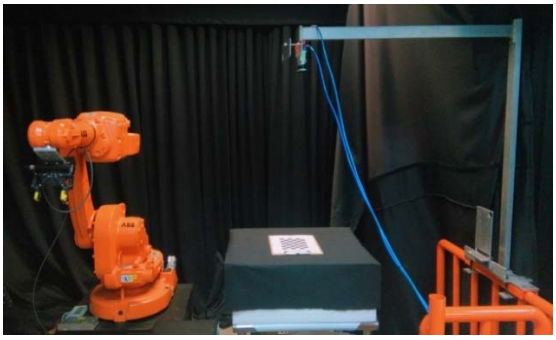

Figure 3. Experimental platform

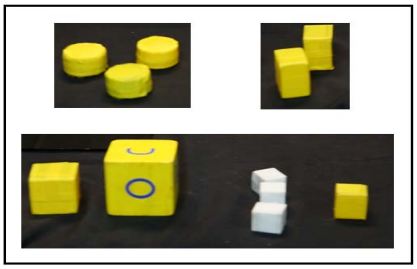

Figure 4. Target objects for picking and placing– cylinders, cuboids, and cubes

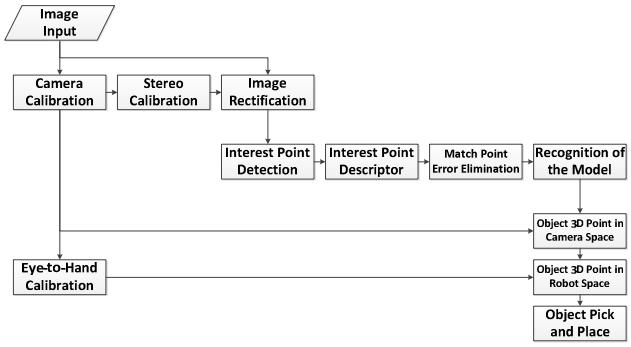

Figure 5. Schematic diagram of the robotic system with eye-to-hand configuration for object pick-and-place

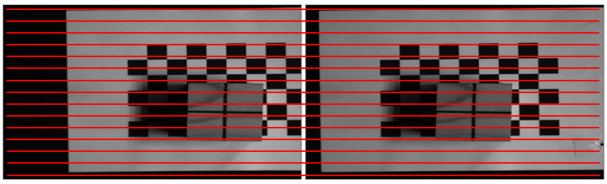

Figure 7. The rectified left and right image plane (with ~~red~~ parallel lines)

Figure 8. Object pick-and-place sequence for eye-to-hand