Object Identification and 3-D Position Calculation Using Eye-in-Hand Single Camera for Robot Gripper

J. Shaw
Institute of Mechatronic Engineering
National Taipei University of Technology
Taipei, Taiwan
jshaw@ntut.edu.tw

K.Y. Cheng
Institute of Manufacturing Engineering
National Taipei University of Technology
Taipei, Taiwan
chengkai1009@gmai.com

***Abstract*—In this paper we designed an intelligent gripper module with vision system for robot manipulators. The vision system in the gripper module used only a single camera, as in the form of eye-in-hand architecture, for identifying a target object in space for the robot manipulator to grasp. SURF algorithm and back projection method of hue-saturation histogram model are applied for correctly identifying the target object. Then the identified object position in space can be located by employing the binocular vision system model using images captured at two different positions. Experimental results show that the vision system can correctly identify the target among different objects and calculate its position with relative error less than 4%. Finally, the intelligent gripper module is attached to a robot manipulator and experiments are carried out to verify its effectiveness in locating objects in space for robot arm to grasp successfully.**

Keywords—robot gripper; SURF; back projection

I. INTRODUCTION

The robot arms have been used for many years, mainly in manufacturing factories. Its reliability in both accuracy and durability plays an important role for its wide-range application in industry. In addition, service robots have been the popular research focus recently which require also robot arms for doing some related service works. Moreover, due to labor cost increase and high quality product demand, robot applications would be one of the best solutions facing many companies. There are many kinds of robot arms with diverse functions depending on the job's need. Robot arms with 4 DOF (such as SCARA robot) and 6 DOF (articulated robot) are very common in the recent market.

The end-effector of a robot arm decides what functions the arms are equipped with. Among which, finger grippers are mostly used resembling dexterous human hands. Two, three, and five fingers are often used in robot gripper design. Shaw and Lee [1] designed a two-finger gripper module, actuated by shape memory alloy with force sensor, to be attached to a robot manipulator as its end-effector. Vision systems, on the other side, are often used with robot arms for purposes such as inspection and guidance (visual servoing) [2]. As compared to a static-and-fixed camera, camera mounted on a robot arm (eye-in-hand architecture) provides more versalities in inspection and assembly. Papanikolopoulos, Khosla, and Kanade [3] used a camera mounted on a robot arm for visual tracking of a moving target, with known depth information. Huang et al. [4] presented a visual servoing approach for macro peg-and-hole alignment with a single eye-in-hand high-speed camera. Tsai [5] studied on automatic material handling of a mobile manipulator with an eye-in-hand vision system. Experiments showed the validity of the vision system. For the eye-in-hand camera to see target objects, different digital image processing algorithms needed be implemented. Artificial markers are commonly attached to target objects for easy computation [4,6]. Without attaching artificial markers to objects, SIFT and SVM algorithms [7] were used to recognize target objects, whereas area and color features and artificial neural network were chosen in [8]. In this paper, we attached a low cost single webcam on top of the gripper module [1] for identifying the target object in space and guiding the robot arm to successfully grasp it, hence completing development of an intelligent robot gripper module with force and vision sensors for any available robot arm. In addition SURF and color features in hue-saturation histogram model are employed for more robust object recognition.

The paper is organized as follows. In Section II, a brief description of system design is given, showing the overall system configuration including the gripper module, robot arm, vision system, and controller. The mathematical methods used for vision system to guide the robot gripper to the target object are introduced in Section III. Experimental results are shown in Section IV, with conclusions in Section V.

II. SYSTEM DESIGN

A. Gripper Module Design

A two-finger gripper module, actuated by shape memory alloy actuator and with force sensor attached to its finger, was designed by the authors [1]. This gripper module was used as an end-effector for a 6-DOF Staubli robot arm and experimental results did show its capability in taking unknown objects with suitable grasping force. This paper aims to develop a vision system for this gripper module so that it has two key features, namely force control and vision system. More specifically, the designed gripper module can be attached to any robot manipulator for visually guiding the robot arm to the target object and picking it up to destination with suitable grasping force without crashing it or losing it.

978-1-4673-8075-1/16/$31.00 ©2016 IEEE

B. System Configuration

A webcam for taking images in front of the gripper is to be fixed at the top cover of the gripper module. The main controller (a laptop computer) processes the images from the webcam to locate the object and send its corresponding control signals to robot controller which in turn moves the robot manipulator to the target object. Finally, the microcontroller in gripper module controls the gripper force to successfully pick up the object and then move it to the place location. The overall system configuration is illustrated in Fig. 1, where USB, Ethernet, and RS232 are employed for respective communications. The corresponding control block diagram is shown in Fig. 2, where three controllers: main controller, robot controller, and gripper microcontroller are used in the system for implementing different tasks.

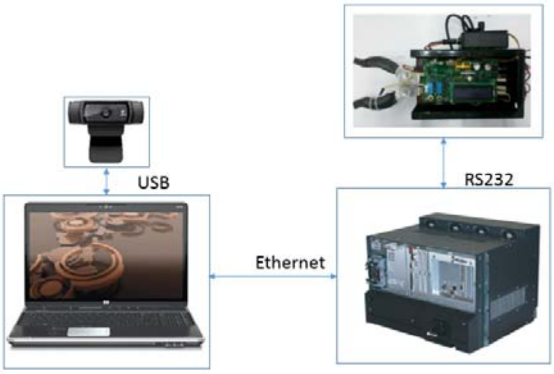


Fig. 1. The system configuration.

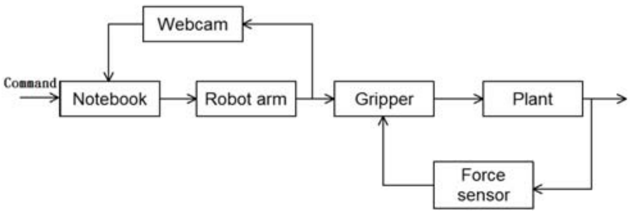


Fig. 2. Control block diagram.

III. THEORETICAL BACKGROUNDS FOR IMAGE PROCESSING

The main controller is used mainly for image processing for identifying the target object based on the command and to compute its position in space so that the robot gripper can be moved to it. To ensure robust and correct object identification, two algorithms effective for object matching are employed: SURF algorithm first for matching object's features, and then back projection method for ensuring the object colors. Finally, the identified object space position can be located by using the binocular vision system model using images captured at two different positions.

A. SURF Algorithm

SURF (Speed-Up Robust Feature) [9] is a upgrade version of SIFT (Scale Invariant Feature Transform) algorithm [10,11] for robust and speed-up considerations. SURF is often applied for real-time object recognition between two images, whose performance is invariant against different image transformations such as translation, rotation, scale and illumination. There are three major steps in SURF algorithm:

1) interest point detection: The object's points of interest can be detected in this step using Fast-Hessian Detector

$$H(p,\sigma)=\begin{bmatrix} L_{xx}(p,\sigma) & L_{xy}(p,\sigma) \\ L_{xy}(p,\sigma) & L_{yy}(p,\sigma) \end{bmatrix} \tag{1}$$

where $L_{()()}(p,\sigma)$ is the second-order derivative of the grayscale image in direction ()() at point p(x, y).

2) local neighborhood descriptor: Every point of interest in this step will be associated with a local descriptor with dimension 64, whose elements are computed based on the sum of Haar wavelet responses. Fig. 3 shows a region centered at an interest point in the pricipal orientation and each sub-region (total 16) will give 4 sums of Haar wavelet responses at 5x5 regularly spaced sample points inside.

3) matching: In this final step, by comparing the descriptors obtained from two images (one for object model, the other for the test), matching pairs can be found. Here, we adopted the *k*-NN (*k*-Nearest Neighbors) algorithm [12] for matching. Essentially, an object is classified by a majority vote of its neighbors, with the object being assigned to the class most common among its *k* nearest neighbors.

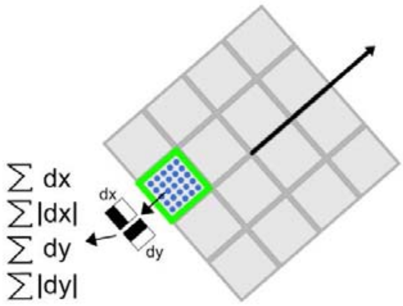


Fig. 3. Formation of a local descriptor.

B. Back Projection Method

The back projection method [13] is to be applied immediately after SURF algorithm to distinguish similar objects (if any) with different colors. A hue-saturation histogram model, having 16 bins in hue (range from 0 to 360) and saturation (range from 0 to 255) space, must be constructed first for a specific object. The histogram model represents probability distribution of the object in hue and saturation space. This hue-saturation histogram model will then be used to identify the possible region of the object in the test image using the back projection method, detailed steps are as follows:

1) For each pixel (*x,y*) in the test image, read its hue and saturation values $(h_{x,y}, s_{x,y})$ and locate the bin coordinates in the histogram model corresponding to the hue and saturation values.

2) Read the value (probability) stored in the bin coordinates of the histogram model.
3) Put this probability value in pixel (*x,y*) of a new image (back projection image). Repeat the process until all pixels are filled. Normalize the back projection image to [0, 255], the region of the object in the image can thus be identified.

C. 3-D Position Calculation

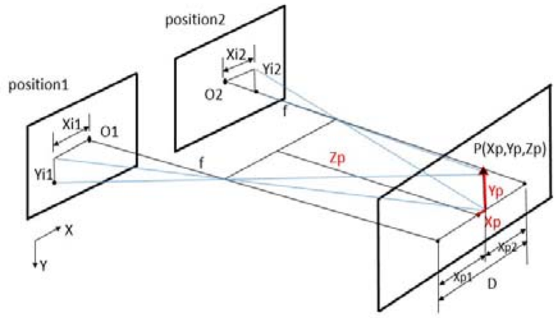
After consecutive applications of SURF and back projection algorithm, the target object in front of the camera can be surely identified. We further calculate its center in terms of image frame coordinates (x_{i1}, y_{i1}), as shown in Fig. 4 at camera position 1. Since space position of the target object is needed for gripper to move to and only a single camera is attached to the gripper module, it is obvious that a second camera position (position 2) is required to take another image to have coordinates (x_{i2}, y_{i2}), as stated by the binocular vision system model. Therefore, robot arm will move camera a distance of D to position CCD2 along *x*-axis. It is noted that the positions at 1 and 2, as well as the distance D, are all known from the robot arm configuration. Finally, the space coordinats of the center of the target object with respective to point O_{CCD} can be computed using triangular similarity theorem as follows:

$$z_p = \frac{D}{\frac{|x_{i1}|}{f_1} + \frac{|x_{i2}|}{f_2}} \tag{2}$$

$$x_p = \frac{1}{2}\left(\frac{x_{i1}}{f_1} - \frac{x_{i2}}{f_2}\right) z_p \tag{3}$$

$$y_p = \frac{z_p}{f_1} y_{i1} = \frac{z_p}{f_2} y_{i2} \tag{4}$$

where $f_1 = f_2$ is the focus length. To verify the accuracies of the method, a target object has been placed at ten different space locations in front of the camera and the relative errors between the actual coordinates and the computed coordinates by (2,3,4) are calculated. The relative errors in each X,Y,Z direction are all within 4%, thus enabing the camera on the gripper module to guide the robot arm to the target object.



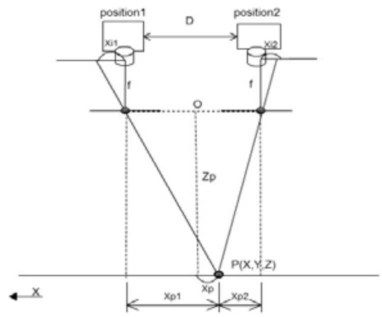


Fig. 4. Binocular vision system model.

IV. EXPERIMENTAL RESULTS

In the experiments, the designed gripper module with a webcam attached to its top cover is installed in the sixth axis of a 6-DOF Staubli robot arm, as shown in Fig. 5. Sample objects to be picked up are placed randomly in the conveyor besides the robot arm. The robot arm at its initial starting position will be commanded to pick up the target object in the conveyor and move it to the place position after the main controller successfully identifies the object and calculates its corresponding coordinates in space.

Fig. 5. Robot arm with gripper attached.

The first stage in the experiment is to use SURF algorithm to identify the target object in the conveyor. Fig. 6 illustrates the target has been successfully matched to its SURF model with respective matching interest points shown. However, the SURF algorithm works only with the grayscale image, and hence might lose key color features. Therefore a second matching algorithm using back projection of the target's hue-saturation histogram model is applied for double safety. A yellow and a white clips at different position and depth in the conveyor among other objects are shown in the upper-left corner of Fig. 7. The target object is the yellow clip. Only after the application of back projection algorithm of the yellow clip's hue-saturation histogram model can the correct target be identified (instead of the white clip). Its 3D space position is thus calculated and shown in the screen of human-machine interface in Fig. 7. The robot arm is then commanded to do the pick-and-place task for the yellow clip (with force control implemented), as shown in Fig. 8.

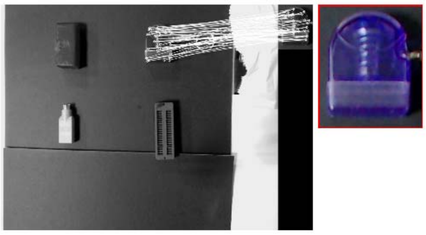


Fig. 6. SURF algorithm to match the target object.

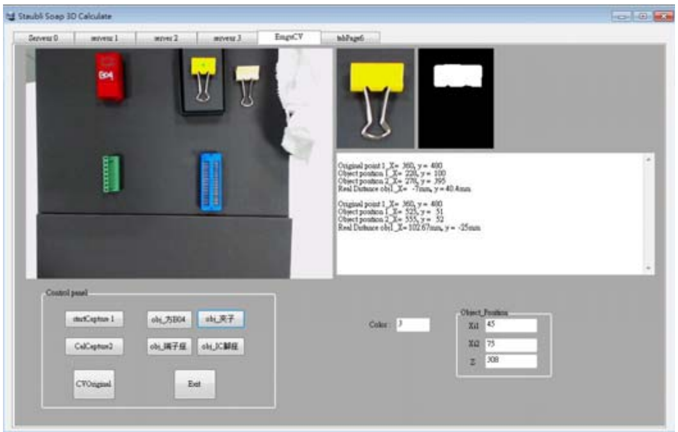
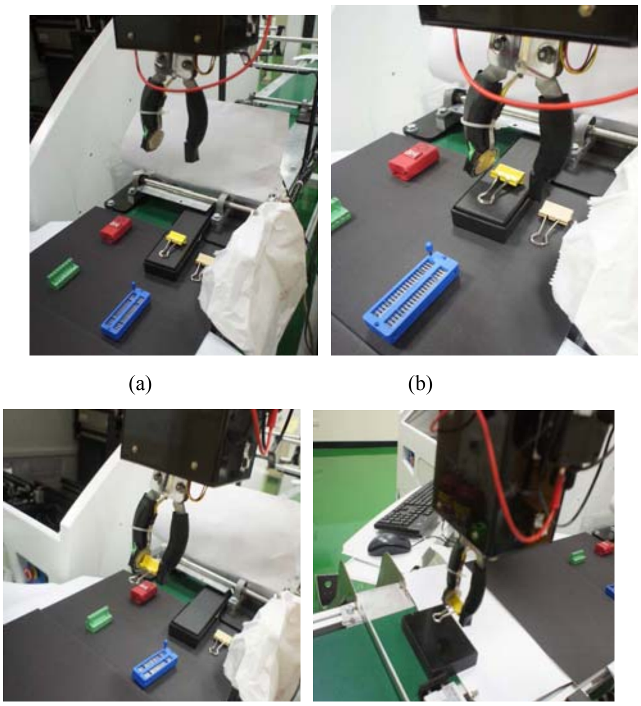


Fig. 7. Screen of human-machine interface for experiment.



(a) (b) (c) (d)

Fig. 8. Snapshop Photos of robot arm doing pick-and-place task.

V. Conclusions

In this paper, an intelligent robot gripper module with force control and a eye-in-hand vision system is developed. This gripper can be attached easily to any commonly seen robot manipulator for industry and service works. The single camera in the gripper module can detect a target object and compute its space position for the robot arm, employing SURF, back projection, and binocular vision system model. Experimental results show effectiveness of the proposed methods for pick-and-place tasks. In the future, orientation of the object needs be calculated also for more robust and versatile object grasping of the gripper.

Acknowledgment

The work was partially supported by Minister of Science and Technology, Taiwan, under grant number MOST 104-2221- E- 027 - 030.

References

[1] J. Shaw and J.L. Lee, "Force control of a robot gripper featuring shape memory alloy actuators," International Conference on Advanced Robotics and Intelligent Systems, Taipei, Taiwan, June 6~8, 2014.

[2] S. Hutchinson, G.D. Hager, and P.I. Corke, "A tutorial on visual servo control," IEEE Transactions on Robotics and Automation, vol. 12, no.5, pp. 651-670, October 1996.

[3] N.P. Papanikolopoulos, P.K. Khosla, and T. Kanade. "Visual tracking of a moving target by a camera mounted on a robot: a combination of control and vision," IEEE Transactions on Robotics and Automation, vol. 9, no. 1, pp. 14-35, 1993.

[4] Shouren Huang, Y. Yamakawa, T. Senoo, M. Ishikawa, "Realizing peg-and-hole alignment with one eye-in-hand high-speed camera," AIM 2013, pp. 1127-1132, 9-12 July 2013.

[5] C.Y. Tsai, "A study on automatic material handling of a mobile manipulator with an eye-in-hand vision system," MOST research report, 2014.

[6] Tingting Wang, Wenfang Xie, Guodong Liu, Yimin Zhao, "Quasi-min-max model predictive control for image-based visual servoing," AIM 2012, pp. 98-103, 11-14 July 2012.

[7] Chyi-Yeu Lin and Edwin Setiawan, "Object orientation recognition based on SIFT and SVM by using stereo camera," IEEE International Conference on Robotics and Biomimetics, Bangkok, Thailand, February 21 - 26, 2009.

[8] H.I. Lin, Yi-Yu Chen, and Yung-Yao Chen, "Robot vision to recognize both object and rotation for robot pick-and-place operation," International Conference on Advanced Robotics and Intelligent Systems, Taipei, Taiwan, June 6~8, 2014.

[9] B. Herbert, E. Andreas, T. Tuytelaars, L.V. Gool, "SURF: speeded up robust features, " Computer Vision and Image Understanding, vol. 110, no. 3, pp. 346-359, 2008.

[10] D.G. Lowe, "Distinctive image features from scale-invariant keypoints," International Journal of Computer Vision, vol. 60, no. 2, pp. 91-110, 2004.

[11] T.S. Caetano, J.J. McAuley, L. Cheng, Q.V. Le, and A.J. Smola, "Learning graph matching", COCS, Beijing University Of Technology, April 2011.

[12] N.S. Altman, An introduction to kernel and nearest-neighbor nonparametric regression, The American Statistician, 1992.

[13] R.C. Gonzalez, R.E. Woods, Digital Image Processing, Addison Wesley, 2009.