Object Pose Estimation for Grasping Based on Robust Center Point Detection

Kai-Tai Song and Che-Hao Chang
Department of Electrical Engineering
National Chiao Tung University, Hsinchu, Taiwan, R.O.C.
ktsong@mail.nctu.edu.tw, changchehao.ece97g@nctu.edu.tw

Abstract—**The objective of this study is to design a grasping system for a mobile manipulator, such that it can find and grasp a target object using vision. Speed up robust feature (SURF) algorithm was adopted to define features of the target object and match features between current image and object database to confirm the target. To strengthen the feature matching results and calculate the necessary reference control point, we adopted RANdom Sample Consensus(RANSAC) algorithm to estimate the planar transformation matrix (Homography matrix) in order to accurately mark the center of target. A control design was developed based-on coordinate estimation for visual servoing of the mobile manipulator. Experiments on a self-constructed mobile manipulator reveal that the proposed method can find and grasp a target object successfully.**

Keywords- Object recognition, visual servo, mobile manipulator, stereo vision

I. INTRODUCTION

Vision is one of the most important perception sensors of human beings. It provides much useful information for human to recognize people and objects. In robotics, computer vision has been widely studied to allow a robot to understand its surroundings and interact with people. One typical application of computer vision is to search and track an interested object in the environment in order to fetch it.

According to the location of camera, vision-based grasping design can be divided into two types, eye-in-hand and eye-on-head. Prats *et al.*[1] presented an eye-in-hand approach to sensor-guided task execution. Task-oriented grasping algorithms were developed to plan a suitable grasp position of an object. However, in eye-in-hand design the camera will not be able to view the environment after the object is grasped. If the camera is placed on the robot head, the robot can perform other tasks after the target object is grasped. Azad *et al.* [2] presented an eye-on-head system that used a stereo camera in order to deal with textured objects as well as objects that can be segmented globally by their shapes.

An object recognition system counts a great deal on successful extraction of sufficient features for distinguishing objects. Features of target object are essential for generating appropriate control signals in the visual-servo control loop. Scale Invariant Feature Transform (SIFT)[3] and Speed up robust feature (SURF)[4] have been widely employed in image-based feature extraction. On the other hand, stereo vision is a powerful tool for depth estimation in object grasping. In [5], Lee *et al*. present a 3D object recognition and pose estimation method based on combining SIFT features and geometric features in an image sequence. In [6], Jang *et al*. present a design of real-time task execution based-on spatial reasoning. A potential field path planning algorithm was used to determine the path for accessing and retracting of a robot arm in delivering an object. In [7], Lee *et al*. present a method of 3D modeling of workspace for manipulative robotic tasks.

Another popular strategy in robotic grasping is to use an image sensor to collect information about the target object as well as the environment. To make a robot recognize object in an intelligent home environment, Baeg *et al.* [8] presented a service robotic system of fast object recognition in a smart home setting. They used MPEG-7 visual descriptors including color and texture. In [9], Aiguo *et al.* a mobile manipulator system is demonstrated as a versatile platform for home service tasks. In their design, one part is the efficient and easy recognition of the environment about geometrical, physical information; another is mobile manipulation based-on such geometrical, physical and additional information.

In many cases, erroneous feature matching still happens when one uses powerful tools such as SIFT to handle feature matching problem. Random sample consensus (RANSAC)[10] algorithm has been employed to remove the outliers in SIFT and increase accuracy of feature tracking during grasping task. In [11], Marquez-Neilia *et al.* introduce a procedure for recognizing and locating planar landmarks in mobile robot navigation, based on detection of a set of interesting points. They used RANSAC to fit a homography and locate the landmarks. Panin and Knoll [12] presented an efficient, robust and real-time system for 3D object pose tracking in image sequences. In [13], Choi *et al.* present a real-time solution for estimating and tracking the 3D pose of a rigid object for image-based visual servo with natural landmarks. In [14], Tomono presents a method of object map building using object models created from image sequences captured by a single camera. For calculating optimal grasping pose, Yamazaki *et al.*[15] described a grasp planning for a mobile manipulator. A grasp planner can find a stable grasp pose from the automatically generated model which contains redundant data and shape error of the object.

From previous related works in mobile manipulation, it is observed that stable object recognition and its 3D pose

Manuscript received Dec. 10, 2010. This work was supported by the National Science Council of Taiwan, under grant NSC-96-2628-E-009-152-MY3.

estimation are crucial in visual servo control and still major challenges. Further, computation burden in image processing poses a major problem in practical applications. In this paper, we want to develop a method to reduce the computation burden and investigate real-time grasping control design. A novel method is proposed to use center point of the object imagery and homography to determine the pose of target object for grasping. By using this method, the robot can be guided successfully to grasp the target object.

II. OBJECT RECOGNITION

For a grasping task, the first step is to find the interested object in the current image. To facilitate real-time visual servoing, we adopted SURF method to extract features from images. SURF not only detects interest feature locations but also extract features around a feature location such that together they can be used to recognize features from different views of an object. SURF features are invariant to orientation, scale, and illumination variation in the image. After the target object database is built from SURF features, the appearance of a target object can be determined in the current image frame. Nearest neighbor algorithm was adopted to match features between database images and the current image. The nearest neighbor is defined as the keypoint with minimum Euclidean distance for descriptor vector such that:

$$\boldsymbol{d} = \left(\sum_{\boldsymbol{i}=1}^{128} (\boldsymbol{Des}_c(\boldsymbol{i}) - \boldsymbol{Des}_d(\boldsymbol{i}))^2 \right)^{1/2} \quad (1)$$

where Des_c is the current image feature descriptor vector and Des_d is the database feature descriptor vector. In this design the successful matching point is the nearest neighbor with less than 0.8 times the distance to the second-nearest neighbor.

A. Robust Feature Extraction using RANSAC

Even with a good feature descriptor like SURF, matching error cannot be avoided. Those feature points which are matched incorrectly will introduce erroneous pose estimation in visual servo control. Therefore, outlier rejection is an important step to increase the robustness and accuracy of grasping control. The idea is to adopt RANSAC to both reject outliers robustly and find the homography matrix between database and current image. For a given plane in world coordinate system, there exists a homography matrix between every feature belong to the plane in an image and the corresponding features in another image. Assuming that

$$p_a = \begin{bmatrix} x_a \\ y_a \\ 1 \end{bmatrix}, \quad (2)$$

represents a feature point coordinate in an image frame (or in a database), there exists a 3x3 homography matrix

$$\mathrm{H}_{ab} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \quad (3)$$

such that

$$\begin{bmatrix} x_b \\ y_b \\ 1 \end{bmatrix} = p_b = H_{ab} p_a, \quad (4)$$

where p_b is the corresponding feature point coordinate of p_a in the current image frame. Note that $\mathbf{H}_{ab}$ has 8 degrees of freedom, since it is defined up to a scaling factor. In order to robustly estimate $\mathbf{H}_{ab}$ for outliers rejection, the RANSAC algorithm randomly selects various sets of matched feature points, estimate the homography between them, and find inliers using back-projection. After several iterations, the inliers as well as a robust $\mathbf{H}_{ab}$ will be found.

We then use the estimated homography matrix to calculate and find the center point of the object in the current image frame. Fig. 1 and Fig. 2 show the feature matching and

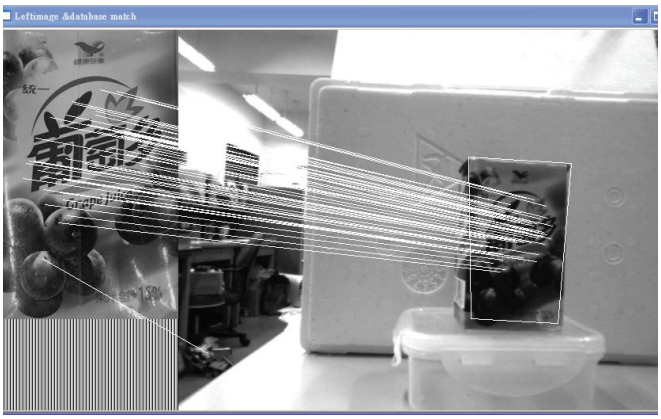


Figure 1. Matched feature points and homography estimation result in left image.

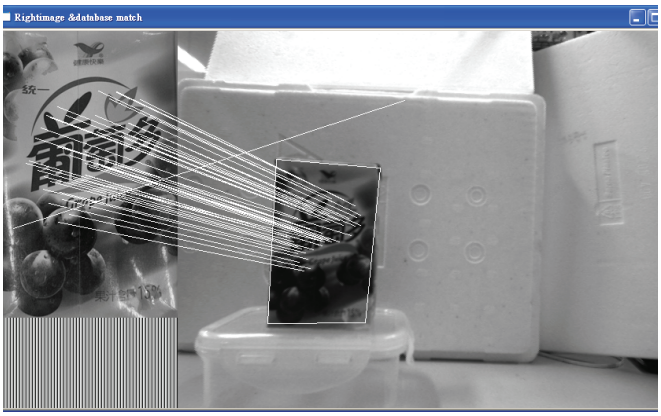


Figure 2. Matched feature points and homography estimation result in right image.

homography estimation results using the sttereo camera. If the location of center point is obtained accurately, the grasping control will be assured.

B. *Define the Center Point*

In the visual servo design, we need to select a robust reference point to guide the robot manipulator. After obtaining the homography matrix, we can use it to find the position of target object in current image. Then we can calculate four corner points of the object based-on the homography matrix. We then calculate the average of the four corners as center point of the object as shown in Fig. 3. The center point in left and right images are employed to calculate the 3D pose of the target. Furthermore, the coordinate of center point will be used as input signal for position-based visual servo control.

III. OBJECT POSE ESTIMATION

In order to get precise result of pose estimation, we adopted stereo vision as the tool for capturing images. Every feature point on the object will be projected on the left and right image plane, and the pairs of feature points are termed corresponding points. The location difference of corresponding points is termed disparity. Based-on the property of disparity, we can calculate the 3D coordinate of any corresponding object features on both image planes.

A. *Stereo Vision*

As shown in Fig. 4, a point in world coordinate and its projected points on both image planes exist a geometry relationship. The following formula to calculate the 3D coordinates from corresponding point on image plane is referenced from[16]. Let (u,v) be the point on image plane, λ is the focal length and B is baseline, we have.

$$z = \lambda - \frac{\lambda * B}{u_l - u_r} \tag{5}$$

$$x = \frac{u}{\lambda}(\lambda - z) \tag{6}$$

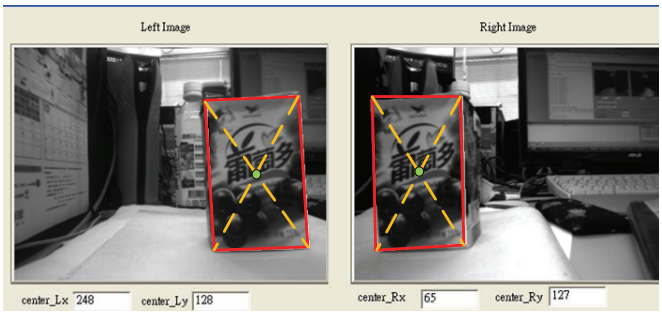


Figure 3. Target in image plane and location of center point.

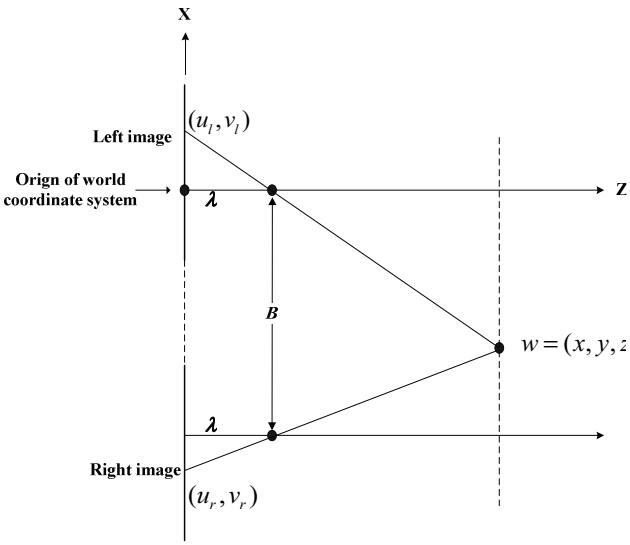


Figure 4. Top view of stereo imaging.

$$y = \frac{v}{\lambda}(\lambda - z) \tag{7}$$

B. *Calibration of 3D Coordinate Equations*

The values in equation were not equal to the world coordinate, a transformation from equation value to real value is needed. Therefore, we present a method to find the transformation for precisely 3D coordinate calculation. First, we recorded the location on both image planes at different distance for the same point, as shown in Fig. 5. The disparity would close to zero, when the distance between camera and object is far enough. Count all data and observe the relationship between equation value and real value. From the statistics we can find out that there exists a linear relationship between equation value and real value. We use curve fitting to find the linear equation of those data. Finally, we can use the corresponding points to get its 3D coordinate.

By using the this method, we can precisely estimate the position of object in world coordinate. Through calibration experiments, the error between estimate value and real value is below 10mm within the optimal recognition range of 1.5m.

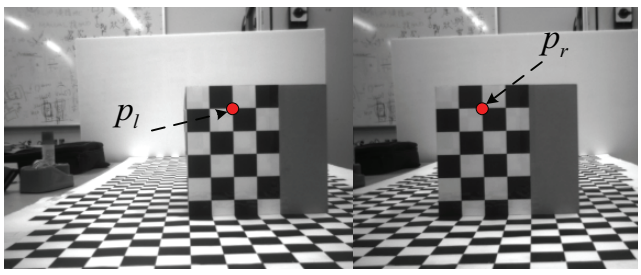


Figure 5. The location of same point on both image planes.

IV. VISUAL SERVO DESIGN

Position-based visual servo(PBVS) approach [17][18]was adopted in this work to verify the effectiveness of object recognition and 3D coordinate estimation algorithms presented above. A stereo camera was installed on top of the robot. In the image-based object grasping task, we compute the relative position between robot and target to guide the robot to move toward the target, and then grasp it. Thanks to center point, the target object position is estimated in real-time for the controller. The 3D coordinate of object's center point is used as the input, thus the visual servo design is simplified. The visual servo controller guides the gripper to pre-grasp point. The gripper then moves along the grasping path guided by the extracted center point. Finally the robot closes the gripper and complete the task.

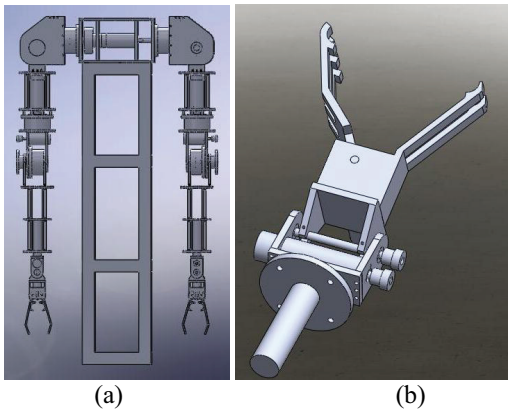
A. The Mobile Manipulator

As shown in Fig. 6, a 6 DOF robotic arm has been designed and constructed in the lab for experiments. Harmonic drives were employed at shoulder and elbow to eliminate backlash and give smoother motion. The EPOS controllers from Maxon were used for joint servo control. We turn the position command as input of the controller.

A typical experiment setup is such that the object stands upright on a desk, and the end-effector will always keep parallel to the ground for grasping the upright object. The gripper has 2 DOF, as shown in Fig 6(b). Because the gripper will not always access the same position relative to the object, we therefore planned a grasp locus for grasping the object. As shown in Fig. 7, when the gripper reaches the pre-grasp point, wrist, pre-grasp point, and object center will be aligned to a straight line. Then the gripper moves along the line with assigned distance, and it finally grasp the object.

B. Omni-Directional Platform

Because this work aims to study mobile manipulation, the dual-arm robot manipulator was installed on an omni-directional mobile platform. The motion model of the omni-directional platform was presented in [19]. Fig. 8 shows the kinematic motion model of the omni-directional mobile robot. When the robot moves after time step Δt , the location of robot



(a) (b)

Figure 6. Design of the experimental dual-arm manipulator (a) the 6 DOF robotic arm, (b) the gripper.

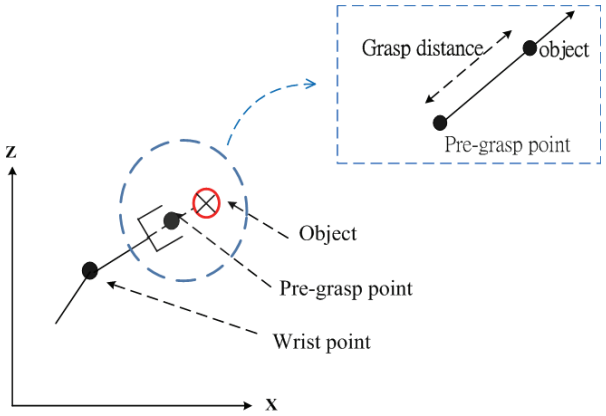


Figure 7. Grasping locus planning.

can be expressed such that.

$$\begin{bmatrix} x_k \\ y_k \\ \theta_k \end{bmatrix} = \begin{bmatrix} 0.5[-\sin(-\frac{\pi}{4}+\theta)v_{1k} - \sin(\frac{3\pi}{4}-\theta)v_{2k} + \sin(-\frac{\pi}{4}+\theta)v_{3k} + \sin(\frac{3\pi}{4}-\theta)v_{4k}]\Delta t \\ 0.5[\cos(-\frac{\pi}{4}+\theta)v_{1k} - \cos(\frac{3\pi}{4}-\theta)v_{2k} - \cos(-\frac{\pi}{4}+\theta)v_{3k} + \cos(\frac{3\pi}{4}-\theta)v_{4k}]\Delta t \\ \frac{1}{4L}(v_{1k}+v_{2k}+v_{3k}+v_{4k})\Delta t \end{bmatrix} + \begin{bmatrix} x_{k-1} \\ y_{k-1} \\ \theta_{k-1} \end{bmatrix} \quad (8)$$

C. Proposed System Architecture

The system architecture of the proposed visual servo and grasping control system is shown in Fig. 9. The system first get feature points from current left and right image by using SURF. It then removes the outliers within the matched pairs. Then the homography matrix between current image and database is computed. For visual servoing, the pair of object center points in both image planes was selected as reference points by homography estimation. Then we compute the 3D coordinate of object center, and get the position command for motion control. The visual servo system then guides the robot successfully to grasp the target object.

V. EXPERIMENTAL RESULTS

Several interesting experiments are presented in this section to examine the proposed visual servo control design.

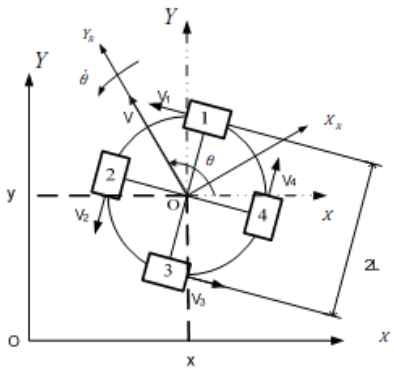


Figure 8. Motion model of the omni-directional platform.

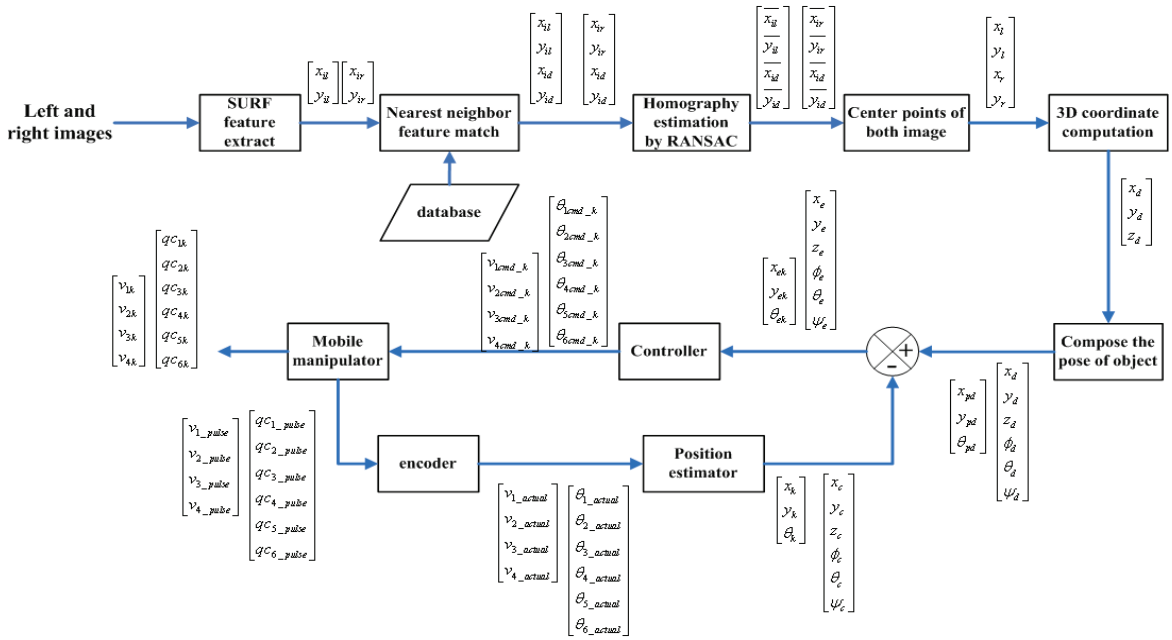


Figure 9. Control architecture of the mobile manipulator.

A. *Grasping of an Object in Manipulator Workspace*

In this experiment, the target object was placed in the workspace of the mobile manipulator. To grasp the object, the robot only needs to calculate the target position and guide the gripper to the grasp pose. In the experiment, the target was 105 cm from the ground and the distance between robot and target was 60cm. First, the object was detected in current image by matching SURF feature pairs between the database and current image. By using RANSAC and Homography, the position of target object was determined in the image plane. As the center of target object in left and right image plane were obtained, the 3D coordinate of the target object were calculated by the pair of center points. The estimated target position was then used as input of visual servo system. The robot then moved to the pre-grasp position and grasped the target. A video clip of the experiment can be found in [20].

B. *Target Outside Manipulator Workspace*

In this experimental, the robot started from a position in front of the target 100 cm away, which is outside of the reach range of the manipulator. When the robot started to track the target object, it would know that the target position was outside workspace. According to the 3D coordinate of target, system would decide the direction and distance for the robot to move. Fig. 10 shows the scene of stereo camera the moment in the beginning when robot stands at a distance away from the object. Fig. 11 shows the image processing result of estimated center points.

When the target object is recognized and the pose of the object is calculated, the robot will move towards the object and come into the workspace for an optimal grasping position. Fig. 12 shows the scene of the stereo camera when it comes to a location with the object within the workspace. Fig. 13 shows the corresponding result of estimated center points. We observe

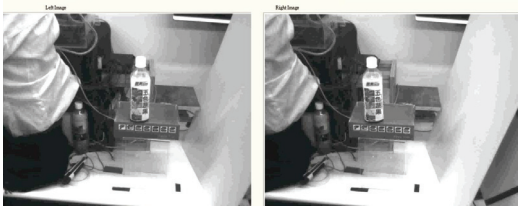


Figure 10. Left and right images of stereo camera at initial position.

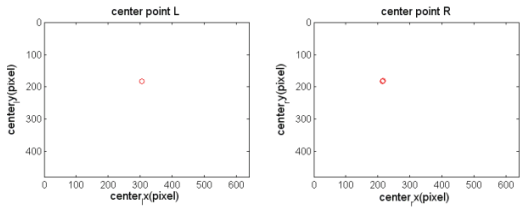


Figure 11. Center point of the object in image plane at initial position.

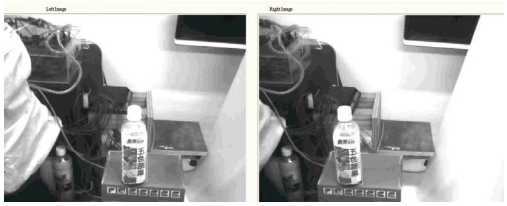


Figure 12. Left and right images of stereo camera at grasp position.

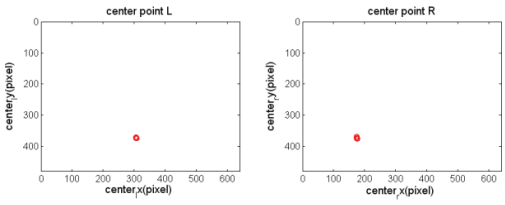


Figure 13. Center point of the object in image plane at grasp position.

that the location of the target moves forward the robot in the image, due to robot motion with a fixed camera on the robot head. Fig. 14 shows the sequence of motion of mobile manipulator in the experiment. Fig. 14(a)-(d) shows the process of moving towards the target object and Fig. 14 (e)-(f) shows the grasping of the target object. A video clip of this experiment can be found in [21]

VI. CONCLUSIONS

With SURF features, the robot can find a target object in the environment. RANSAC and Homography effectively enhance correct matching of feature pairs. These tools substantially increase the overall stability of the system. The proposed 3D coordinate estimate method can accurately obtain the position of object in the environment. Using the center point, the computation of 3D pose is greatly reduced. The center point successfully guide robot move toward the object. Experimental results reveal that the mobile manipulator can find and grasp a target successfully. In the future, a SLAM algorithm will be integrated to investigate mobile manipulation control design in a home environment.

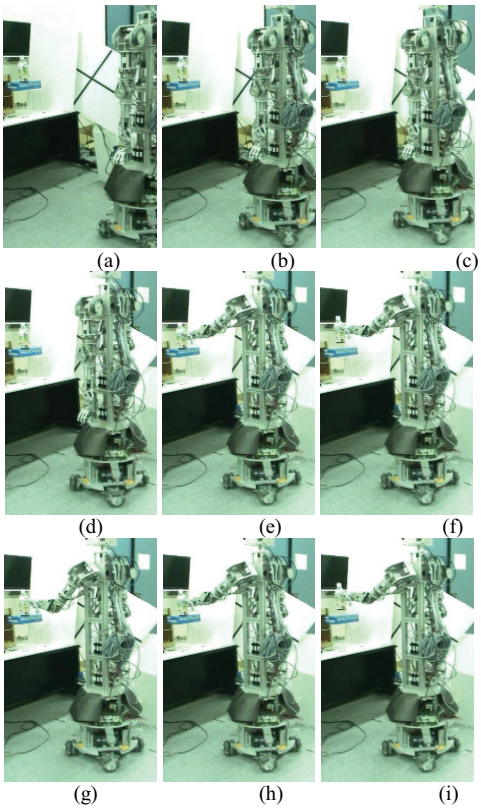


Figure 14. The sequence of grasping experiment.

REFERENCES

[1] M. Prats, P. Martinet, A. P. del Pobil and Sukhan Lee, "VIsion Force Control in Task-Oriented Grasping and Manipulation," in *Proceedings of IEEE International Conference on Intelligent Robots and Systems*, San Diego, CA, USA, 2007, pp. 1320-1325.

[2] P. Azad, T. Asfour and R. Dillmann, "Stereo-Based 6D Object Localization for Grasping with Humanoid Robot Systems," in *Proceedings of IEEE International Conference on Intelligent Robots and Systems,* San Diego, CA, USA, 2007, pp. 919-924.

[3] D. G. Lowe, "Distinctive Image Features from Scale-invariant Keypoints," International Journal of Computer Vision, 2004, pp.91-110.

[4] H. Bay, T. Tuytelaars and L. V. Gool, "SURF: Speeded Up Robust Features," *in Proc. of the 9th European Conf. on Computer Vision,* Graz Austria, 2006, pp. 404-417.

[5] Sukhan Lee, Eunyoung Kim and Yeonchool Park, "3D Object Recognition Using Multiple Features for Robotic Manipulation," in *Proceedings of IEEE International Conference on Robotics and Automation,* Orlando, Florida, USA, 2006, pp. 3768-3774.

[6] Han-Young Jang, Moradi Hadi, Suyeon Hong, Sukhan Lee and JungHyun Hong, "Spatial Reasoning for Real-time Robotic Manipulation," in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*, Beijing, China, 2006, pp. 2632-2637.

[7] Sukhan Lee, Daesik Jang, Eunyoung Kim, Suyeon Hong and JungHyun Hong, "A Real-Time 3D Workspace Modeling with Stereo Camera," in *Proceedings of IEEE International Conference on Intelligent Robots and Systems*, Edmonton, Alberta, Canada, 2005, pp. 2140-2147.

[8] S. H. Baeg, J. H. Park, J. Koh, K. W. Park and M. H. Baeg, "An Object Recognition System for a Smart Home Environment on the Basis of Color and Texture Descriptors," in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*, San Diego, CA, USA, 2007, pp. 901-906.

[9] M. Aiguo, X. Zhaoxian, T. Yoshida, M. Yamashiro, T. Chao and M. Shimojo, "Home Service by a Mobile Manipulator System -System Configuration and Basic Experiments," in *Proceedings of IEEE International Conference on Information and Automation*, Zhangjiajie, China, 2008, pp. 464-469.

[10] M. Fischler and R. Bolles, "Random Sampling Consensus: a Paradigm for Model Ftting with Application to Image Analysis and Automated Cartography," *Commun. Assoc. Comp*. Mach., 1981, vol. 24, pp. 381-395.

[11] P. Marquez-Neila, J. Garcia Miro, J. M. Buenaposada and L. Baumela, "Improving RANSAC for Fast Landmark Recognition," in *Proc. of Computer Society Conference on Computer Vision and Pattern Recognition*, Anchorage, AK, 2008, pp. 1-8.

[12] Giorgio Panin and Alois Knoll, "Fully Automatic Real-Time 3D Object Tracking using Active Contour and Appearance Models," *International journal of multimedia*, 2006, vol.1, no.1, pp.62-70.

[13] Changhyun Choi, Seung-Min Baek and Sukhan Lee, "Real-time 3D Object Pose Estimation and Tracking for Natural Landmark Based Visual Servo," in *Proceedings of IEEE International Conference on Intelligent Robots and Systems*, Nice, France, 2008, pp. 3983-3989.

[14] M. Tomono, "3-D Object Map Building Using Dense Object Models with SIFT-based Recognition Features," in *Proceedings of IEEE International Conference on Intelligent Robots and Systems*, Beijing, China, 2006, pp. 1885-1890.

[15] K. Yamazaki, M. Tomono, T. Tsubouchi and S. Yuta, "A Grasp Planning for Picking up an Unknown Object for a Mobile Manipulator," in *Proceedings of IEEE International Conference on Robotics and Automation*, Orlando, Florida, USA, 2006, pp. 2143-2149.

[16] K.S. Fu, R.C. Gonzalez and C.S.G. Lee, *Robotics Control, Sensing, Vision, and Intelligence*,McGraw-Hill Book Company, 1987.

[17] Geoffrey Taylor and Lindsay Kleeman, *Visual Perception and Robotic Manipulation*, Springer, 2006.

[18] S. Hutchinson, G.D. Hager and P.I. Corke, "A Tutorial on Visual Servo Control," in *IEEE Transactions on Robotics and Automation*, 1996, Vol. 12, No. 5, pp. 651-670.

[19] Kai-Tai Song and Chao-Wu Wang, "Self-Localization and Control of an Omni-Directional Mobile Robot Based on an Omni-Directional Camera," in *Proceedings of 7th Asian Control Conference*, Hong Kong, China, 2009, pp.899-904.

[20] http://isci.cn.nctu.edu.tw/video/ASCC2011/1/

[21] http://isci.cn.nctu.edu.tw/video/ASCC2011/2/