Guiding Robot Motion Using Zooming and Focusing

Jiang-Yu Zheng, Toshiaki Sakai and Norihiro Abe

Faculty of Computer Science and Systems Engineering
Kyushu Institute of Technology, Japan

Abstract

This work explores hand-eye coordination of robot by using a camera with functions of zooming and focusing. Under the guidance of the camera, a robot hand can arrive at an object to manipulate it in the 3D space. The hand movement is decomposed into the move parallel to the image plane and the move in depth, which are controlled from image position and focus value of the hand. The focal value is put on the destination object and the hand move in depth tries to increase the image sharpness of hand. The zooming, focusing, and 2D image position control cooperate with each other and work in feedback fashion. The fovea control of eye realized by zooming from wide-angle to tele-angle brings both flexibility and accuracy in locating hand position.

1. Introduction

This work aims at controlling robot hand by a robot eye to manipulate objects. The problem to solve is to locate the robot hand at a destination in the 3D space with accuracy and flexibility. A camera with zooming and focusing functions is employed, which gives a compact system and needs less on site calibration. We explore the zooming and focusing mechanism of the camera for robot control.

The conventional visual servoing methods applied in a 3D space usually uses multi-cameras including stereo to locate a robot hand. Such multi-view-points measure usually needs more than two cameras or moves a camera to different positions [5,6]. Stereo needs a device that is at least wider than the camera baseline; otherwise the produced accuracy in depth is not enough for robot positioning. In this work, we use one camera with controlled zoom and focus to locate the robot hand. It can further be directed to different directions driven by a small motor.

Depth from focus or defocus approaches have been studied in vision[7-13]. Most of them have a very short range in which the depth could be measured accurately[14]. Beyond such a range, the accuracy of depth from focus is not as good as stereo. Using a fixed lens with a long focal length to yield accuracy in depth for distant objects has also been studied[7,10]. However, such a lens always limits the size of field of view. Robot manipulation is hard to perform if the eye is unable to catch a global view in which both destination and the hand are visible. In order to know global spatial relation between the hand and objects and acquire accuracy in depth as well, we use zoom function of the camera in this work.

The basic idea to control the hand is visual feedback. We control the hand motion to let it approach a destination that is an object or a position with marks given in the 2D image. At the same time, if the camera focuses on the destination, the image of the hand should become clear as the hand moves towards the destination, which means its depth becomes close to that of the destination. The zoom changes from wide-angle to tele-angle which increases the resolution of 2D images and sensitivity of focus measure so that the positioning accuracy of hand is improved. If a destination is indicated in the viewed image, the hand will arrive at that position automatically.

Through the zooming process, the camera can understand hand location and moves the hand to the destination precisely. We will reduce the number of images taken for focus searching by coarse to fine zooming so as to achieve speedy guidance of the hand.

In the following, we will first analyze the characteristics of the camera zoom and calibrate a zoom camera. Based on the analysis and the calibrated data, a strategy of motion guidance of the robot hand is designed. Finally, experiments on moving robot hand to real objects will be shown. The accuracy achieved will be 1cm in depth and 1mm along the image plane.

2. Depth from Focus with Zooming

2.1 Depth from Focus

Let us first look at the relations of focus, zoom, field of view, depth, accuracy, etc. The simplest model of a thin lens is described by

$$\frac{1}{f}=\frac{1}{F}+\frac{1}{D} \tag{1}$$

where D is the distance of a focused point on an object from the lens, and F is the focus value. The focal length of the lens changed by zoom is denoted as f here and we have F>f. The principle of the image formation is shown in Fig. 1. A shift of the image plane from the in focused position will blur image. The focus value F is determined by changing the image plane to a position where the focused point is clearest in the image. Various ways to evaluate the clearness or sharpness of view have been considered[14]. A simple parameter used here is the differential value of intensity which gives contrast at edge positions in the image.

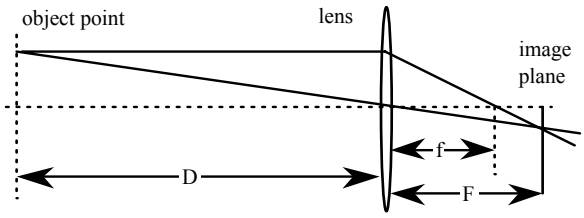


Fig. 1 A simple lens model of depth from focus.

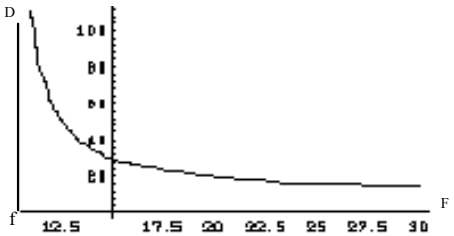
From Eq. 1, we have

$$D = \frac{Ff}{F-f}, \quad \text{and} \quad F = \frac{Df}{D-f} \tag{2}$$

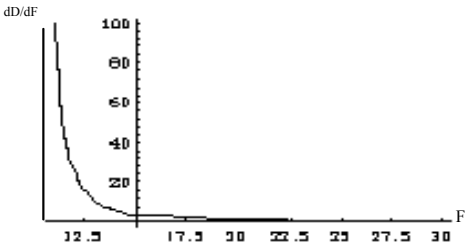
If the distance of the focused point from the camera is estimated as d (the measure of D), its error due to the inaccurate focus value can be computed by

$$\left|\frac{\partial d}{\partial F}\right| = \frac{f^2}{(F-f)^2} \tag{3}$$

The corresponding focus-distance graph and error distribution due to the inaccurate focus a can be given in Fig. 2(a-b). From Fig. 2(b), we can know that the error tends higher for distant object (small F value when focused) than close one.



(a) a graph describing the relation between focus and measured distance of a 3D point.



(b) relative error of measured distance of a point due to a deviation of focus value from the true one.

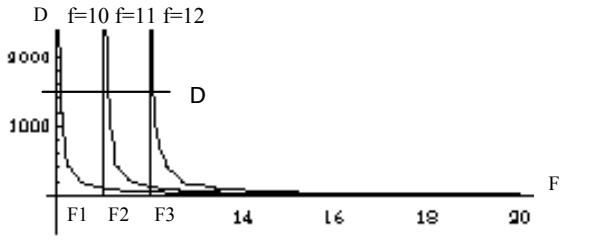
Fig. 2 Focus-distance relation and error distribution from focus value.

2.2 Accuracy of Depth Related to Zooming

If the zoom changes, we have a sequence of focus-distance graphs given in Fig. 3(a) which is obtained by shifting graph of Fig. 2(a). According to the focal length, the graph for tele angle is on the right. For a particular distance D>>f, the infocused position of F under different zooming values are slightly different. We denote them as F(f). We have to refine the focus position when the zoom changes from wide angle to tele angle. The error computed from Eq.(3) is no longer a constant. It is related to the zoom value, which becomes

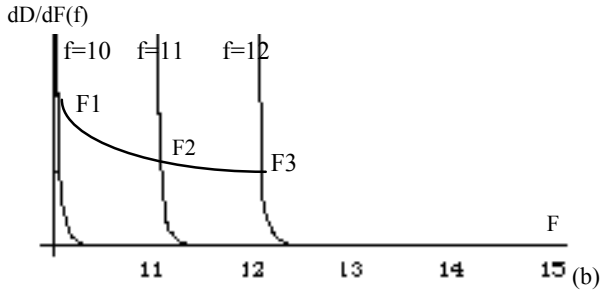
$$\left|\frac{\partial d}{\partial F}\right| = \frac{f^2}{(F-f)^2} = \frac{f^2}{\left(\frac{Df}{D-f} - f\right)^2} = (D-f)^2 \tag{4}$$

It decreases when the focal length increases (the zoom approaches tele-angle) as illustrated in Fig. 3(b). That means the accuracy of depth measured from the focus value is better at tele-angle than at wide angle.



(a)

focus distance relation changing with zooming.



(b)

error of measured distance from focus value is reduced when using tele-angle of zoom.

Fig. 3 Focus-distance relations and error distributions for different zoom values.

2.3 Sensitivity of Focus Value under Zooming

In order to evaluate sharpness of features viewed in the images, the differential values at an edge point are calculated for different focus values. For an edge taken at different zoom values, the distributions of focus values are also different.

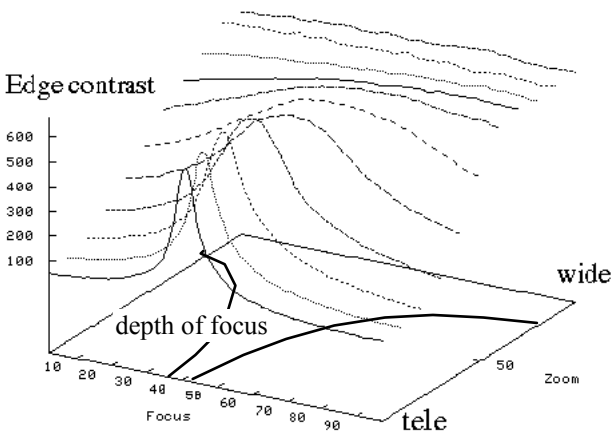


Fig. 4 sharpness distribution of an edge or edge contrast changing with focus and zoom. Three axes indicate values of zoom, focus, and contrast.

Figure 4 shows the focus-contrast graphs for different zoom values calibrated by using a step edge at a given distance. At a wide angle, the edge is sharp over a wide range of focus values. The peak of contrast that reflects the distance is hard to locate for the flat distribution taken with a wide angle. It gradually becomes distinguishable when the zoom tends to tele-angle. At a tele-angle, the image blurs immediately if the focus shifts away from the in focused position. We can therefore locate the peak of contrast for the desired focus value. The range of focus

where a target looks sharp is described by depth of focus.

If the depth of focus is short, edges outside it is blurred so that following targets in the image may become difficult. The in focused position is sensitive to be detected, which yields an accurate focus value.

2.4 Field of View Depending on Zooming

When the zoom moves to tele-angle, the field of view becomes small and the resolution of scene becomes high. Figure 5 gives the zoom-scale graph calibrated for the camera we use. When the zoom value changes, objects have expansions in their positions from the image center. The moved feature points can be predicted so that tracking hand and other objects during zooming is more stable and needs low cost.

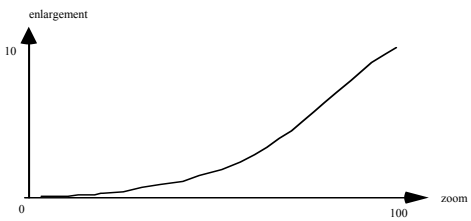


Fig. 5 Zoom-scale graph reflects magnification of a zoom value.

3 Robot Guidance with Zoom and Focus

3.1 Multi-modules in Hand Motion Control

Figure 6 shows the principle of hand guidance. The motion towards the destination can be decomposed into vectors in two directions. One is the vector in the plane parallel to the image plane and the other is the vector in depth. The motion is controlled by visual feedback in two different spaces; one is the 2D image space and the other is the space of focus value. The camera controls the hand until its image projection arrives that of destination. If the focus is set at the destination, the hand image becomes sharpest when it reaches the same depth with the destination. In order to do this, images that focus on the destination are continuously taken as the hand moves.

The robot position control using focus and 2D image are not enough to bring the hand to destination precisely. If the focal length is short, the camera has a large depth of focus There is no remarkable change between the hand and the destination in their edge sharpness even they are far away from each other in depth. On the contrary, if the focal length is long, the hand may go out of the small field of view or becomes completely blurred so that tracking it in continuous images is no longer possible. To solve that problem, we use the zoom in this motion control.

The zoom, focus and 2D image distance between the hand and a destination are controlled separately; each of them performs feedback control. They all have an explicit objective and at the same time coordinate with each other. The following gives the detail stop conditions of each module (Fig. 7).

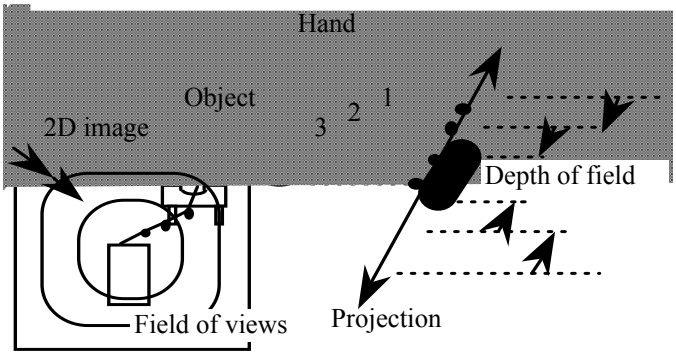


Fig. 6 The hand approaching destination under the guidance of camera.

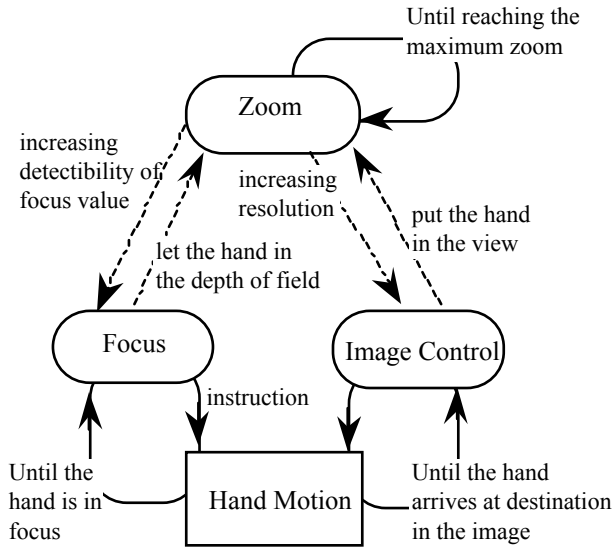


Fig. 7 Three cooperate modules in guiding the robot hand to the destination. Their processes work with feedback for until arriving their stopping conditions.

2D image control: the image distance between hand and destination is measured. The robot hand moves in the direction planned in the image until its projection arrives the planned image position.

Focus: it selects the focus value on which the destination is in focus. Then the robot hand moves in depth until it is also in focus (at the same depth with the destination).

Zoom: it increases gradually until reaching the maximum value; each move keeps the hand inside the field of view and, at the same time, not being blurred too much to be followed in the images. The move from wide angle to tele-angle increases spatial resolution both in the XY plane and in focus value (depth of field becomes small so that the edge contrast is sensitive to small focus change in the depth of focus). It eventually increases accuracy in depth measure.

Three processes repeatedly work until all of them reach their stop conditions. This forms a hand locating task. More general tasks could be decomposed into such low level tasks. The indication of destination could be considered by using object recognition, but in this work we assign destination with mouse through a man-machine interface.

3.2 Feedback Control of Hand Position

According to the calibrated zoom-scale relation, the zoom process selects a focal length (zoom value) as long as possible which can keep the hand in the field of view. As the zoom approaches to the focal length, the image of hand may be blurred since the hand has not been in focused at the increased zoom level. The hand is tracked in the continuous images and its image position and size are prediced. The zooming process stops when the contrasts of edges in the hand area become low (the hand image blurs) or the expected focal length is reached.

The control of the hand in the XY plane is simple. A path of hand is drawn to the destination in the image. The hand moves along the path parallel to the image plane and its image is tracked continuously (the resolution and focus are the same so that the tracking has less problem). The feedback control moves the hand until its projection arrives the planned position in the image, which is under the resolution determined by the current zoom.

The focusing process then becomes active. It changes focus value searching the position where the destination of hand is in focus. The contrast values obtained during the change of focus can tell if the hand is in front of the destination or behind it. The destination is then precisely in focused.

Now, the hand moves in depth. The move is tracked in the image and stops when the hand is in focus at that zoom level. In order to focus at the destination and make the hand in focus, we evaluate the average value of several biggest contrasts in the window that covers the hand. The contrast is obtained from computing differential values using Sobel operator. The biggest contrast values are taken from edge points. If the average of the biggest values changes from increasing to decreasing, there must have a peak in the contrast distribution where the hand or destination is sharpest in the view. We then locate the peak and obtain the corresponding focus value.

The move step of hand in depth is determined from the zoom value of that time. Because for a wide angle, contrast distribution has no significant global peak reflecting the true focus value but has many local peaks due to disturbance (Fig. 4), the hand has a big step to move (long interval of moving distance) so as to avoid stopping at any local peak of contrast value. If a peak is reached with a big step, the search of fine location of real peak at the binary divided position of the step is followed. Even though, the accuracy of the location is limited by the zoom value at that stage. The further improvement is at following stages when the zoom value increases.

The tracking of the hand and the destination object is by correlation and edge matching. When the size of hand and destination become large during zooming, the window covering them is also increased. The contrasts are then evaluated in large windows.

The feedback control of the hand needs no complicated calibration at the robot working site. We do not measure the exact distance between the robot hand and the destination. Only moving direction is given to the hand each time from the camera. The destination is reached in a flexible way.

4. Experiments

In our experiment, the camera position and direction is known in the motion space of the robot hand through a very simple calibration. The camera coordinates system therefore is known for the robot arm. A computer driven zooming lens (Canon J10¥10REA-11 PZF1) has focal length from 10 mm to 100 mm. The maximum magnification of the field of view is about 10. The range for imaging is 1.2m~•. Image size is 640¥480 pixels. We assume the hand grasps object from the top. In order to achieve both efficiency and accuracy, the hand moving distance instructed by one feedback loop is different for different zoom value. For both motions in depth and the XY plane, the moving distances (steps) are coarse for wide angle of zoom and fine for tele-angle of zoom.

The highest spatial resolution of using the zoom lens estimated from the calibration are 10mm in depth and 0.16mm in the XY plane when objects are put at the distance of 1m (increasing object distance may decrease accuracy a little) and zoom has the maximum value. The accuracy of the robot move is 0.32mm. Using a zoom with longer focus length will increase the accuracy both in depth and the 2D space parallel to the image.

Figure 8 gives a sequence of images when the robot hand approaches a destination. A scale-invariant pattern is attached on the hand for tracking its images. We hence use a window smaller than that for contrast evaluation in hand tracking. The real achieved position accuracy at the maximum tele-angle is 1.0 cm in depth and about 0.5 mm in the XY plane of the camera, which is not so big when the camera is 1.5m away from the object. It is lower than expected accuracy because of the position error in image processing and robot hand controlling. The error comes from inaccurate focus value, which is further related to the inaccuracy of zoom lens. We have tested that the focal length is not constant in the whole view. At the same time, locating the same aera on the object to compute constrast for focus value is important and should be done precisely.

To apply this method to robot manipulation, surface texture is needed on the object surface for focus. Beside a case with printed pattern, we have also tried some shiny metal which has highlights on the surface (Fig. 9) and some bolt with distinct shading and shadow on it. Their results of locating robot are similar with that shown in Fig. 8.

5. Conclusion

In this paper, we introduced a vision approach and an eye system to control motion of a robot hand in the 3D space. By employing a zoom in the approach of depth from focus, the robot eye can not only observe a wide environment and select objects to focus, but also achieve a high accuracy in robot depth control. Visual servoing using this type of eye is better than conventional stereo and multiple view points measure in the compactness.

The robot hand is controlled in a feedback way so that there is no heavy calibration and the move of the hand is qualitatively instructed. Multi-processes work on zoom, focus, and 2D localization and cooperate with each other.

The path is planned based on 2D image and depth information from zooming and focusing. We are going to use this hand-eye coordinate system to do manipulation in an unstructured environment.

References

[1] S. Tsuji, and J. Y. Zheng, “Visual path planning by a mobile robot”, IJCAI, 1987.

[2] D.H. Ballard, "Reference frame for animate vision", Proc. IJCAI-89, pp.1635-1641, 1989.

[3] P. J. Burt, "Smart sensing within a pyramic vision machine", IEEE Proceedings on Computer Vision, pp.1006-1015, 1988.

[4] R. Bajcsy, "Active perception", Proc. IEEE special issue on Computer Vision, August 1988, pp.996-1005.

[5] J. Y. Zheng, Q. Chen, S. Tsuji, and F. Kishino, "Active camera control for manipulation", CVPR91, pp.413-418, 1991.

[6] J. Y. Zheng, Q. Chen and S. Tsuji, "Active camera guided manipulation", 1991 IEEE Int. Conf. Robotics and Automation, Vol. 1, pp. 632-638, 1991.

[7] H. N. Nair, C. V. Stewart, “Robust focus ranging”, CVPR92, pp. 309-314, 1992.

[8] P. Grossman, “Depth from focus”, Pattern Recognition Letters, vol. 5, pp. 63-69, 1987.

[9] N. Asada, H, Fujiwara, and T. Matuyama, “Edge and depth from focus”, 1st ACCV, pp. 83-86, 1993.

[10] S. K. Nayar, “shape from focus system”, CVPR, pp.302-308, 1992.

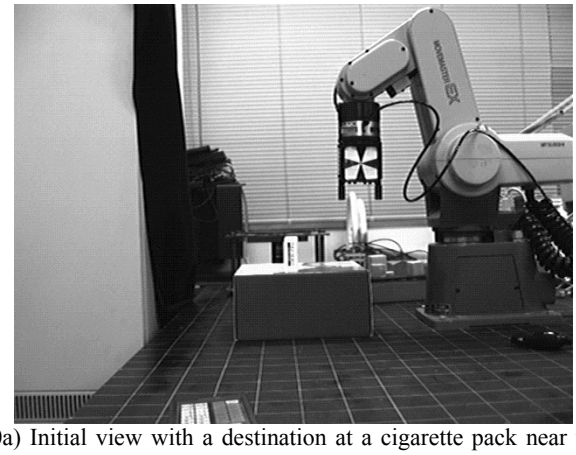
[11] A. P. Pentland, “A new sence for depth of field”, PAMI, vol. 9, pp. 523-531, 1987.

[12] E. Krotkov, “Focusing”, IJCV, vol.1, pp. 223-237, 1987.

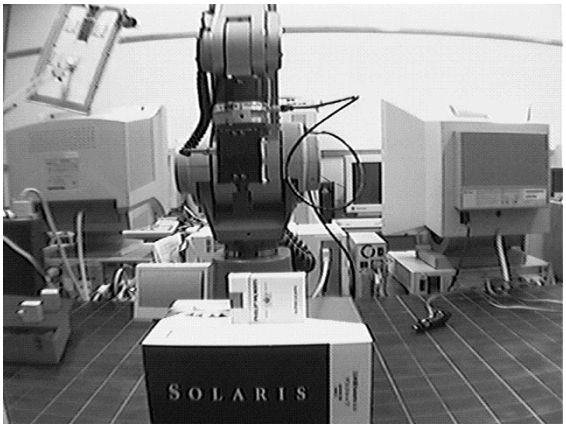
[13] A. Witkin, "Scale-space filtering", in IJCAI, Karlsruhe, pp.1019-1021. 1983.

[14] S. K. Nayar, Watanabe, Noguchi, "Real-time focus range sensor", 5th ICCV, pp.995-1001. 1995.

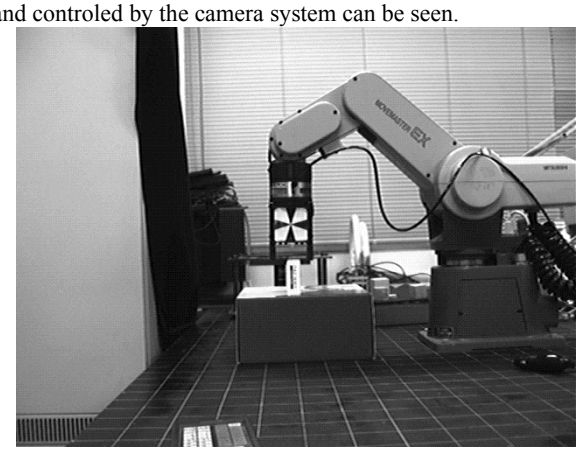
[15] N. Asada, H. Fujiwara, T. Matsuyama, "Seeing behind the scene: analysis of photometric properties of occluding edges by the reversed projection blurring model", 5th ICCV, pp.150-155, 1995.



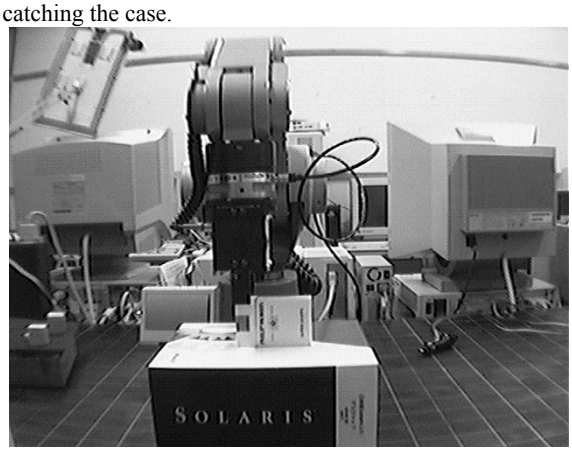
(0a) Initial view with a destination at a cigarette pack near the center of the image. The robot hand is recognized at a sharp and scale-invariant mark attached on it.



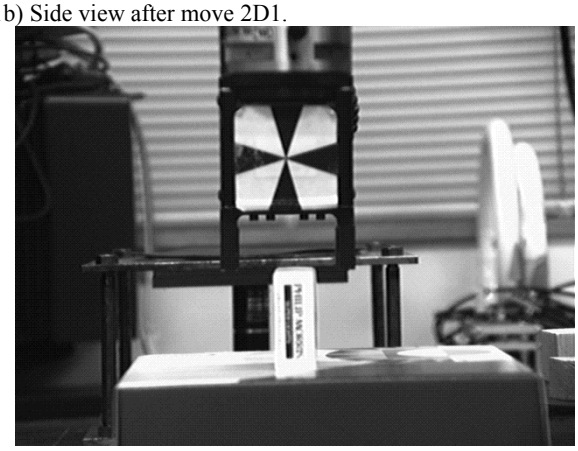
(0b) The side view of the hand where the depth move of the hand controled by the camera system can be seen.



(1a) 2D1: Moving down in the XY plane to the position possible to catching the case.

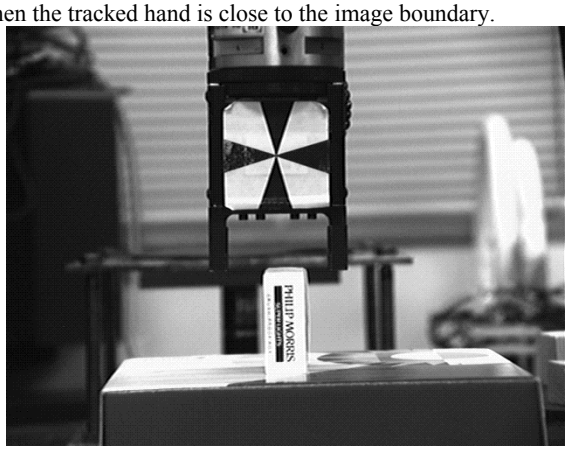


(1b) Side view after move 2D1.

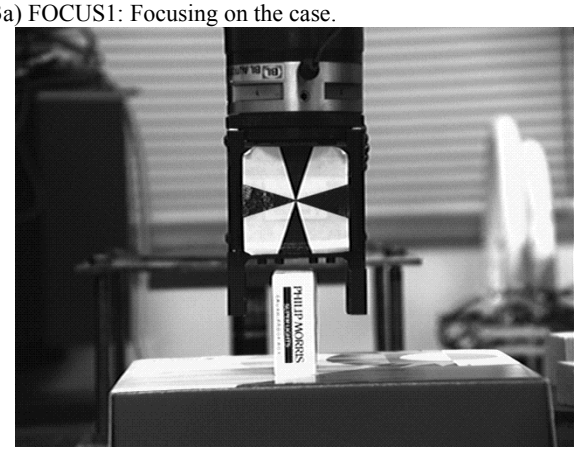


(2a) ZOOM1: Zooming the hand and the destination. It stops

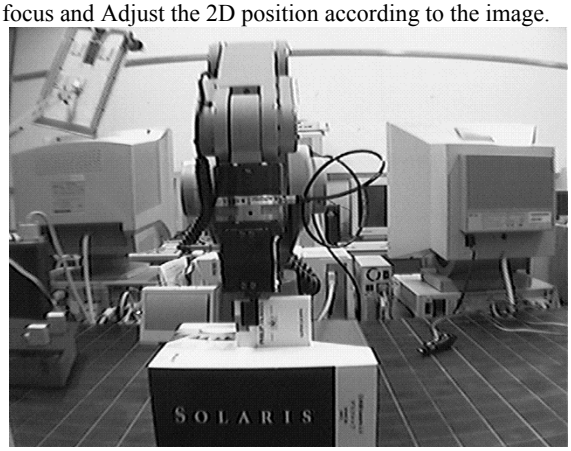
when the tracked hand is close to the image boundary.



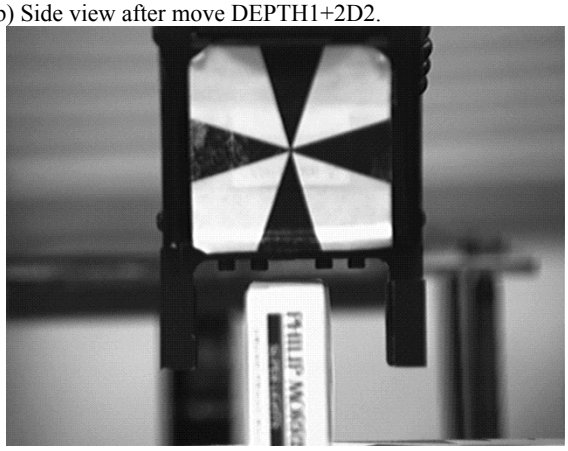
(3a) FOCUS1: Focusing on the case.



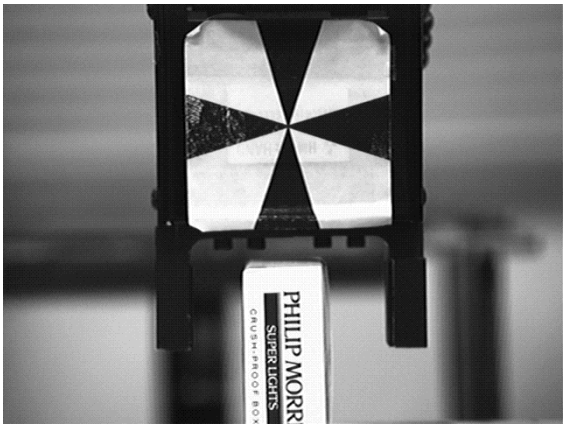
(4a) DEPTH1+2D2: Hand move in depth to make its image best in focus and Adjust the 2D position according to the image.



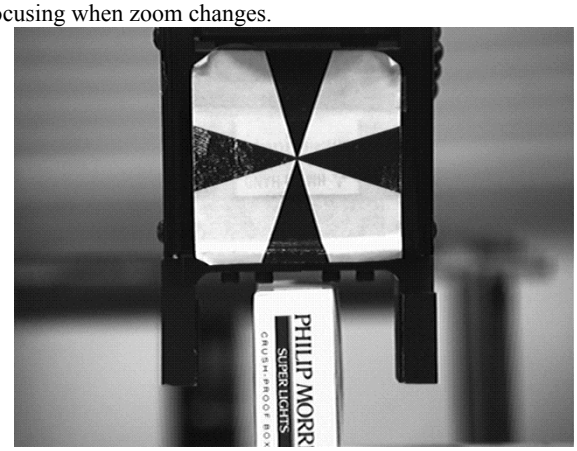
(4b) Side view after move DEPTH1+2D2.



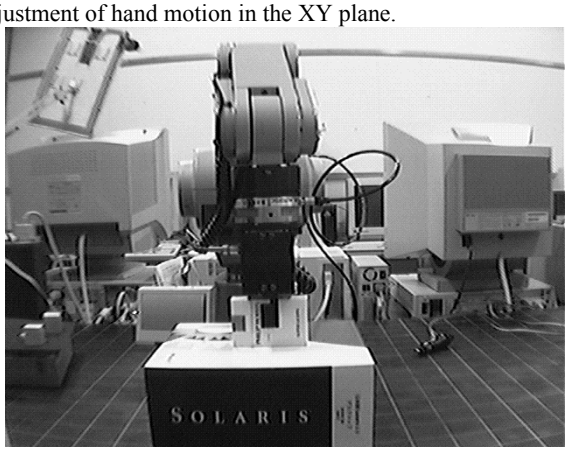
(5a) ZOOM2: The zoom enlarges the view again.



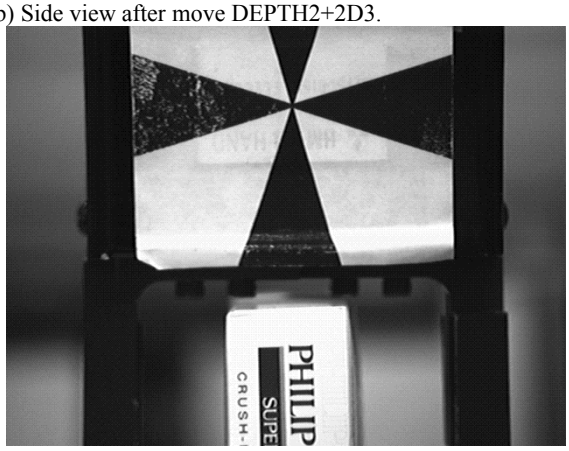
(6a) FOCUS2: Focus on destination again to improve the focusing when zoom changes.



(7a) DEPTH2+2D3: Hand motion in depth again and fine adjustment of hand motion in the XY plane.

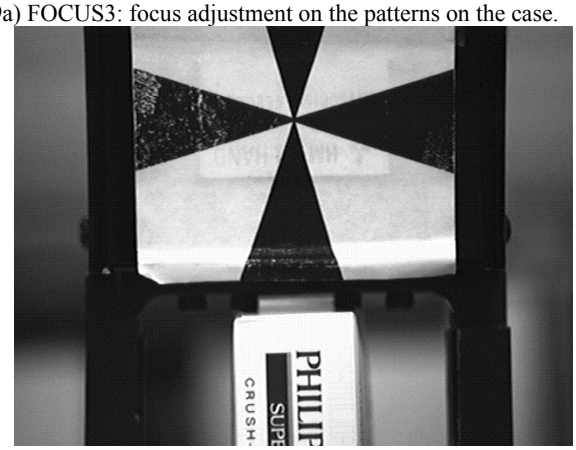


(7b) Side view after move DEPTH2+2D3.

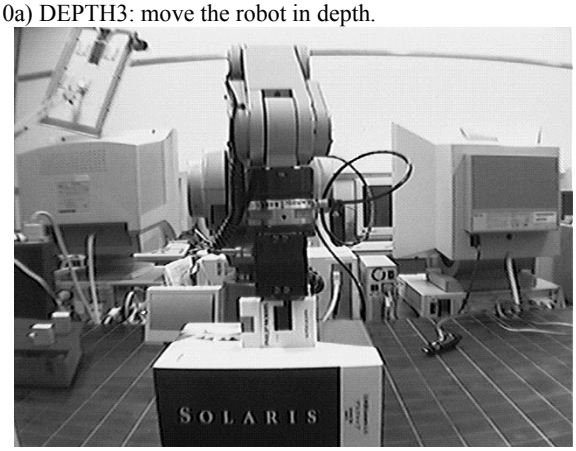


(8a) ZOOM3: zooming in to the hand mark.

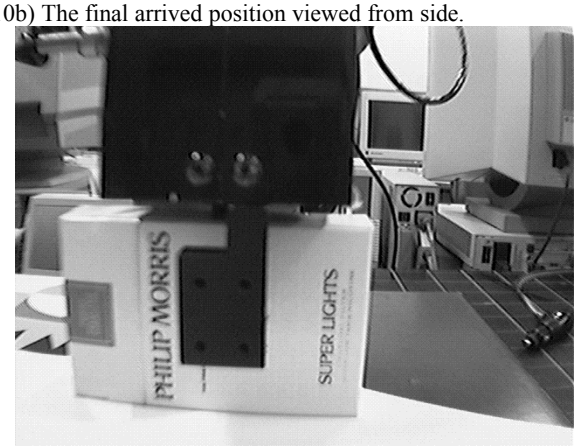
(9a) FOCUS3: focus adjustment on the patterns on the case.



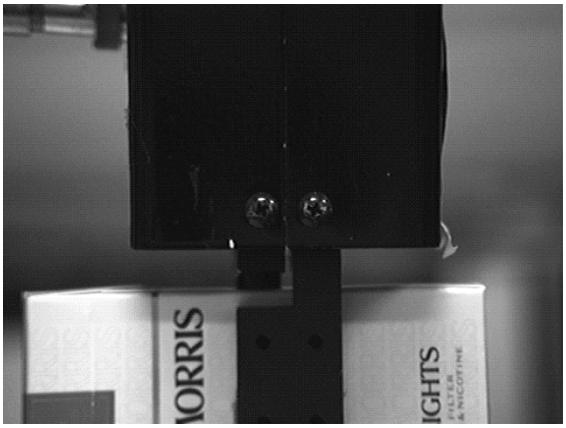
(10a) DEPTH3: move the robot in depth.



(10b) The final arrived position viewed from side.

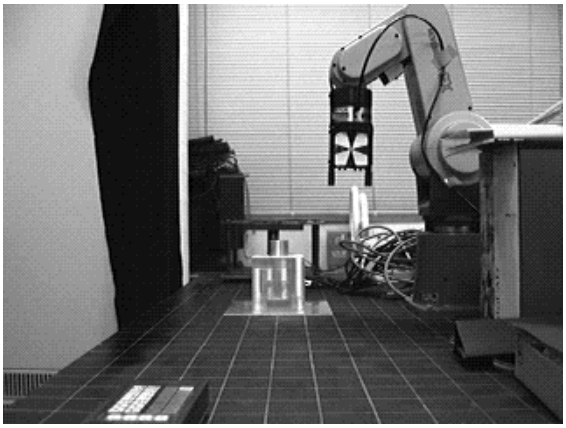


(10c) The front surface of the hand should fit the surface of object to grasp. The robot can pick up the case.

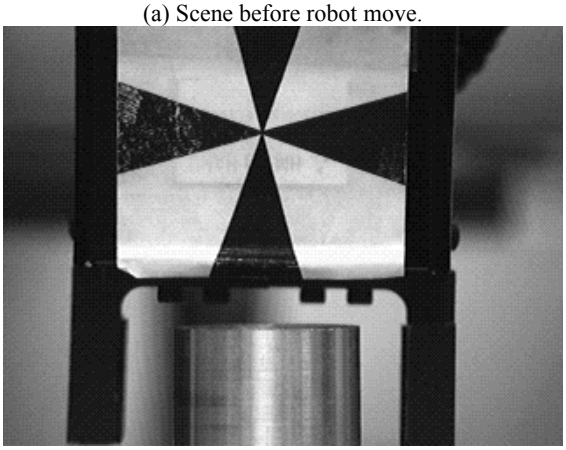


(11) The robot catches the case.

Fig. 8 A sequence of images showing how the robot hand approaches the destination at another object controlled by zoom and focus. The accuracy in depth is 1cm in this case.



(a) Scene before robot move.



(b) Final position of the robot.

Fig. 9 Grasp a cylindrical part using zoom and focus. The contrast is measured at the highlight on the surface.