UNOCCULUDED OBJECT GRASPING BY USING VISUAL DATA

Muhammet Ali ARSERİM [*], Yakup DEMİR [2], Ayşegül UÇAR[3]

[*1] Electrical and Electronics Engineering Department, Dicle University,Diyarbakir, Turkey

[2] Electrical and Electronics Engineering Department, Firat University, Elazig, Turkey

[3] Mechatronics Engineering Department, Firat University, Elazig, Turkey

E-mail: marserim@dicle.edu.tr

Automatic grasping objects can become important in the areas such as industrial processes, processes which are dangerous for human, or the operations which should be executed in the places, small for people work. In this study, it is aimed to design a robotic system for grasping unocculuded certain objects by using visual data. For this aim an experimental process was implemented.

Visual data process can be divided in two main parts: identification and three dimensional positioning. Identification issue suffers from several conditions as rotation, camera position, and location of the subject in the frame. Also obtaining the features invariant from these conditions is important. Therefore Zernike moment method can be used to overcome these negativities. In order to identify the objects an artificial neural network was used to classify the objects by using Zernike moment coefficients.

In the experimental system a parallel axis stereovision subsystem, a DSP-FPGA embedded media processor, and five-axis robot arm were used. The success rate of artificial neural network was 98%. After identifying the objects, a sequential algebra were performed in the DSP part of the media processor and the position of the object according to robot arm reference point was extracted. After all, desired object in the instant frame was grasped and placed in different location by the robot arm.

Key words*: Zernike Moment Method, Stereovision, DSP-FPGA Embedded System, Robot Arm, Artificial Neural Networks.Introduction*

1. Introduction

In autonomous systems robot manipulators can be used for picking, grasping, moving objects, and several tasks in different working areas. Defining coordinates of a target point can be done by using visual data. For this aim stereovision systems can be used for extracting three dimensional locations of the objects.

Identification based on visual data is investigated is two categories: using boundary of the object, and regional data of the object in the image frame [1]. Also the second approach is appropriate for both geometric, and Zernike moment method, used in this study.

Zernike moment method provides an advantage of defining coefficients which are invariant from translation, rotation, and scaling concepts. Therefore these coefficients can be used for classification for the objects [2-4]. Thus Artificial Neural Network (ANN) can be used for classification of objects by using Zernike moment coefficients as the inputs.

Using parallel axis stereovision systems simplifies the calculation of disparity between two frames. In such stereovision systems, reflection of the object in the associated frames displaces only on horizontal axis [5]. So distance from object to focus of chosen camera can be determined by using geometrical calculations.

Using special hardware for signal processing is important. They can provide rapid responses for processing and robustness. Controlling robot arm by visual processing in several problems is necessary. Since amount of visual data is too much, several analysis techniques should be used to reduce this amount. From this point Zernike moment method can be a good alternative for image processing.

2. Material and Method

In this study an experimental system was implemented as seen in Fig. 1. This system was consisted of a baseline stereovision block, which had two identical pine-hole cameras, a DSP-FPGA embedded media processor, and a robot arm with its controller unit.

Fig. 1. Experimental system.

The embedded media processor was SUNDANCE SMT339 board and it was a commercial product of SUNDANCE firm. It has a DSP (TMS320DM642), and a FPGA (Virtex-4 XC4VFX60-10) [6]. The FPGA component of this card was used to preprocessing of the raw image data and to send the video information to imaging devices. Also it was used to send coordinate information to robot arm controller unit. All the calculations were performed in DSP processor. During the study the software (DVL) written for the Sundance products was for only one camera whether it had two camera inputs. The related part of the software for using a camera was rearranged for the first time to achieve using two cameras for stereovision in this study.

Robot arm and its controller unit were SCORBOT-ER VPlus and Controller-A unit from Intelitek Company respectively. It is an articulated and 5-axis arm with gripper and it is designed as an educational arm. Its programming language is Advanced Control Language (ACL) and it can be directed via either Cartesian coordinates or joint angles. Also controller unit produces 20Khz PWM signals to drive joint servo dc motors [7].

2.1. Baseline Stereovision

As mentioned above image planes of two cameras are placed on a baseline, shown in Fig. 2 and if the cameras are identical disparity calculation is simplified as.

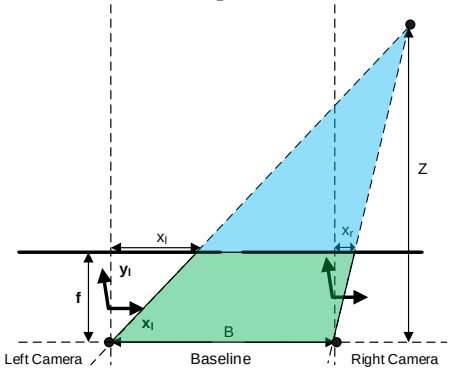

Fig. 2. Parallel stereovision.

$$\frac{Z-f}{Z} = \frac{B-(x_l-x_r)}{B} \tag{1}$$

where $d=x_l - x_r$. After rearranging (1), distance of the target point to camera focal point is calculated as;

$$Z = Bf/d \tag{2}$$

The other axial components of the point determined as

$$X = \frac{Zx_l}{f}, Y = \frac{Zy}{f} \tag{3}$$

2.2. Moment Method

Generally moments are used to determine several quantities by using distance with respect to a reference point [8]. Also moment method can be used to extract features from the images. Thus binary or grey level images can be regarded as a two dimensional density distribution function.

2.2.1. Geometric Moments

For a two dimensional continuous function, f(x,y), (p+q)th order geometric moment(mpq) is defined as [8]

$$m_{pq} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^p y^q f(x,y)\,dx\,dy \tag{4}$$

If this function is discrete and associated with a NxM piksel discritized image the integral form translates to summation equation, written in (5):

$$m_{pq} = \sum_{y=0}^{M-1}\sum_{x=0}^{N-1} x^p y^q f(x,y) \tag{5}$$

The 0th order moment, m_{00},

$$m_{00} = \sum_{y=0}^{M-1}\sum_{x=0}^{N-1} xy\, f(x,y) \tag{6}$$

is used to determine total mass of the function whereas first order moments, m_{10} and m_{01}, are used to obtain center of mass coordinates [9]:

$$\bar{x} = \frac{m_{10}}{m_{00}}, \quad \bar{y} = \frac{m_{01}}{m_{00}} \tag{7}$$

Pixel amounts of the object in the frame can change due to distance from camera or localization. In order to provide invariance from translation, shifting the reference point to center of mass leads to central moment coefficients as:

$$m_{pq} = \sum_{y=0}^{M-1} \sum_{x=0}^{N-1} (x - \bar{x})^p (y - \bar{y})^q f(x, y) \tag{8}$$

Also normalized central moments can be obtained by using

$$\mu_{pq} = \frac{m_{pq}}{m_{00}^{(2+p+q)/2}} \tag{9}$$

2.2.2. Zernike Moment Method

Teague [9] proposed first time to use orthogonal moments(Legendre and Zernike Moments) in image process. Zernike moment method process can be explained as projecting image on complex Zernike polynomials. Also this method is invariant from rotation by its nature.

Zenike polynomials of nth order is defined as

$$V_{pq}(x, y) = R_{pq}(\rho) e^{jq\theta} \tag{10}$$

and here $R_{pq(\rho)}$ is a real-valued radial polynomial, given as;

$$\sum_{\substack{m=0 \\ p-|q| \text{ çift}}}^{p-|q|/2} (-1)^m \frac{(p-m)!}{m! \left[\frac{(p-2m+|q|)}{2}\right]! \left[\frac{(p-2m-|q|)}{2}\right]!} \rho^{p-2m} \tag{11}$$

If k=p-2m is chosen, (11) can be transformed to (12)

$$R_{pq}(\rho) == \sum_{\substack{k=|q| \\ p-|q| \text{ çift}}}^{p} (-1)^{\frac{(p-k)}{2}} \frac{\left[\frac{(p+k)}{2}\right]!}{\left[\frac{(p-k)}{2}\right]! \left[\frac{(k+|q|)}{2}\right]! \left[\frac{(k-|q|)}{2}\right]!} \rho^k \tag{12}$$

After these definings, nth order, and q repeated Zernike moment of a two dimensional function, f(x,y) can be determined as

$$Z_{pq} = \frac{p+1}{\pi} \iint V_{pq}^*(x, y) f(x, y)\, dx\, dy \tag{13}$$

where the integral limits satisfy the x2+y2<1 condition. However the above integral can't be applied to discrete image function, so the integration is replaced by summation [10].

$$Z_{pq} = \lambda(p, N) \sum_{j=0}^{N-1} \sum_{i=0}^{N-1} f(i, j) V_{pq}^*(\rho_{ij}, \theta_{ij}) \tag{14}$$

Also two common techniques are used for mapping the square image to unit circle as: placing circle inside the image and placing the image inside the circle [11]. The second technique, which is shown in Fig. 3, can be used to avoid information loss according to first technique due to remaining part of the image outside of the unit circle. If NxN pixel digital image is used then r, and θ in the Fig. 3 can be determined as

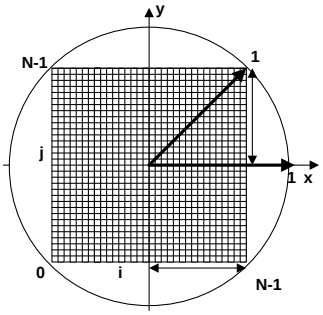

Fig. 3. Mapping the image inside the unit circle.

$$r = \sqrt{(c_1 i + c_2)^2 + (c_1 j + c_2)^2} \tag{15}$$

$$\theta = tan^{-1}\left(\frac{c_1 j + c_2}{c_1 i + c_2}\right) \quad , \quad \lambda(p,N) = \frac{2(p+1)}{\pi(N-1)^2} \tag{16}$$

where $c_1=\sqrt{2}/(N-1)$, and $c_2=-1/\sqrt{2}$.

2.2.3. Obtaining Zernike Moment Coefficients from Geometric Moments

Relationship between Zernike moment coefficients and geometric moments can be expressed as [12]

$$V_{pq} = \frac{p+1}{\pi} \sum_{k=|q|}^{p} \sum_{j=0}^{s} \sum_{m=0}^{|q|} w^m \binom{s}{j} \binom{|q|}{m} B_{p|q|k} M_{k-2j-m,2j+m} \tag{17}$$

where s=(k-|q|)/2, p-k=even numbers, and w=-j for m>0 and w=+j for m<0. Also Mk-2j-m,2j+m term denotes geometric moment. If f(x,y) function of a NxN digital image is considered as f(xi,yi), then Zernike moments can be written as [12]

$$Z_{pq} = \frac{p+1}{\pi} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} f(x_i, y_j) \int_{x_i - \frac{\Delta x_i}{2}}^{x_i + \frac{\Delta x_i}{2}} \int_{y_j - \frac{\Delta y_j}{2}}^{y_j + \frac{\Delta y_j}{2}} V_{pq}^*(x,y)\, dx\, dy \tag{18}$$

where Δxi=xi+1-xi and Δyi=yi+1-yi values are the distance between two sequential pixels. It is proposed that choosing sampling points as the midpoint of pixels decreases geometric errors of zero order approximation. If an NxN square image function is defined in the [-1 1]x[-1 1] interval, then geometric moments related with these function, M_{pq}, are expressed as

$$M_{pq} = \int_{-1}^{1} \int_{-1}^{1} x^p y^q f(x,y)\, dx\, dy \tag{19}$$

Also if the image is digital and defined only in (xi,yi) points, (19) can be translated to (20) [12]

$$M_{pq} = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} f(x_i, y_j) \int_{x_i - \frac{\Delta x_i}{2}}^{x_i + \frac{\Delta x_i}{2}} \int_{y_j - \frac{\Delta y_j}{2}}^{y_j + \frac{\Delta y_j}{2}} x^p y^q\, dx\, dy = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} f(x_i, y_j) h_p(x_i) h_q(y_j) \tag{20}$$

Here these two functions, hp(xi) and hq(yj), are independent from each other and their integrals can be defined as [12]

$$h_p(x_i) = \int_{x_i - \frac{\Delta x_i}{2}}^{x_i + \frac{\Delta x_i}{2}} x^p\, dx = \left[\frac{x^{p+1}}{p+1}\right]_{x_i - \frac{\Delta x_i}{2}}^{x_i + \frac{\Delta x_i}{2}}, h_q(y_j) = \int_{y_j - \frac{\Delta y_j}{2}}^{y_j + \frac{\Delta y_j}{2}} y^q\, dy = \left[\frac{y^{q+1}}{q+1}\right]_{y_j - \frac{\Delta y_j}{2}}^{y_j + \frac{\Delta y_j}{2}} \tag{21}$$

It can be seen from (21) that any term related with image function isn't exist in the above functions. So hp(xi), and hq(yj) can be pre calculated, and stored before. Also calculation process can be speed up. Therefore Zernike moments can be determined from geometric moments by using [13]

$$V_{pq} = \frac{p+1}{\pi} \sum_{k=|q|}^{p} \sum_{j=0}^{s} \sum_{m=0}^{|q|} w^m \binom{s}{j} \binom{|q|}{m} B_{p|q|k} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} f(x_i, y_j) h_p(x_i) h_q(y_j) \quad (22)$$

3. Experimental Study

In this study an experimental system, consisted of a parallel axis stereovision block, an image processor, a robot arm, and its controller unit, was implemented to achieve autonomous grasping by using visual data. Block scheme of the whole system is given in Fig. 4.

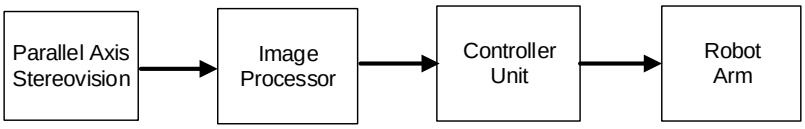

Fig. 4. Block scheme of experimental system.

Here for pattern recognition and acquiring the position of the object, Zernike moment method was used. Also Zernike moment coefficients were determined by using geometric moment method expressed as [11] and [12]. Algorithm associated with Zernike moment method implemented in DSP [14] was checked with a sample data in [12] and the results are given in Table 1. As seen from Table 1, the algorithm written in DSP gives accurate results.

Table 1. Validating data of the algorithm written for Zernike moment method.

p	q	Zpq [12]	Zpq (acquired by DSP)
0	0	5.4113	5.411268
1	1	0	0
2	0	-5.4113	-5.41127
2	2	0	0
3	1	-0.3376-1.3505i	-0.337618-1.350474i
3	3	0.3376-1.3505i	0.337619-1.350474i
4	0	-1.8038	-1.80375
4	2	0	0
4	4	-1.8038	-1.80376
5	1	0.2321+0.9285i	0.232113+0.928450i
5	3	-0.4959+1.9835i	-0.495878+1.983509i
5	5	0.2638+1.0551i	0.263764+1.055058i
6	0	1.8038	1.803745
6	2	0	0
6	4	1.8038	1.803752
6	6	0	0
7	1	0.5495+2.1980i	0.549504+2.198043i
7	3	0.1328-0.5311i	0.132763-0.531048i
7	5	-0.4669-1.8675i	-0.466861-1.867451i
7	7	-0.2154+0.8616i	-0.215408+0.861631i
8	0	1.0823	1.082241
8	2	0	0
8	4	1.0823	1.082268
8	6	0	0
8	8	1.0823	1.082255

In the experimental study three different objects, seen in the Fig. 5, were used. Also acquired video format by FPGA processor was BT.656 form [15]. In addition Y components of the video frames were processed in 128x128 pixels dimensions as in Fig. 6.

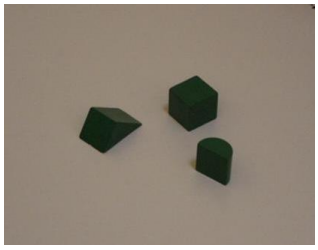

Fig. 5. The objects used in the study.

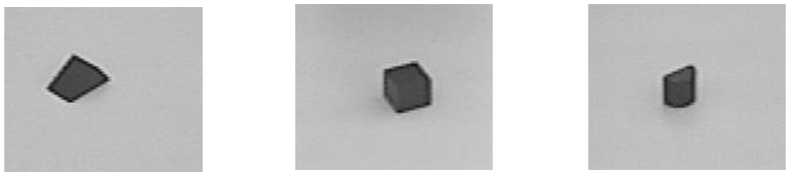

Fig. 6. Gray images of the objects.

The images of objects were grabbed as stereovision pairs. Since the cameras were located in parallel axis, right frame was seen as a shifted image of left frame in x axis (Fig. 7).

Fig. 7. Stereo images; (a) image from right cam (b) image from left cam.

Binarization process was applied to stereo image pairs to eliminate several artifacts as illumination and an example of this is shown in Fig. 8.

Fig. 8. Thresholded stereo image pair.

In this study a feature vector which was consisted of absolute values of Zernike moment coefficients up to 8th order were extracted for object recognition. However due to normalization, and translation to center Geometric μ10, μ01 ve μ00 moments were determined as 0, 0, and 1 respectively. So Z_{00} and Z_{11} coefficients were excluded from the feature vector. Feature vector was formed as;

$$X=[\ Z_{20},\ Z_{22}\ \ Z_{31},\ Z_{33},\ Z_{40},\ Z_{42},\ Z_{44},\ Z_{51},\ Z_{53},\ Z_{55},Z_{60},\ Z_{62},\ Z_{64},\ Z_{66},\ Z_{71},\ Z_{73},\ Z_{75},\ Z_{77},\ Z_{80},\ Z_{82},\ Z_{84,}\ Z_{86,}\ Z_{88}] \quad (23)$$

However from the samples it was seen that some of the coefficients were very close to 0, and these affected the ANN, used in classification negatively. With respect to this new feature vector was formed as;

$$X=[\ Z_{20},\ Z_{22},\ Z_{40},\ Z_{42},\ Z_{51},\ Z_{53},\ Z_{60},\ Z_{62},\ Z_{71},\ Z_{73},\ Z_{80},\ Z_{82},\ Z_{84}] \quad (24)$$

Training performance of ANN was shown in Fig. 9

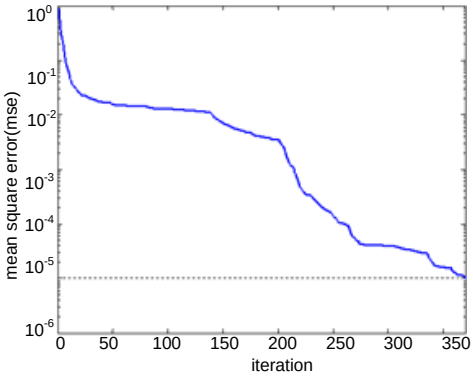

Fig. 9. Performance curve of ANN for training data.

The success rates were % 100, and % 98.33 for training, and test data respectively. Also Table 2 shows the recognition performance of the ANN for the test data of objects, used in the study.

Table 2. Confusion matrix.

Test Objects	Used Amount	Correct	False
1	40	40	0
2	40	39	1
3	40	39	1

After recognition process, finding the real coordinate points of the object problem was solved. For this issue geometric moment center point coordinates, given in (7) were used. Pixel equivalents of these geometric centers were found in each associated frame by using (25) and (26).

$$x_{cm=}64 + 64 \times \sqrt{2} \times \mu_{10} \quad (25)$$

$$x_{cm=}64 + 64 \times \sqrt{2} \times \mu_{10} \quad (26)$$

In addition these pixel coordinates were translated to metric units then distance between left camera focal point, and center point of the object in Z and, X directions were found by these pixel coordinates. Distance between reference point of the table which cameras were fixed and the object was shown in Fig. 10 and given in (27). Also the focal length of cameras was found 7.6cm experimentally. So the parameters as disparity, Z, and X were given in (28)-(30);

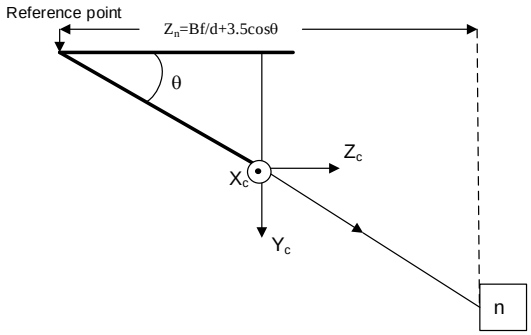

Fig. 10. Demonstration of coordinate references of the cameras fixation table and object.

$$Z\_n = Bf/d + 3.5cos\,\theta \tag{27}$$

Also the geometry of cameras in horizontal axis was shown in Fig. 11 and distances were given in (29)-(30);

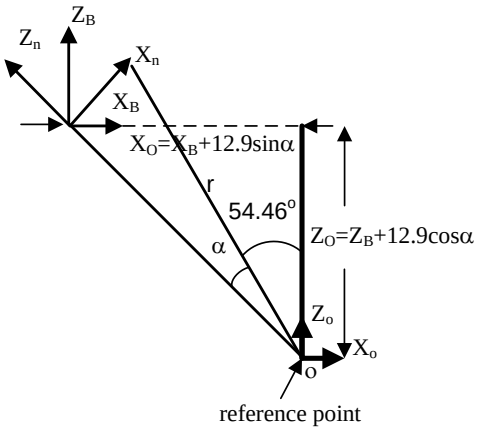

Fig. 11. Horizontal coordinate plane representation of left camera and table reference point.

$$r = \sqrt{((l_1 + l_2)^2 + l_3^2)} = 12.9cm\,, \qquad l_1 = 4cm\; l_2 = 3cm\; l_3 = 10.5cm \tag{28}$$

$$X_B = X\,cos \propto -Z\,sin \propto -r\,sin(54.46+\propto) \tag{29}$$

$$Z_B = X\,sin \propto +Z\,cos \propto +r\,cos(54.46 + \propto) \tag{30}$$

In addition coordinate transformation between the table of which cameras were fixed and robot arm was shown in Fig. 12 and the transformation matrices were given in (31)-(32) (There was a small angle difference as 6.7° between the experimental table coordinate reference and robot arm coordinate plane) ;

$$\begin{vmatrix} X_M \\ Y_M \\ Z_M \end{vmatrix} = \begin{vmatrix} -1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & -1 & 0 \end{vmatrix} \begin{vmatrix} X_N \\ Y_N \\ Z_N \end{vmatrix} + \begin{vmatrix} -23.5 \\ 94 \\ 52.6 \end{vmatrix} \tag{31}$$

$$\begin{vmatrix} X_{robot} \\ Y_{Mrobot} \\ Z_{Mrobot} \end{vmatrix} = \begin{vmatrix} cos\,6.7^o & sin\,6.7^o & 0 \\ -sin\,6.7^o & cos\,6.7^o & 0 \\ 0 & 0 & 1 \end{vmatrix} \begin{vmatrix} X_M \\ Y_M \\ Z_M \end{vmatrix} \tag{32}$$

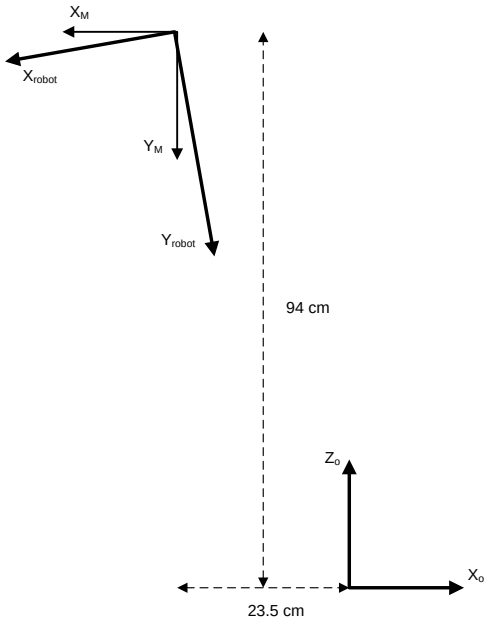

Fig. 12. Coordinate planes of robot arm and cameras representation.

After all objects were successfully by the robot arm grabbed, and an illustration was given in Fig. 13:

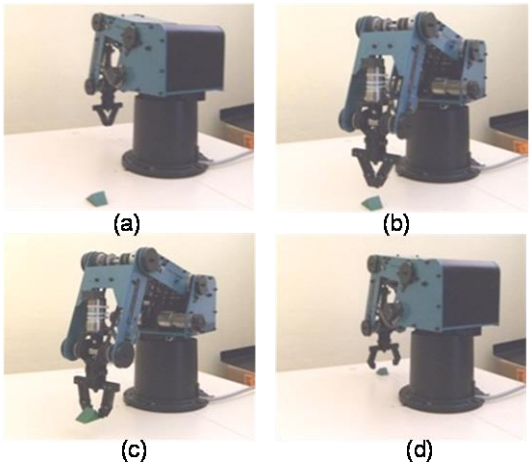

Fig. 13. Action of robot arm for grabbing objects.

4. Conclusions

This study was implemented to carry out aim of recognize, and locate certain objects by visual data and then to grasp them by a robot arm. The method used in this study is included in region based recognition so any artifact can affect the performance of the system.

Parallel stereovision cameras reduce geometric complexity and provide the difference in a direction. Also using pinhole cameras as in the study eliminates the calculations regarding to camera lens. It was seen that coefficients of Zernike moment method for each object were consistent regarding with different appearances of objects in video frame. Also performance of the recognition was very good as 98.33 %.

Program written in this study was tested with an equivalent written in MATLAB. It was seen that the computer determined it faster than DSP. However if this program would be formed in FPGA part of the embedded card maybe it would operates faster. Because calculation done by FPGA could be parallel.

Acknowledgement

This study was partially supported by Turkey Scientific Research Office (TUBITAK) with the 107E170 project.

References

[1] A.D. Kulkarni, Computer Vision and Fuzzy-Neural Systems, Prentice Hall, 2001, pp. 509.

[2] K. Huebner, BADGr—A toolbox for box-based approximation, decomposition and GRasping, Robotics and Autonomous Systs. 60, (2012), 3, pp. 367–376.

[3] Z. Iscan, Z. Dokur , T. Ölmez, Tumor detection by using Zernike moments on segmented magnetic resonance brain images, Expert Syst. Appl., 37, (2010), 3, pp. 2540–2549.

[4] S.M. Lajevardi, Z. M. Hussain, Higher order orthogonal moments for invariant facial expression recognition, Digital Signal Process. 20, (2010), 6, pp. 1771-1779.

[5] W.L.D. Lui, R. Jarvis, Eye-Full Tower: A GPU-based variable multibaseline omnidirectional stereovision system with automatic baseline selection for outdoor mobile robot navigation, Robotics and Autonomous Systs. 58, (2010), 6, pp. 747–761.

[6] http://www.sundance.com/docs/SMT339%20User%20Guide.pdf

[7] Intelitek, Scorbot-er 5Plus User Manuel, 1996, pp. 144.

[8] R.J. Prokop, A.P. Reeves, A survey of moment-based techniques for unoccluded object representation and recognition, CVGIP: sGraph. Models Image Process. 54 , (1992), 5, pp. 438–460.

[9] M.R. Teague, Image analysis via the general theory of moments, J. Opt. Soc. Am. 70, (1980), 8, pp. 920–930.

[10] M.K. Hu, 1962, Visual pattern recognition by moment invariants, IRE Trans. Inf. Theory 8, (1962), 2, pp. 179–187.

[11] C.Y. Wee, R. Paramesran, R. Mukundan, A comparative analysis of algorithms for fast computation of Zernike moments, Pattern Recognit. 36 (3) (2003) 731–742.

[12] C.Y. Wee, R. Paramesran, On the computational aspects of Zernike moments, Image Vision Comput. 25 (6) (2007) 967–980.

[13] K.M. Hosny, Fast computation of accurate Zernike moments, J. Real-Time Image Proc. 3 (2) (2008) 97–107.

[15] M.A. Arserim, Object recognition and robot arm control by intelligent methods, Ph.D. Thesis, Department of Electrical and Electronics Engineering, University of Firat Turkey, 2009

[16] www.intersil.com/data/an/an9728.pdf