



## Humanoid Robot Learning How To Track and Grip

El Sayed M. Saad<sup>1</sup>, Medhat H. Awadalla<sup>1,2</sup>, Hosam Eldin I. Ali<sup>1</sup>, Rasha F. A. Mostafa<sup>1</sup>

<sup>1</sup> Department of Communications, Electronics and Computers,  
Helwan University, faculty of Engineering  
Cairo, Egypt

<sup>2</sup> Department of Electrical and computer engineering, SQU, Oman

### ABSTRACT

Interaction with its environment is a key requisite for designing a humanoid robot especially to have the ability to recognize and manipulate unknown objects and it is crucial to successfully work in natural environments. However visual object recognition still remains a challenging problem. To get the robot capable of identifying the geometric shapes and colors of the objects, this paper proposes new approach using neuro Zernike moments. Furthermore, the paper proposes a natural language understanding system, where the robot will be able to effectively communicate with human through a dialogue developed in Arabic language. The developed dialogue and a dynamic object model are used for learning semantic categories, object descriptions, and new words acquisition for object learning. In this paper, a robot will be developed to interact with the users performing some specified actions. Moreover, integration between the proposed vision and natural language understanding systems has been presented. Finally, a hardware circuit is designed and Q-learning technique is presented assisting the robot to track and grip objects. Intensive experiments have been conducted indoor to address the validity of the complete system. Qualitative comparison among different techniques is accomplished. The achieved results show that the overall system performance of the proposed system outperforms in terms of accuracy and response time.

### Indexing terms/Keywords

Zernike moments, neural network, natural language understanding system, Gripping system and Reinforcement Learning



## Council for Innovative Research

Peer Review Research Publishing System

Journal: INTERNATIONAL JOURNAL OF COMPUTERS & TECHNOLOGY

Vol 11 No 7

[editor@cirworld.com](mailto:editor@cirworld.com)

[www.cirworld.com](http://www.cirworld.com), [member.cirworld.com](http://member.cirworld.com)

## INTRODUCTION

Autonomous robots can perform desired tasks in unstructured environments without continuous human guidance. Many kinds of robots have some degree of autonomy. Different robots can be autonomous in different ways. A high degree of autonomy is particularly desirable in fields such as space exploration, cleaning floors, mowing lawns, waste water treatment, localization and recognition of objects. Since the proper system is the reliable vision system, so the question arises here is: how to develop a robot that can see like a human? For many applications in robot vision interested in detecting objects by giving it a distinctive color from the surrounding environment, for an instance, recognizing the ball in pitch between two teams of humanoid soccer robot team [1, 2], using laser, sonar, camera or learning to classify objects into categories in human development. Such ability is crucial for robots that should operate in human environments where object categorization skills are required to recognize complex object categories (e.g., metal objects, empty bottles, etc.) [3], or other approaches used for grouping objects with the same color and shapes [4] or separating objects of different shapes [5]. However, in this paper, robots will learn how to distinguish among different geometric shapes, rectangular, circle, and triangular objects picked up via a camera mounted on the robot arm and identify their colors using different techniques. However in this paper, robots will learn how to distinguish among different geometric shapes of square, rectangular, circle, and triangular objects picked up via a camera mounted on the robot and also identify their colors.

The paper also focuses on the distinction of robot to a command given by user in Arabic language. Using the Arabic language syntax for imperative sentence and establishment of dialogue to identify the objects that not exist in the database. In this paper, we address learning of unknown objects in dialogue, which enables a robot to acquire information about unknown objects, and store this information in a knowledge base. A typical problem will be raised is that non-trivial information must be communicated, such as when the user enters an imperative syntax error, or there are new words in the written sentence that cannot be understood by the system. Thus, the dialogue system should conduct dialogue strategies for learning in such way that the information about the object can successfully be communicated. In addition, it has to cope with new words learning on writing, grammatical and semantic levels to achieve the learning goal. It should create a model of the object's semantics, which describes the type, color, shape, properties of the object and its function. All previous data will be addressed using Arabic language. Moreover, this paper adds question syntax of Arabic language.

For controlling and enhancing the motion of humanoid robot to autonomously grasp objects, a new hardware circuit is designed and Q-learning technique is presented assisting the robot to find, change its state and grasp objects.

The remainder of this paper is organized as follows: Section 2 gives an overview of the proposed system comprising the proposed vision system architecture, the natural language understanding system architecture, the integration of them and shows an overview of the grasping system. Section 3 presents experiments and discussion. Finally section 4 concludes the paper.

## 2. The Proposed System:-

The proposed system, as shown in figure (1), consists of the following sub-systems:

- A. Vision Sub-system.
- B. Natural Language Understanding Sub-system.
- C. User-Interface Dialogue Manager.
- D. Arabic Text to Speech Sub-system.
- E. Motion Sub-System (Learning How to track and grip).

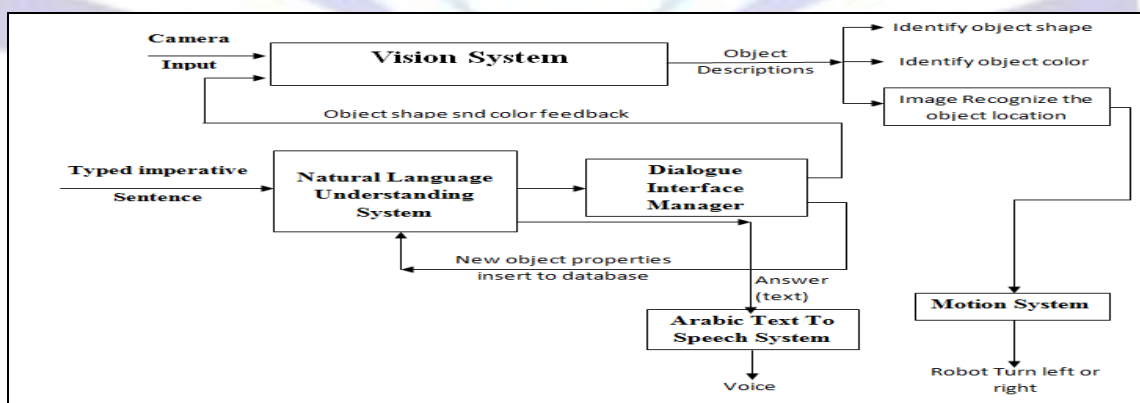


Fig.1. System Overview

### 2. A. Vision Sub-System:

There are things that attract any child such as colors and geometric shapes of objects. Once the child taught how to distinguish between them using one word that defines each shape and color then he could recognize them by himself later on. The main aim of the proposed system is to make robots behave as child, once it is learned the skill of how to

recognize object's shape such as square, rectangle, circle, or triangle, and its color such as red, black, white, blue, green,...etc., seen by its camera. It could detect the object's shape and its color by itself seen later in any image taken by his camera at different places. The proposed vision system has examined many approaches such as hough transform, harris corner detection and neural network but this paper provide more accurate and new technique using Zernike moments in new way. The following procedure is demonstrated in the flowchart shown in figure 2.

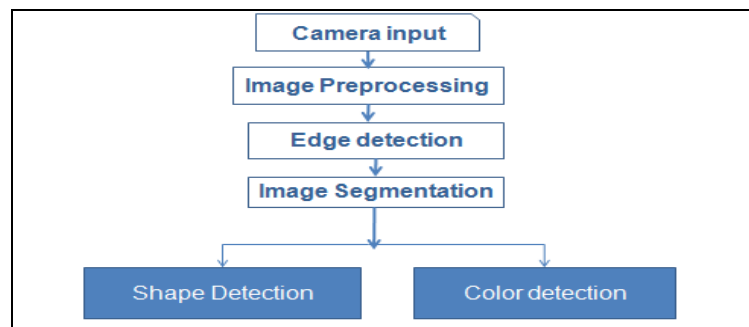


Fig.2.The object and color detection program flow chart

### 1- Acquiring image

Images taken from the robot's camera can easily be fed to our program using developed function. This function makes it possible to assign a variable as a video input. Image processing cannot be performed on a video input, so single frames should be extracted from the video with a frame grabber. A so-called snapshot is taken out of the video input and this single image is used for object shape and color detection.

### 2-Image preprocessing

The images of the object from webcam that used in our experiments have different qualities such as brightness, contrast, or noise. We overcome these problems as follow:-

**a. Noise**, image frame taken from camera is a color image. Images are coded using RGB-space. In the RGB color space, each color is described as a combination of three main colors, namely Red, Green, and Blue. This color space can be visualized as a 3d matrix. Each image is converted into black and white as shown in figure (3.a, 3.b), then the image is filtered to remove any added noise due to lighting using Median filtering to reduce "salt and pepper" noise.



Fig.3.Result of converting image to black and white

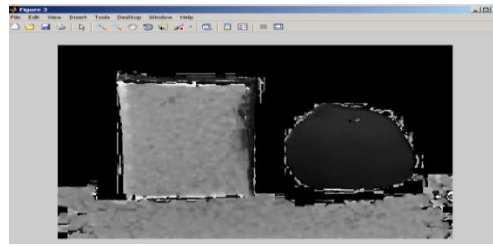
**b. Brightness**, to adjust the brightness of the image, the intensity operator with the following methods is used. The original image matrix X of N x M is the array of fuzzy singleton of each member represented by the degree of brightness levels  $p, p=0,1,2,3...P-1$  with p values between 0 and 255 can be written as follows[6].

$$X = \begin{pmatrix} \mu_{11}/x_{11} & \mu_{12}/x_{12} & \dots & \mu_{1M}/x_{1M} \\ \mu_{21}/x_{21} & \mu_{22}/x_{22} & \dots & \mu_{2M}/x_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{N1}/x_{N1} & \mu_{N2}/x_{N2} & \dots & \mu_{NM}/x_{NM} \end{pmatrix}$$

Where  $0 \leq \mu_{mn} \leq 1$  is intensity,  $m=1,2,...M, n=1,2,...N$  and  $\mu_{11}/x_{11}$  means intensity value of first pixel  $x_{11}$  of matrix X.

$$\begin{aligned} T_1(\mu_{mn}) &= T_1'(\mu_{mn}) = 2\mu_{mn}^2 & 0 \leq \mu_{mn} \leq 0.5 \\ &= T_1^*(\mu_{mn}) = 1 - 2(1 - \mu_{mn})^2 & 0.5 \leq \mu_{mn} \leq 1 \end{aligned} \quad (1)$$

Equation (1) is used to adjust each pixel of image brightness as follows, the values greater than 0.5 increases the brightness of the pixels, the values less than 0.5 decreases the brightness of the pixels. The procedure of the fuzzy image enhancement is shown in Figure 4.



a. Light affection on an image

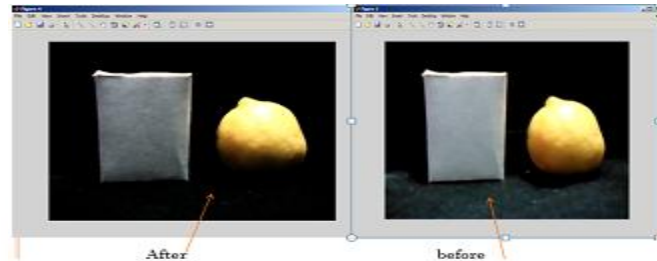


Fig.4.Reduction of the brightness on an image

### 3- Edge detection

The black and white image is converted into edge image, we developed function takes a binary image  $I$  as its input, and returns a binary image  $BW$  of the same size as  $I$ , with 1's where the function finds edges in  $I$  and 0's elsewhere. Different methods has been examined and chose the best one as seen in table 2 the best approach for edge detection in our application using edge based color Method which will be illustrated as follow:-

#### Edge Detection based on Color Method

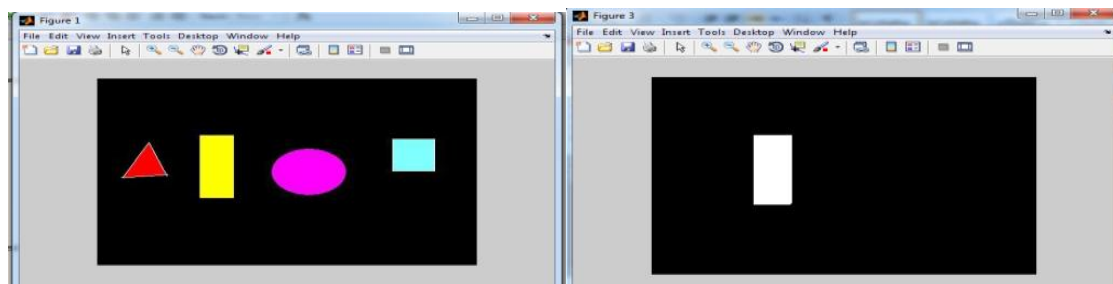
As viewed previously in flowchart in figure 2, color of object is detected after separating each object alone and found its color using RGB system. Identification the color of the object can be achieved through the borders of each object, separating the object into an individual image, and then the color can be determined by creating a color map function, where each color can be visualized as a 3d matrix. Finding the average of this matrix can identify the color of the object. This method has low efficiency. So modification made to identifying the object's color achieved more powerful to propose the use of the color to detect the object location and boundary using HSV system to save the fuzzification preprocessing time which more accurate and powerful system. Fuzzification stages that detect the object edge using fuzzy sobel [6] takes long time. This time will be a challenging problem when robot arm get learned to grasp object. Robot will take a long time to take action according to the object's location. To tackle this problem, a new technique is proposed. This technique does not begin with shape detection then detect its color after using fuzzy sobel edge, however it detects object location and its edge by its color then it recognizes its shape using the following techniques. This algorithm is as follows:-

*Step1:* Color image taken by camera.

*Step2:* Convert color system from rgb system to hsv system.

*Step3:* Develop a function to detect object location by its color.

For example, user needs to find object with rectangle shape and yellow color, the developed system will search for all objects with yellow color using color hsv ranges in table 1, and extracts them in new black and white images which will then be used to find the rectangle image using one of the following techniques as shown in figure 5. So we save fuzzy sobel edge detection time needed for preprocessed image.



(a)

(b)

Fig.5. Example of an image taken and yellow object is extracted in a separate image

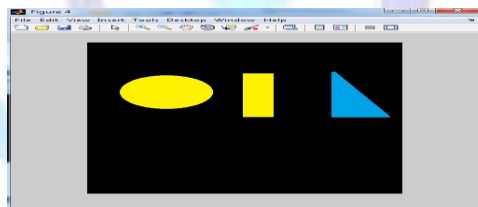
**Table 1. HSV colors ranges**

	H(Hue) Range	S(Saturation) Range	V(Value) Range
<b>Black</b>			$\geq 0 \& \leq 0.19$
<b>White</b>		$\geq 0 \& \leq 0.049$	
<b>Yellow</b>	$\geq 0.084 \& \leq 0.25$		
<b>Green</b>	$> 0.25 \& \leq 0.4167$		
<b>Cyan</b>	$> 0.4167 \& \leq 0.5833$		
<b>Blue</b>	$> 0.5833 \& \leq 0.75$		
<b>Magnetic</b>	$> 0.75 \& \leq 0.9167$		
<b>Red</b>	$> 0.083 \& \leq 0.9168$		

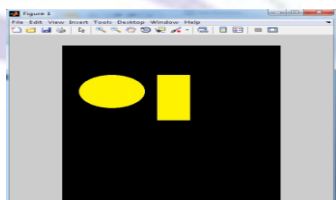
The proposed edge detection method based on fuzzification preprocessing is compared with object extraction based on color to extract object and the edge detection based on the ordinary sobel, shown in table 2. Noted that, the proposed object detection based color is fastest and there is tiny change in time with increasing number of objects in an image. From the previous figure, it is noted that, the object with required color was detected without the need for shape detection. However, shape detection will be needed in case of there are more than one object with the same color as seen in figure 6 So after detected object based its color, the following shape detection algorithm used.

**Table.2. comparison among fuzzy sobel, sobel, and object extraction based on color**

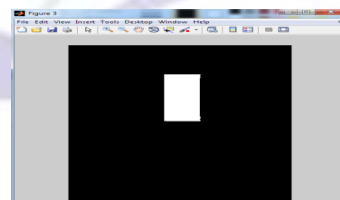
	Object edge detection based on fuzzy sobel	Object edge detection based on sobel	Object edge detection based on color
Time for one object per image	48.8273 sec	0.1049 sec	0.1090 sec
Time for four object per image	72.5782 sec	0.2035 sec	0.1099 sec



(a) Original image



(b) After edge based yellow color



(c) After used shape detected rectangle shape

**Fig.6. Example show shape detection important**

#### 4- Image Segmentation

The previous binary image is used to find the boundary of each object. We developed a function traces the exterior boundaries of objects, as well as boundaries of holes inside these objects, in the binary image BW. Also descends into the outermost objects (parents) and traces their children (objects completely enclosed by the parents). BW must be a binary image where nonzero pixels belong to an object and 0 pixels constitute the background returning indices of each object boundary. These indices are used to segment the black and white image to a set of images, each of which contains only one item.

## 5-Neuro Zernike Moments

Neural network technique mentioned in [6] provided highest percentage compared to all previous techniques such as neuro hough transform, neuro harris corner detection, and neuro wavelet transform as will be seen in the following experimental result. To enhance vision system to detect the shape of the object regardless of its rotation, we trained neural network with Zernike moments vector of the object shape. In general, Zernike moments are often used efficiently as shape descriptors of image objects, such as logos or trademarks that cannot be defined by a single contour. Shape identification and feature extraction are the main concern of our vision system. Object parameters are mostly dependent on spatio-temporal relationships among the pixels. However feature extraction is a complex phenomenon which needs to be addressed from the invariance property, irrespective of position and orientation. Zernike moments are used as shape descriptors and identified as rotation invariant due to Orthogonality property.

To compute the Zernike moments of a digital image, the range of the image should be mapped to the unit circle first with its origin at the image's center. The pixels falling outside the unit circle are discarded in the computation process. In our implementation of Zernike moments, we used binary images. So the following steps are necessary to extract features of any shape image using Neuro Zernike moments:-

- (a) Converts color image into the binary image.
- (b) Map over a unit disc image and convert into polar coordinate as seen in figure 7.a,b.The center of the image and disk is same

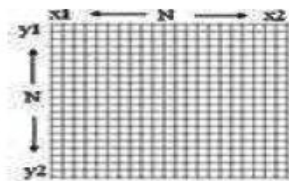


Fig .7( a) NxN pixels size image.

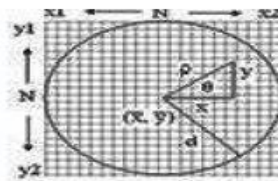


Fig.7.(b) Unit Circle Mapped image bitmap onto NxN pixel

Where  $x_1, x_2$  are X-axis dimensions and  $y_1, y_2$  are Y-axis dimensions of the pixel rectangle.  $\bar{x}, \bar{y}$  is the center of the unit disk,  $\rho$  is polar value and  $\theta$  is polar angle.

- (c) Compute the distance (d) using equation (2)

$$d = \sqrt{(x^2 - \bar{x})^2 + (y^2 - \bar{y})^2} \quad (2)$$

- (d) Compute the distance vector  $\rho$  and angle  $\theta$  for any (x,y) pixel in  $f(x,y)$  in polar coordinates as

$$\rho = \sqrt{(x - \bar{x})^2 + (y - \bar{y})^2} / d$$

and

$$\theta = \tan^{-1} \left[ \frac{x - \bar{x}}{y - \bar{y}} \right]$$

This step maps pixel coordinate  $(x_1, x_2)$  to  $(-1, +1)$  and  $(y_1, y_2)$  to  $(-1, +1)$  in polar. In this way, almost all the pixels in image bound box as given in figure 3.21 above are inside unit circle except some foreground pixels.

- (e) Compute a set of complex polynomials  $\{V_{nm}(x, y)\}$  which form a complete orthogonal set over the unit disk of  $x^2 + y^2 \leq 1$  in polar coordinates and Zernike moments using equations 3,4,5. Moreover, reconstruction of Zernike moments can be used to determine the amount of moments necessary to make an accurate descriptor using equation 6.

$$Z_{nm} = ((n + 1)/\pi) \sum_x \sum_y V^*(\rho, \theta) f(x, y) \quad x^2 + y^2 \leq 1 \quad (3)$$

Where:

$|m| \leq n, n - |m|$  is even.

The Zernike polynomial  $V_{nm}(\rho, \theta)$  is defined as a function of the radial polynomial:

$$V_{nm}(\rho, \theta) = R_{nm}(\rho) e^{im\theta} \quad (4)$$

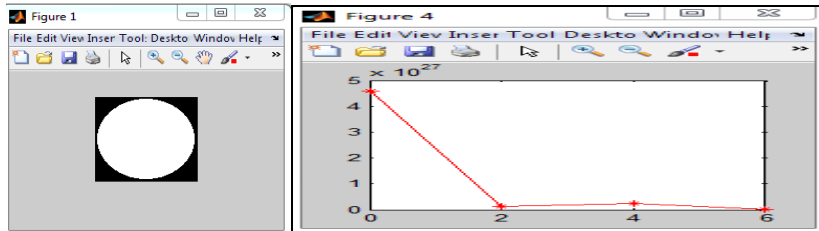
And the radial polynomial is:

$$R_{nm}(\rho) = \sum_{s=0}^{\frac{n-|m|}{2}} \frac{\rho^{n-2s} (-1)^s (n-s)!}{s! \left(\frac{n-2s+|m|}{2}\right)! \left(\frac{n-2s-|m|}{2}\right)!} \quad (5)$$

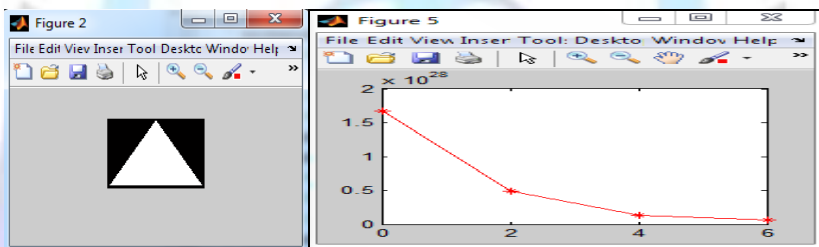
Suppose we know all Zernike moments  $Z_{nm}$  of  $f(x, y)$  up to order  $N$ , we can reconstruct the image by:

$$\hat{f}(x,y) = \sum_{n=0}^N \sum_m Z_{nm} V_{nm}(x,y) \quad (6)$$

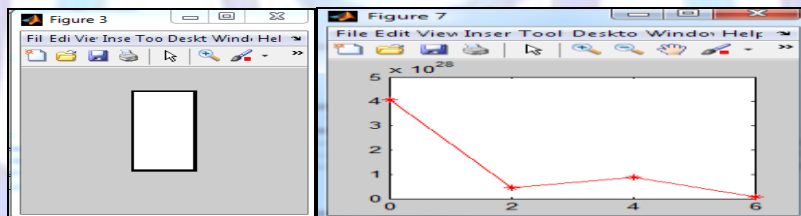
(f) Choosing order of  $(n=8,10,120,140)$  each with four different repetitions  $(m=0,2,4,6)$ . Order of Zernike is choosing according to how could differentiate between the three shapes. Find out that, calculating Zernike moments for fixed order with different repetitions and plotting Zernike moments results producing different curves for each shape, for example rectangle has different curve than circle or triangle as seen in figure 8. But when calculating Zernike moments for fixed repetition with different orders producing the same curves for the three shapes which couldn't allow us to differentiate between the three shapes as seen in figure 9.



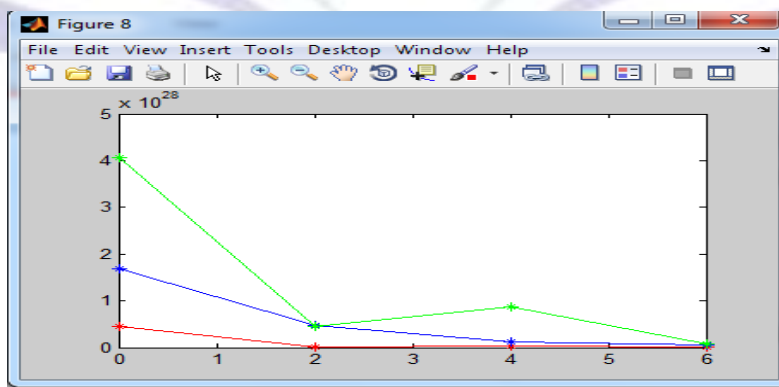
(a) Magnitude of Zernike moments for circle



(b) Magnitude of Zernike moments for triangle

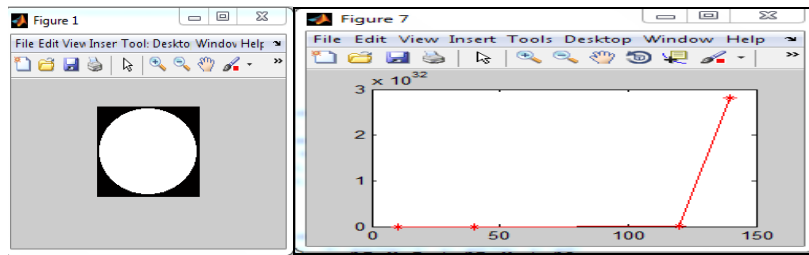


(c) Magnitude of Zernike moments for rectangle

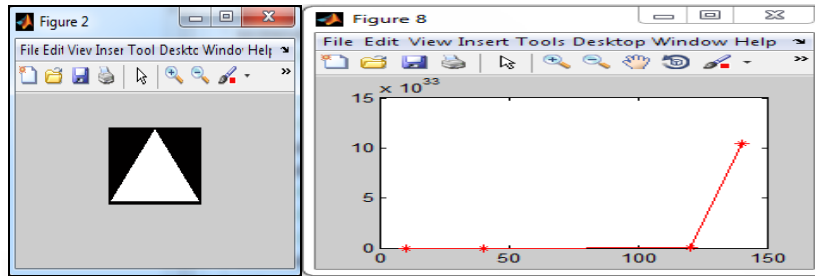


(d) Zernike moments result of three shapes, red for circle, blue for triangle and green for rectangle

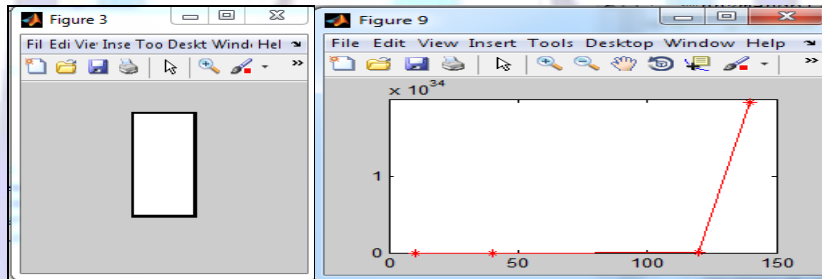
**Fig.8 Zernike moments results of  $n=120$  with  $m=0,2,4,6$**



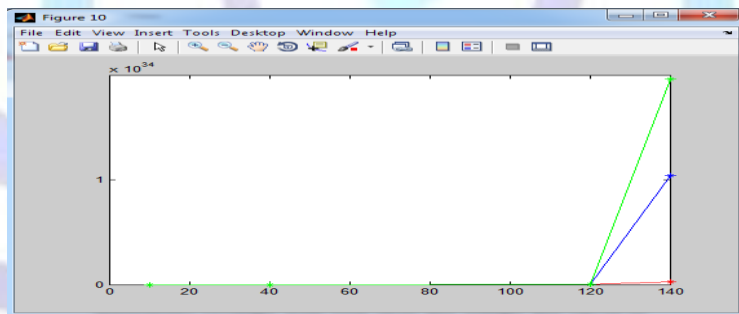
(a) Magnitude of Zernike moments for circle



(b) Magnitude of Zernike moments for triangle



(c) Magnitude of Zernike moments for rectangle



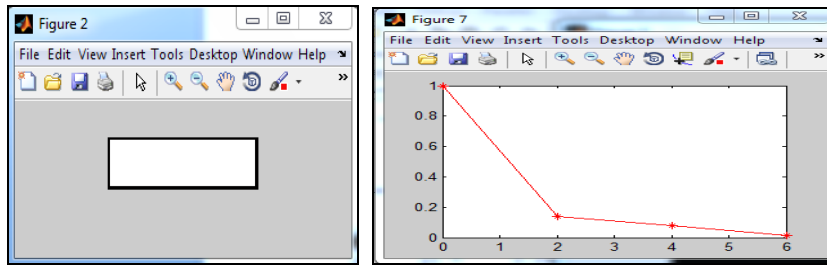
(d) Zernike moments result of three shapes, red for circle, blue for triangle and green for rectangle

**Fig. 9. Zernike moments results of  $n=10,40,120,140$  with  $m=2$**

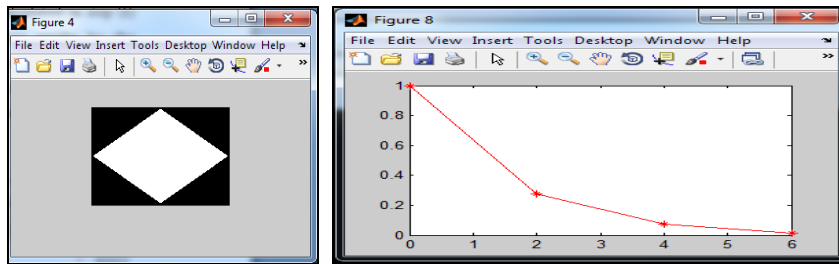
It is noted that, curves of Zernike moments result with fixed order and changing repetition differentiate rectangle from both circle and triangle at high order and curves of Zernike moments with fixed order and changing repetition differentiate circle from both triangle and rectangle at low order. So choosing both higher order and lower one allow to differentiate between shapes. But curves of Zernike moments results with fixed repetition and change order produced the identical curves with all shapes so we couldn't use them to identify the shape of object.

(g) Rescale Zernike moments calculated in step (f) by dividing all Zernike moment results by the maximum one which make all values between zero and one. So when we use different scales for the same shape will produce the same curve as seen in figure 10.

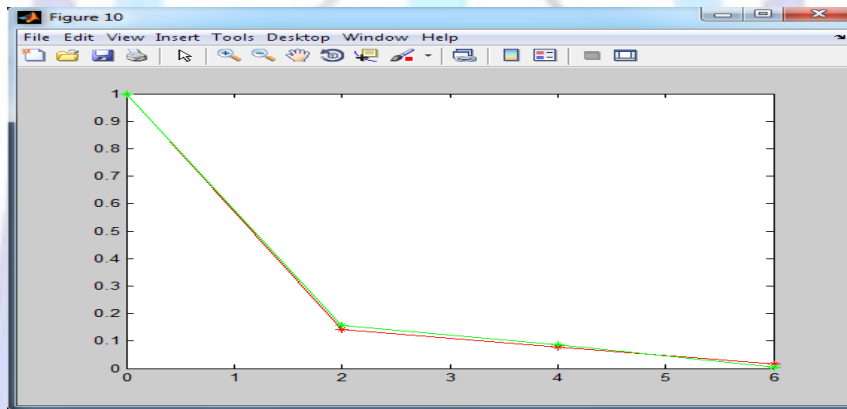




(a)



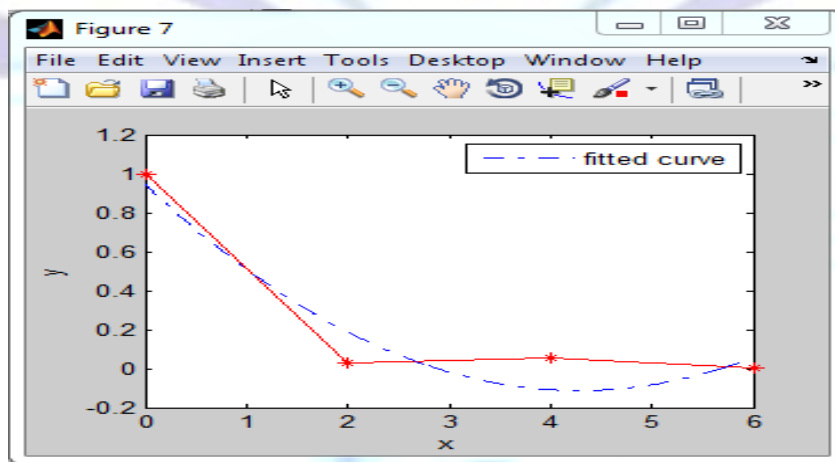
(b)



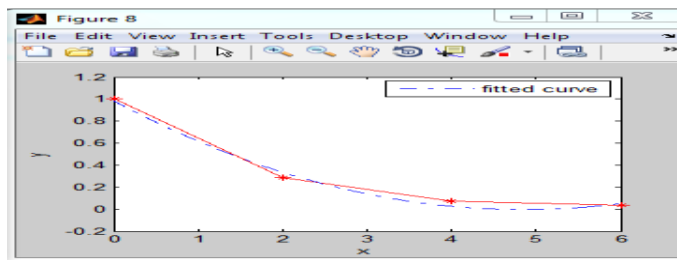
**Fig. 10. Rescale Zernike moments results of the same shape with different scale**

It is noted that, the magnitude of Zernike moments results were rotation independent.

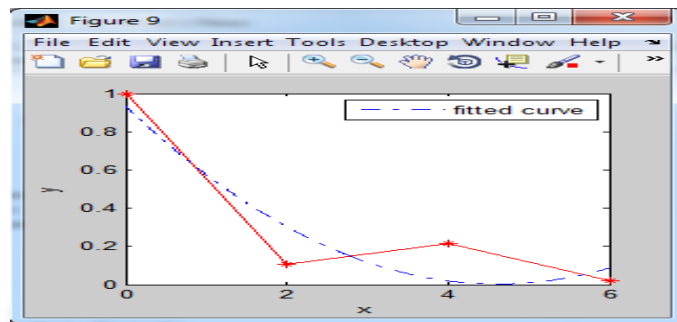
- (h) Calculate four polynomial functions that represent curves for magnitude of Zernike moments results calculated as seen in figure 11 with dashed lines represents polynomial function. And four polynomial functions that represent curves for phase of Zernike moments results.



(a) Polynomial function represents magnitude of zernike moments for circle shape



(b) Polynomial function represents magnitude of zernike moments for triangle shape



(c) Polynomial function represents magnitude of zernike moments for rectangle shape

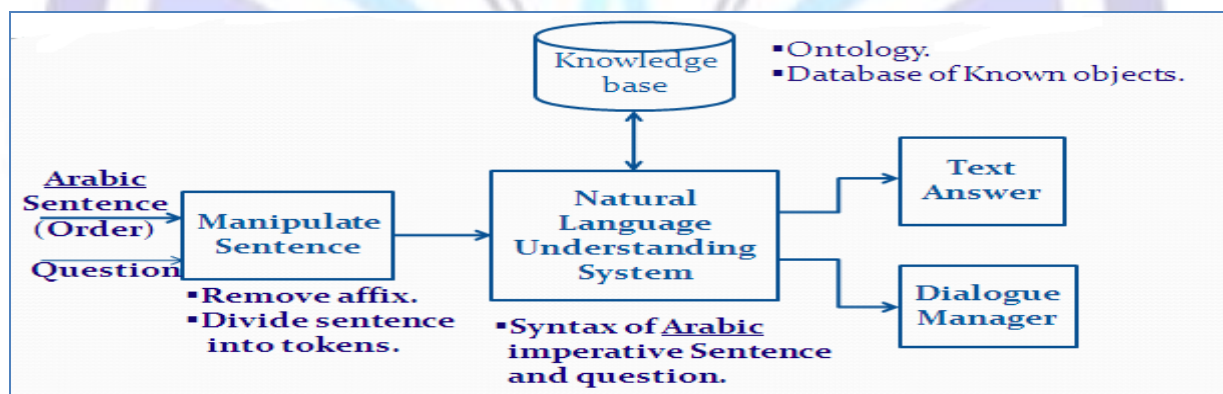
**Fig.11. polynomial function represents magnitude of zernike moments for three shapes**

(i) Calculating three coefficients for each polynomial function has been calculated in step (h).

The previous polynomials' coefficients vector has been used to train the neural network. This solves our input size problem of neural network. The problem of varying coefficients vectors size which used in the previous techniques mentioned in [6] reduced to 24 neurons and is not depend on the size of captured image or shape rotation as will see in the experimental results.

## 2. B. Natural Language Understanding System Sub-system

An interactive learning for artificial systems has been addressed in several systems. However, the number of approaches that allow interactive knowledge acquisition for humanoid robots is still comparably small [7]. This paper focuses on how to establish a dialogue between the user and the robot especially if some of the commands to the robots are not pre-defined. Furthermore, the paper concentrates on how the robot will understand the commands on syntax of the Arabic language, to address these issues, the natural language understanding system is proposed as shown in figure 12. The proposed system has the following components:



**Fig.12. Natural Language Understanding System Overview**

a- The form of imperative sentence syntax

First sentence is entered into the system in the form of imperative syntax; imperative syntax of the Arabic language takes more than one form as shown in figure 13.a. The input sentence is divided into a set of tokens, and then the affix such as "ل" added at the beginning of the word or "ى" added at the end of the word is omitted. Moreover, the proposed system has ability to take a question from user about the place of the object using a question tool of "أين". The question is entered into the system in the form of question syntax; the question syntax of the Arabic language takes more than one form as in the shown in figure 13.b, it begins with "أين".



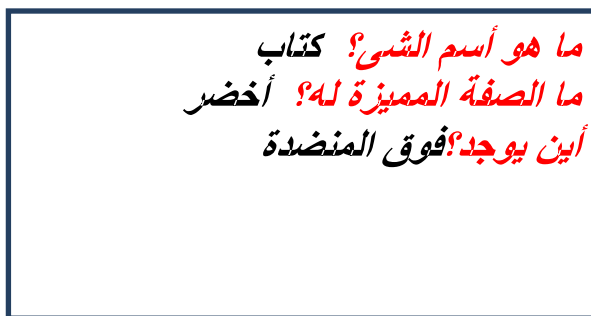


Fig.15. Dialogue sample

## 2. C. User-Interface Dialogue Manager:-

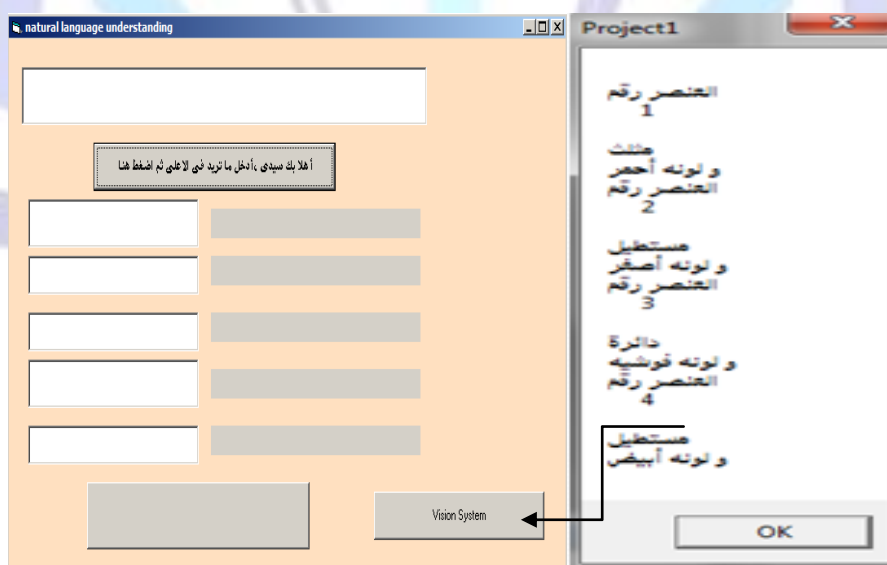
Merging between the vision and the natural language understanding systems is implemented by regular usage of GUI using Visual BASIC 6.0, where it appears to the user interface screen. One of the major problems faced most of previous researchers is the linking between Matlab environment and Visual Basic to have a reasonable interface for the users. In this paper, all programs have been developed in Prolog, and the Dynamic Link Library is implemented in such way that it can deal directly with Visual Basic programs. The developed interface gives the user different capabilities to choose as follows:

1- The user can ask the robot to describe the scene in the front of itself in terms of the shapes of the objects and their colors by using the developed vision program just by pressing a key called the vision system in the developed interface shown in figure16.a.

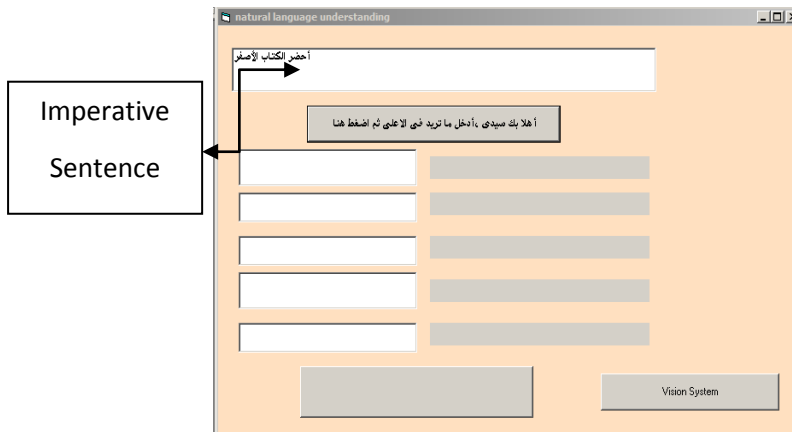
2-The user can ask the robot to identify or fetch a particular object, the robot will use the developed Natural Language Understanding System to search about the object in its database. If it is recognized, the robot writes to the user that object is found. Then the object features will be sent to the vision system to recognize its shape and Color, as shown in figures 16.b and 16.c.

3-The user can ask the robot about the object's place, the robot will use the developed Natural Language Understanding System to search about the object in its database. If it is recognized, the robot writes to the user that the object place as shown in figures 16.e and 16.f. Then the object will be sent to the vision system to recognize its shape and Color.

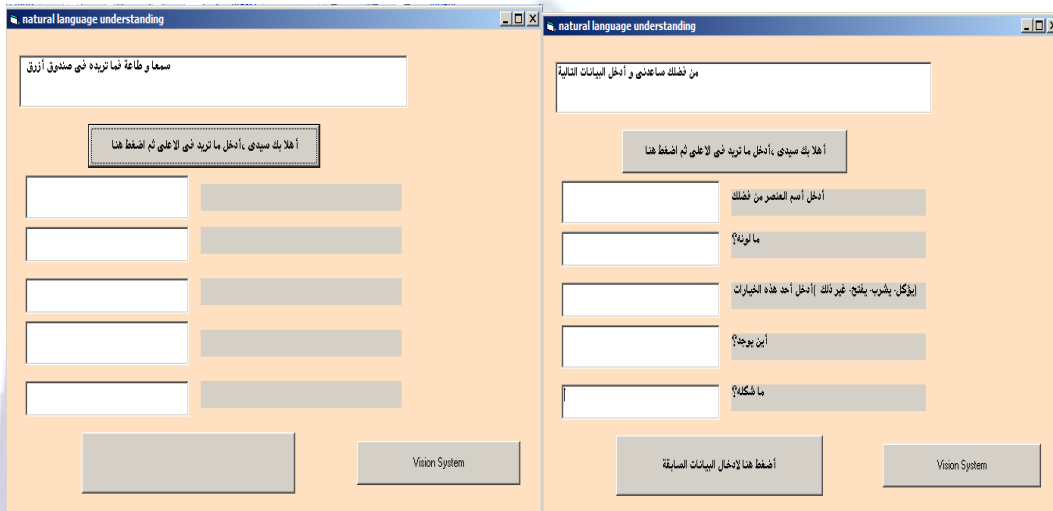
In all the above cases, if the robot could not recognize the object because it is not in its database or the user entered a syntax error, or, robot does not find object's place in its database, a dialogue between the user and the robot will start, the user will answer some questions as shown in figure 16.d, and then the database will be adapted to accommodate the new information. If the same command sent to the robot, or the robot is asked again for identifying the same object, the system will be able to identify it.



(a) Result from calling vision system

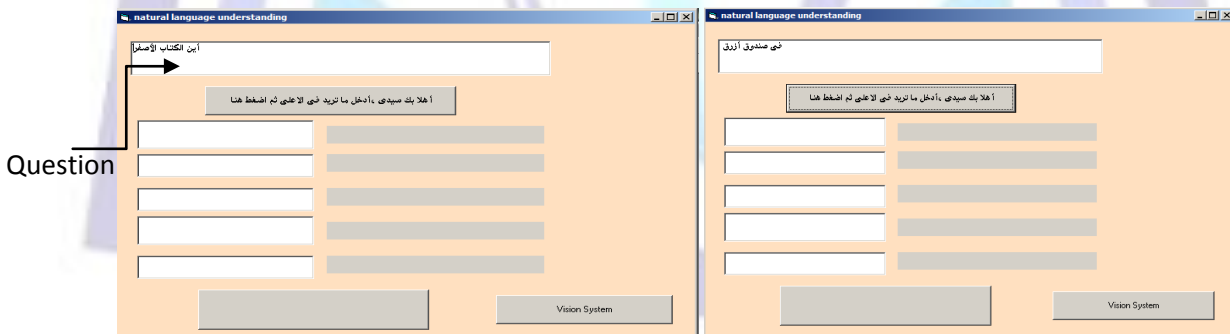


(b)



(c) Result when the object founded

(d) result when the object not founded

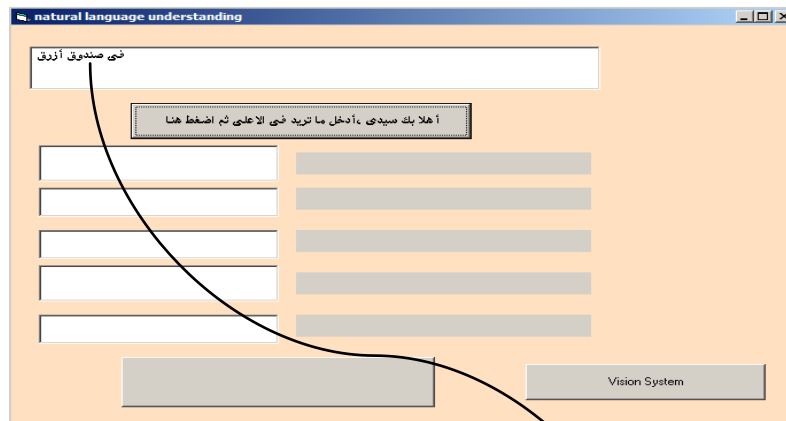


(e)The user ask the robot about the object's place (f) the robot writes to the user the object place

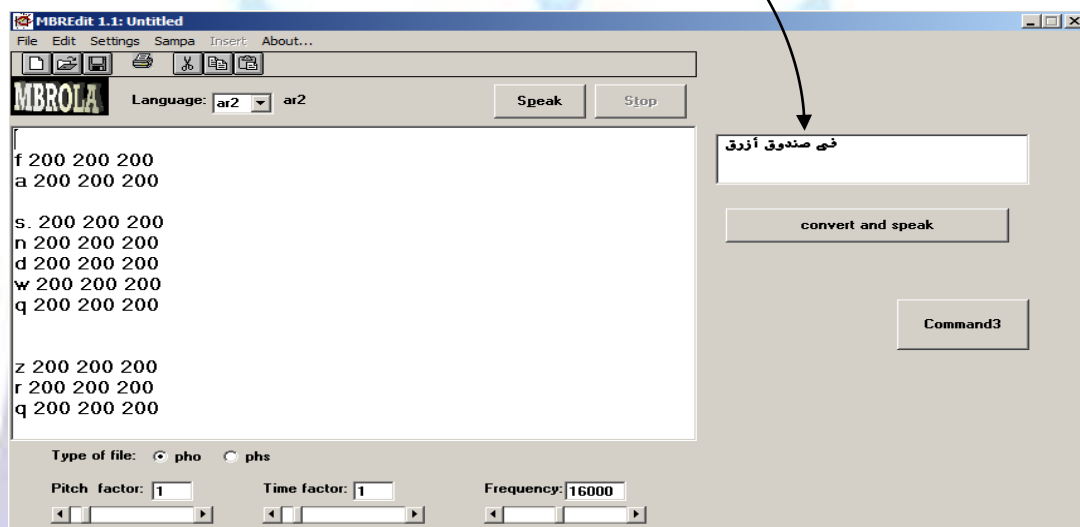
**Fig.16. Example of the proposed system scenario**

## 2. D. Text to speech Sub system (Talking System)

The previously mentioned subsystems are interdependent. It is not enough to equip the robot with basic functionalities for dialogue comprehension and production to make it interact naturally in situated dialogues. We also need to find meaningful ways to relate language to be spoken, and enable the robot to use its perceptual experience to continuously learn and adapt itself to the environment. And cover speech recognition, where the dialogue between the user and robot will be voice based. For this purpose we used a tool to convert Arabic text of robot answer to voice, this tool called MbrolaTools35. This tool is well described in [13]. We used it to convert the answer of robot appeared to user to speech heard by user. When answer is written to user it will hear at the same time, this by pass this text to MbrolaTools35 as seen in Figure (17.a, 17.b), however the screen appears in figure 18.b will be invisible for the user. User will hear the voice immediately when it appears at text box. Also the dialogue between user and robot shown in figure 16.d changed to be not only based on text but also voice based.



(a) Text Answer appear to user



(b) Converting robot answer from text to speech

Fig.17. Example of converting text to speech mechanism

## 2. D. Motion Sub-System (Learning How to track and grip):

In highly structured settings, modern-day robots can be scripted to perform a wide variety of tasks with mind boggling precision and repeatability. However, outside of carefully controlled settings, robotic capabilities are much more limited. Indeed, the ability to even merely grasp amodest variety of previously unseen rigid objects in real world cluttered environments is considered a highly nontrivial task [14]. While Reinforcement Learning (RL) is not traditionally designed for interactive supervisory input from a human teacher, several works in both robot and software agents have adapted it for human input by letting a human trainer control the reward signal. This paper proposed a trained robot arm to grasp object specify certain shape and color using reinforcement learning such as q-learning [15]. Interactive Q-Learning modified to incorporate interactive human guidance in addition to feedback algorithm has been chosen to save time of training using ordinary q-learning algorithm. The algorithm has been used is as follow:-

- Step 1: while learning do
- Step 2: while waiting for guidance do
- Step 3: if receive human guidance message then
- Step 4:  $g = \text{guide-object}$
- Step 5: end if
- Step 6: end while
- Step 7: if received guidance then
- Step 8:  $a = \text{random selection of actions containing } g$

Step 9: else

Step 10:  $a$  = random selection weighted by  $Q[s, a]$  values

Step 11: end if

Step 12: execute  $a$ , and transition to  $s'$  (small delay to allow for human reward)

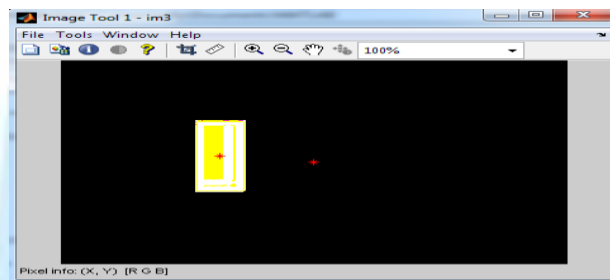
Step 13: sense reward,  $r$

Step 14: update Q-value:

$$Q[s, a] \leftarrow Q[s, a] + \alpha(r + \gamma (\max_{a'} Q[s', a']) - Q[s, a])$$

Step 15: end while

The proposed states has been used due to the difference between the object location and center of image (camera located at robot arm) as shown in figure [18], the state of yellow rectangle object is in state 1 needs the robot to turn left. Actions that robot can do are up, down, left, and right.



**Fig.18.Object in state 1 due to image center**

- The states have been chosen as follows:-
  - State 1:- if  $e_x >$  threshold .
  - State 2:- if  $e_x <$  threshold.
  - State 3:- if  $e_y >$  threshold.
  - State 4:- if  $e_y <$  threshold.
  - State 5:- if  $e_x =$  threshold and  $e_y =$  threshold.

Where  $e_x =$  difference between x-coordinate of object and x-coordinate of image center.

$e_y =$  difference between y-coordinate of object and y-coordinate of image center.

Threshold is chosen according to distance between camera mounted on robot arm and the gripper of robot arm.

- Actions have been chosen as follows:-
  - Action 1:- robot arm turn left.
  - Action 2:- robot arm turn right.
  - Action 3:- robot arm up.
  - Action 4:- robot arm down.
  - Action 5:- robot arm catch object.

- Reward ( $r$ ) is one value from ( 1 , -1 or 0 ) according to the progress of robot arm finding object.

	a1	a2	a3	a4	a5
S1	1	-1	-1	-1	0
S2	-1	1	-1	-1	0
S3	-1	-1	1	-1	0
S4	-1	-1	-1	1	0
S5	-1	-1	-1	-1	1

- $\gamma$  is learning rate has been chosen = .8 its value between [0,1].

The q-learning matrix initially zeros and then after the robot get trained from two different trainers the q-learning matrix was as follows:-

q =				
2	0	-1	0	0
0	3	0	0	0
2	0	3	2	2
0	0	0	1	0
0	0	0	0	4

q =				
5	0	1	0	-1
-1	12	0	0	0
0	0	10	0	-1
0	0	0	8	0
0	0	0	0	11

Trainer 1

Trainer 2

After robot arm learned how to find, verify its state and action for catching object, robot will be tested to follow q- learning matrix to find the appropriate action without intervention of human.The robot arm and hardware circuit connected of our system seen in figure [19].

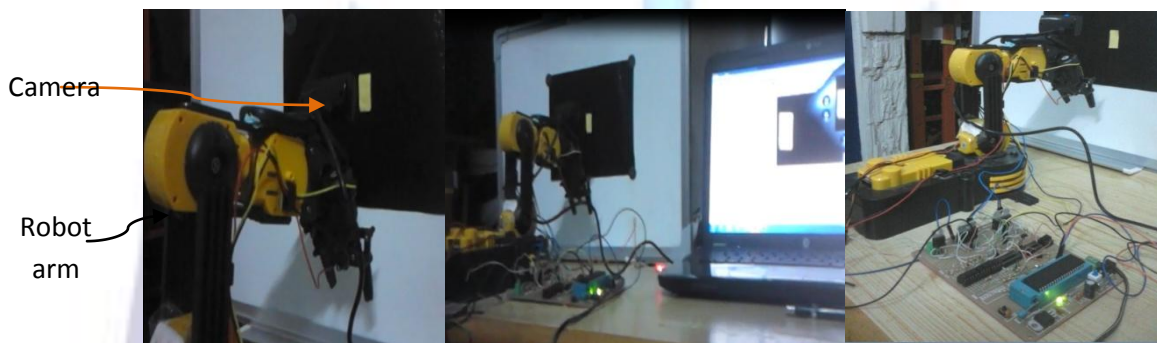


Fig.19. Robot arm and its circuit

### 3. Experiments and Discussion

Intensive experiments have been conducted to address the validity of the proposed systems. First, we have tested the vision program in several stages to check its accuracy, initially it is tested on still images, and then simulated using Webots 6.2.4 simulator. Background with one color, black, is chosen. The achieved results from the vision program have a reasonable precision, although the vision affected by the camera angle, and lighting. Furthermore, the achieved results that compared to related work in [6, 16, 17] in table (3) showed that the proposed neuro Zernike moments provided the best result and it is rotational independent and we made it scale independent also. This thesis provides techniques not only to detect the shape of the object but also enhance the color detection system in [18] which replaced the rgb system with hsv system. Comparison between the vision system techniques mentioned in [6,16,18] and the proposed system mentioned in this paper seen in figure 20, 21, when all previous techniques for color and shape detection applied in detection of 60 practical objects image, it is observed that hsv better than rgb system and is not affected by light like rgb system.

Table.3 PERCENTAGE OF DETECT THE SAME OBJECTS 100 times

Object \ Technique	Circle	Triangle	Rectangle
Hough Transform[17]	98	96	97
Neuro HT	98	97	96
Harris Corner Detection[17]	99	95	95
Neuro HCD	99	96	95
Neural Network[17]	99	99	99
Neuro WT	99	99	98
Neuro Zernike moments	100	100	100



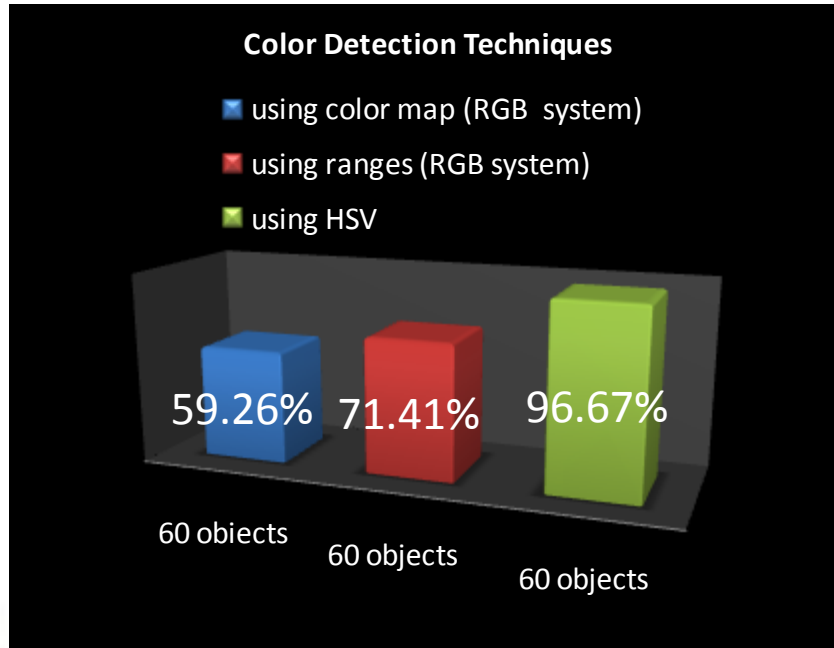


Fig.20. Color techniques comparison

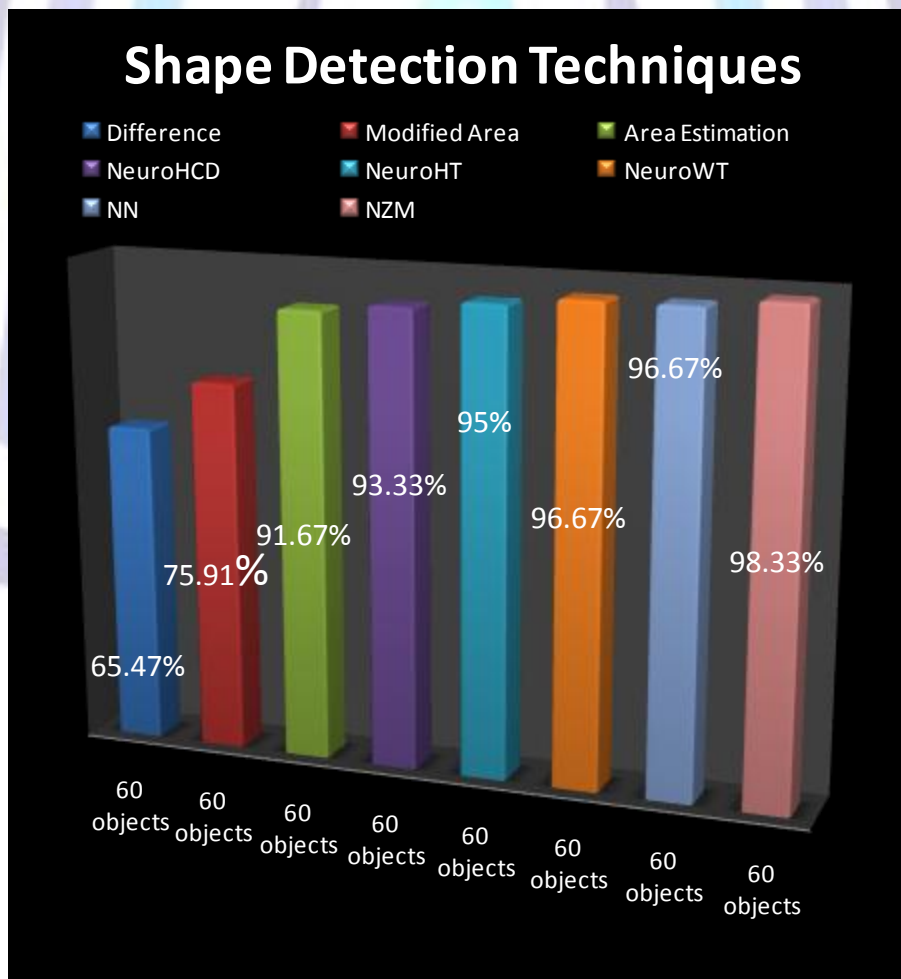


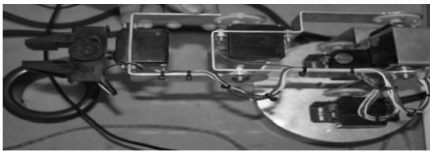
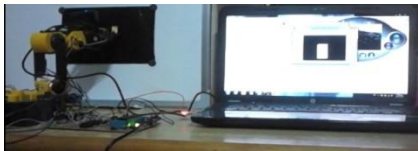
Fig.21. Shape Detection techniques comparison

#### 4. Conclusion

We concluded that to get the robot capable of identifying the geometric shapes and colors of the objects, this system examined different approaches such as , Neuro Hough Transform, Neuro Harris corner detection, Neuro Wavelet transform and Neuro Zernike moments. Hybrid technique between neural network and Zernike moments has a remarkable performance for object identification and shape recognition. Zernike moments have some desirable properties such as rotation invariance, robustness to noise, expression efficiency, and multi-level representation for describing various shapes of patterns so our proposed system recognizes not only shapes but also the shapes even if they have some deformations providing high efficiency than any other techniques.

The algorithm implemented with the robot arm to catch the objects. For controlling and enhancing the motion of humanoid robot to autonomously grasp objects, a new hardware circuit is designed and Q-learning technique is presented assisting the robot to find, change its state and catch objects. The performance of vision system based on simulation was 100% however practically was 98.33%, while it is very fast and scalable. Also, object edge detection based color has been used to enforce the whole operation for finding object location then catches it by learnt robot arm. Our system provided Arabic dialogue interface system that allows dialogue between robot and user. This makes robot really assistant to human in real world environment. Finally our system compared to system proposed in research papers in [6,16,17] in table (4).

**Table 4 Comparison between the proposed system and the work of [6,16,17]**

	<b>J. Toontham 2013,2012, T. Kitti 2011</b>	<b>Proposed System</b>
Edge detection	Using Fuzzification	Using Fuzzification and fastest edge based color
Vision System	Object detected by shape only ,color detection not focused in papers	Object detected by shape and color Even with some deformation
Shape detection	Using HT,NN,HCD	Using new techniques such as NN decreasing its complexity and its problems , Neuro WT, NeuroHT, NeuroHCD, difference and Neuro Zernike momenets
Image size	Dependent on application and image taken	<ul style="list-style-type: none"> <li>• Scale independent</li> <li>• independent on image size</li> <li>• rotation independent</li> </ul>
Natural language Understanding System	Was not his goal	Allow communication through dialogue between robot and user
Talking System	Was not his goal	Use Arabic text to speech tool allow robot answers convert to speech
Motion System	Using The robot arm is a revolute robot type (RRR) with microcontroller circuit to separate objects 	We using OWI-535 wired control robot arm kit with hardware circuit control robot arm 
Learning Algorithm	Was not his goal	Used to learning robot arm to trace and grip

#### References:-

[1] M. Vukobratović, "Humanoid Robotics, Past, Present State, Future", Director Robotics Center, Mihailo Pupin Institute, 11000 Belgrade, P.O. Box 15, Serbia, E-mail: vuk@robot.imp.bg.ac.yu, SISY 2006 • 4th Serbian-Hungarian Joint Symposium on Intelligent Systems, pp 13-27.



- [2] V. Graefe, R. Bischoff, "Past, Present and Future of Intelligent Robots", Intelligent Robots Lab, LRT 6, Bundeswehr University Muenchen, 85577 Neubiberg, Germany, <http://www.UniBw-Muenchen.de/campus/LRT6,CIRA> 2003, Kobe, pp 1-10.
- [3] C. Pasca, "History of Robotics", University of Ottawa, ENRICHMENT MINI-COURSE, Robotics – Intelligent Connection of the Perception to Action, May 5, 2003, pp1-46.
- [4] R. JARVIS, "INTELLIGENT ROBOTICS: PAST, PRESENT AND FUTURE", International Journal of Computer Science and Applications, Vol. 5, No. 3, pp 23 – 35, 2008.
- [5] M. Takizawa, Y. Makihara, N. Shimada, J. Miura and Y. Shirai, "A Service Robot with Interactive Vision- Object Recognition Using Dialog with User -", Osaka University, 2-1 Yamadaoka, Suita, Osaka 565-0871, Japan, E-mail: shimada@eng.osaka-u.ac.jp, 2003.
- [6] J. Toontham and Ch. Thongchaisuratkrul, "An Object Recognition and Identification System Using the Hough Transform Method", International Journal of Information and Electronics Engineering, Vol. 3, No. 1, January 2013.
- [7] H. Holzapfel, D. Neubig, A. Waibel, "A dialogue approach to learning object descriptions and semantic categories", Contents lists available at ScienceDirect, Robotics and Autonomous Systems 56 (2008) 1004\_1013.
- [8] J. Carbonell, Towards a self-extending parser, in: Annual Meeting of the Association for Computational Linguistics, 1979.
- [9] R. Becher, P. Steinhaus, R. Zöllner, R. Dillmann, "Design and implementation of an interactive object modelling system", in: Proceedings of ISR 2006 and Robotik 2006, Düsseldorf, 2006.
- [10] M. Khalifa, V. Liu, "KNOWLEDGE ACQUISITION THROUGH COMPUTER-MEDIATED DISCUSSIONS: POTENTIAL OF SEMANTIC NETWORK REPRESENTATIONS AND EFFECT OF CONCEPTUAL FACILITATION RESTRICTIVENESS", Twenty-Sixth International Conference on Information Systems, 2005, pp 221-232.
- [11] P. Tanwar, T. V. Prasad, M. S. Aswal, "Comparative Study of Three Declarative Knowledge Representation Techniques", Poonam Tanwar et. al. / (IJCSSE) International Journal on Computer Science and Engineering Vol. 02, No. 07, 2010, 2274-2281.
- [12] S. H. Uweil, B. Wrede, and G. Sagerer, "Robust Speech Understanding for Multi-Modal Human-Robot Communication", Faculty of Technology, Applied Computer Science Bielefeld University, 33594 Bielefeld, Germany, 2006.
- [13] Al. Ramsay, H. Mansour, "Towards including prosody in a text-to-speech system for modern standard Arabic", Received 13 March 2006; received in revised form 22 June 2007; accepted 22 June 2007 Available online 6 August 2007, Science Direct, Computer Speech and Language 22 (2008) 84–103.
- [14] J. Maitin-Shepard, M. Cusumano-Towner, J. Lei and P. Abbeel. Cloth Grasp Point Detection based on Multiple-View Geometric Cues with Application to Robotic Towel Folding, 2010.
- [15] A. L. Thomaz, C. Breazeal, "Teachable robots: Understanding human teaching behavior to build more effective robot learners", Interactive Computing, Georgia Institute of Technology, USA, MIT Media Laboratory, USA, Received 9 August 2006; received in revised form 25 September 2007; accepted 27 September 2007
- [16] T. Kitti, T. Jaruwan, and T. Chaiyaporn, "An Object Recognition and Identification System Using the Harris Corner Detection Method", International Journal of Machine Learning and Computing, Vol. 2, No. 4, August 2012.
- [17] J. Toontham, Ch. Thongchaisuratkrul, "The Comparison of Object Recognition and Identification by Using Image Processing Base on the Neural Network, the Hough transform and the Harris Corner Detection", International Journal of Machine Learning and Computing, Vol. 1, No. 4, October 2011.
- [18] E. M. Saad, M. H. Awadalla, H. E. I. Ali, R. F. A. Mostafa, "Object Manipulation Using a Humanoid Robot", International Journal of Engineering and Advanced Technology (IJEAT), ISSN: 2249 – 8958, Volume-1, Issue-6, August 2012.

### Author' biography with Photo

**El Sayed M. Saad** is a Professor of Electronic Circuits, Faculty of Engineering, Univ. of Helwan. He received his B.Sc. degree in Electrical Engineering (Communication section) from Cairo Univ., his Dipl.-Ing. Degree and Dr.-Ing degree from Stuttgart Univ., West Germany, at 1967, 1977 and 1981 respectively. He became an Associate Prof. and a Professor in 1985, and 1990 respectively. He was an International scientific member of the ECCTD, 1983. He is Author and/or Coauthor of 132 scientific papers. He is a member of the national Radio Science Committee, member of the scientific consultant committee in the Egyptian Eng. Syndicate for Electrical Engineers, till 1 May 1995, Member of the Egyptian Eng. Syndicate, Member of the European Circuit Society (ECS), and Member of the Society of Electrical Engineering (SEE).

**Medhat H. Awadalla** is an associate professor at Communication and Computer Department, Helwan University. He obtained his PhD from university of Cardiff, UK. Msc and Bsc from Helwan university, Egypt. His research interest includes cloud computing,

sensor networks, high performance computing and real time systems.



**Hossam Eldin I. Ali** received his B.Sc. degree in Communications & Electronics Engineering, his M.Sc. degree in Computer Engineering, and his Ph.D. degree in Computer Engineering from Helwan University, Cairo, Egypt, in 2000, 2004, and 2009 respectively. He has three published papers at M.Sc. degree, and seven published papers at Ph.D. degree. He is currently a Teacher at Electronics, Communication & Computer Department, Faculty of Engineering, Helwan University.

**Rasha F. A. Mostafa** is an Assistant Teacher at Communication and Computer Department, Helwan University. Received her B.Sc. degree in Computer Engineering, and her M.Sc. degree in computer Engineering from Helwan University, Cairo, Egypt in 2001 and 2008, respectively. Her research interest includes Artificial intelligent, Robot,, Security and operating System.

