Interactive Learning for Humanoid Robot

El Sayed M. Saad[1], Medhat H. Awadalla[1, 2], Hosam Eldin I. Ali[1], Rasha F. A. Mostafa[1]

[1] Department of Communications, Electronics and Computers, Helwan University, faculty of Engineering Cairo, Egypt
[2]Department of Electrical and computer engineering, SQU, Oman

Abstract
Acquiring new knowledge through interactive learning mechanisms is a key ability for humanoid robots in a natural environment. Such learning mechanisms need to be performed autonomously and through interaction with the environment or with other agents/humans. This paper proposes a vision system, where robot can learn how to identify the geometric shapes and colors of the objects. Furthermore, the paper proposes a natural language understanding system, where the robot can learn 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. Moreover, integration between the proposed vision and natural language understanding systems has been presented. Intensive experiments have been conducted indoor to address the validity of the proposed system. The achieved results show that the overall system performance is high compared with the related literature to the theme of this paper.

Keywords: *Vision System, Object Detection, Color detection, Natural Language Understanding, knowledge Representation, Semantic Networks*

1. Introduction

Humans are the most advanced creatures of the nature. It is believed that humanoid robots will be the most advanced creatures of humans. Among the man-made creatures such as automobile hand-phones and multimedia devices, robots of future will hopefully be the most ideal assistants to human beings [1]. In the future we will see "personal robots" that will entertain, comfort and serve people in their private lives and homes. While presently robotic servants or butlers exist only in the form of early prototypes in a few research laboratories, they are expected to become as ubiquitous as PCs in the future [2-4].

An important aspect of humanoid robots in a natural environment is the ability to acquire new knowledge through learning mechanisms, which enhances an artificial system with the ability to adapt to a change or new environment. In contrast to most learning algorithms applied in machine learning today, which mainly work with offline learning on training samples, such learning mechanisms need to be performed autonomously, and through interaction with the environment or with other agents/humans. Here, in this paper, the proposed vision system and dialogue offer appropriate means.

The fact that robots have to be autonomous in such a way that they have to do everything without the intervention of humans. Since the proper system is the good vision system, so the question arises here is: how to make a robot can see like a human? The only way is using of a camera. For many applications in robot vision interested in locating the object by giving it a distinctive color from the surrounding environment as an application to recognize the ball in pitch between two teams of humanoid soccer robot team [5, 6], using laser, sonar, or using camera for robot vision system, or learning to classify objects into categories in human development. Such ability is crucial for robots that have to operate in human environments where object categorization skills are required to recognize complex object categories (e.g., metal objects, empty bottles, etc.) [7]. 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 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, that entered sentence interaction results in errors, new words occur in writing sentence that cannot be understood by the system. Thus, the dialogue system needs to 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 needs to create a model of the object's semantics, which describes the type, color, shape, and properties of the object and what the object can be used for. All previous data will be addressed using Arabic language.

The remainder of the 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, and the integration of vision system and the natural language understanding system. Section 3 presents experiments and discussion. Section 4 concludes the paper.

2. The Proposed System

Our proposed system consists of two parts, stating as seen in figure (1) as follows:-

A. Vision System.
B. Natural Language Understanding System.
C. Merging System.

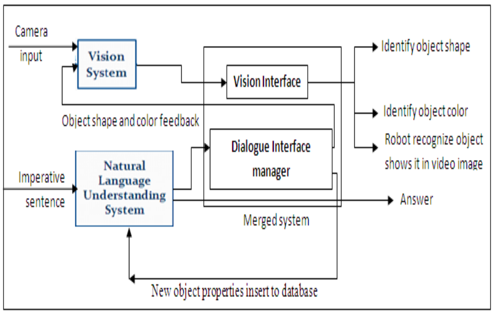


Fig.1.System overview

2. A. Vision 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; he could recognize them by himself later. The main aim of the proposed system is to make robot behaves as child, once it is learned the skill of how to recognize object 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 the following procedure and is demonstrated in the flowchart shown in figure2.

The vision system procedure:
a- Acquiring image,
Images taken from the robot's camera can easily be fed to Matlab program using the *'videoinput'* 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 have to be extracted from the video with a frame grabber [8]. A so-called snapshot is taken out of the video input and this single image is used for object and color detection.

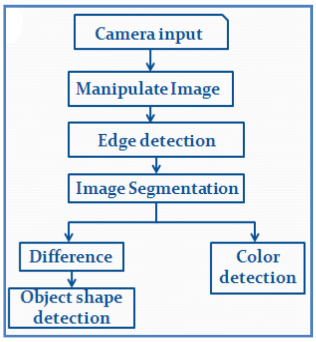


Fig.2.Object and color detection flow chart

b- Image manipulation,
Image frame taken from a camera is a color image. In Matlab, image automatically is coded using the 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 then the image is filtered to remove any added noise due to lighting. A well-known noise filter is the median filter. In Matlab, this filter can be used with the *'medfilt2'*-function [8], as shown in figure 3.


Fig.3. Converting image to black and white

c- Edge detection
The black and white image is converted into edge image, as shown in figure (4), using 'edge'-function [8].

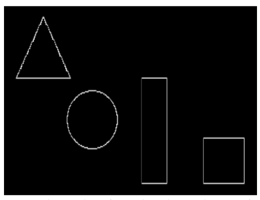


Fig.4 The obtained edge detection

d- Image Segmentation,
The Previous image is used to find the boundary of each object by using *'boundaries'*-function [8]. These indices are used to cut the black and white image to the set of images, each of which contains only one item, as shown in figure 5, each image is then used to determine the object's shape and color.

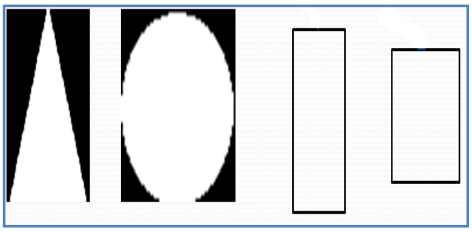


Fig.5.The segmented object image

e- How to identify the shape of an object?
There are many ways to determine the location of the element, including the mean and variance of number of ones in an image. However, the previous methods cannot determine the shape of the object, accordingly, a new method is proposed that calculates the difference in the number of ones in each row for the middle of the line of the image, if it always increases, the object's shape is triangle or if it increases and then decreases, the object's shape is circle and if it does not change, the object is a square or a rectangle, it depends on the dimensions of the image as shown in figure 6. All written functions are based on *'diff'*-function [8].

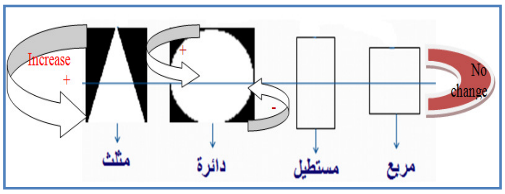


Fig.6. Object shape identification

f- How to identify the color of an object?
Identifying the object color can be achieved through the usage of the borders of each object, which it is obtained previously and separating each object's color into a separate image, and then finding the color by creating a color map using *'colormap'* function [8], where each color can be visualized as a 3d matrix. Finding the average of this matrix can identify the color of the object from the table (1) shown below.

Table.1: Colormap

	B	***G***	***R***
Black	0	0	0
White	1	1	1
Red	0	0	1
Blue	1	0	0
Green	0	1	0
Cyan	1	1	0
Magnetic	0	1	1

The previous procedure has been applied to the image shown in fig. 7, and the achieved results confirm the validity of the proposed approach. All results identify the color and shape of the object given to the user in Arabic.

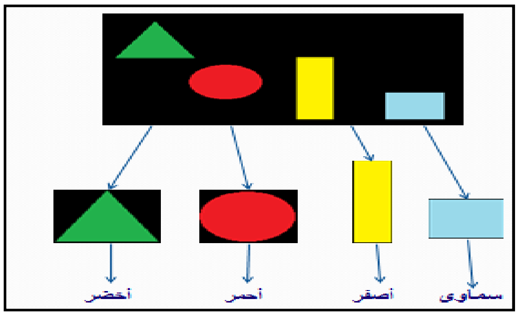


Fig.7.Object color identification

2. B. Natural Language Understanding System

Interactive learning for artificial systems has been studied in several systems. However, the number of approaches that allow interactive knowledge acquisition for humanoid robots is still comparably small [9]. 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 8. The proposed system has the following components:

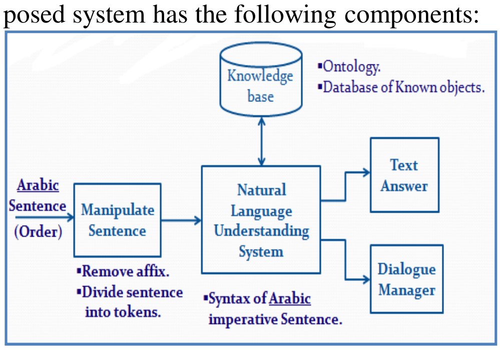


Fig.8. Natural Language Understanding System Overview

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 in the following figure 9. The input sentence is divided into a set of tokens, and then the affix such as"ال"added at the beginning of the word or "ى" added at the end of the word is omitted.

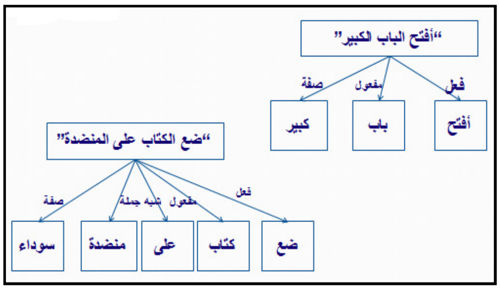


Fig.9. Example of the form of command syntax

b- Ontology

Our ontology is inspired by the work in[9,10], however it is applied to the Arabic language, as shown in figure 10, and also we have added that the object is classified by its color and shape, linking between this program and the vision program will be explained later in this paper. Knowledge representation is an area in artificial intelligence that focuses on the design of formalisms which can explicitly represent knowledge about a particular domain, and the development of reasoning methods for inferring implicit knowledge from the represented explicit knowledge. Semantic network form a family of knowledge representation formalisms which can be used to represent and reason with conceptual knowledge about a domain of interest. For the classification of each object and storing its data in the database, we use semantic network [11]**.** A semantic network is a simple representation scheme that uses a graph of labeled nodes and labeled, directed arcs to encode knowledge [12, 13].Information type and semantic categories of objects are modeled in ontology. The object ontology provides inheritance information and defines properties that can be associated with objects.

c- Dialogue,

Dialogue begins when the robot does not find the object in its database, or the user enters an imperative syntax error. Dialogue is in a form of questions and answers words, mutual between the user and robot in Arabic as shown in figure 11 and it is different from the work in [9] where the answer is yes ,or no only, and in English. There are ten commands the user can use them such as, "أحضر", "أفتح","أغلق","أملى", and,"أنظر".

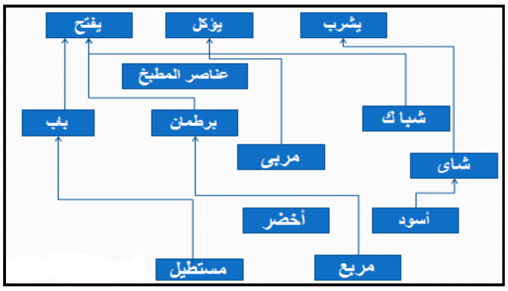


Fig.10. Ontology organization with functional concepts, type hierarchy and properties.

Fig.11. Dialogue sample

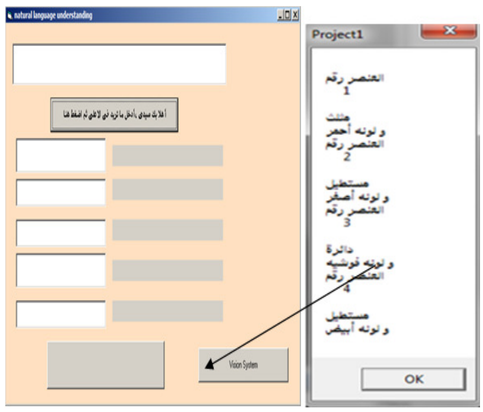
2. C. Merging System

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 that 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 can deal directly with Visual Basic programs. The developed interface gives the user different capabilities to choose as follows: -

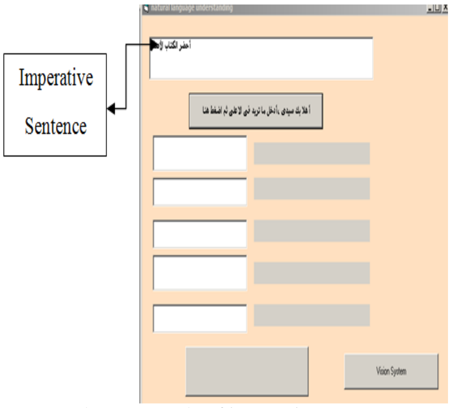
1- The user can ask the robot to describe the scene in the front of itself in terms of the shapes of the objects and their colors by using the developed vision program just by pressing a key called the vision system in the developed interface shown in figure12.a.

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 the object is found. Then the object will be sent to the vision system to recognize its shape and Color, as shown in figures 12.b and 12.c.

When the robot could not recognize the object because it is not in its database or the user entered a syntax error, a dialogue between the user and the robot will start, the user will answer some questions as shown in figure 12.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 vision system will be able to identify it.



(a) Result from calling vision system



(b) Example of imperative sentence

(c) Results when the object found

(d) Results if the object is not found

Fig.12. Example of the proposed system scenario

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 pictures from a camera installed on the laptop, the destination in an interview of an embodiment of the kitchen, and the kitchen components, such as door, window, a cup and also a piece of cheese cooked. Background with one color, black, is chosen. The achieved results from the vision program have a good precision as seen in figure 13, although the vision affected by the camera resolution, and lighting. Our camera resolution used was (640*480). Furthermore, the achieved results show that the accuracy of natural language understanding program is very high even for different users as long as they know the basics of Arabic.

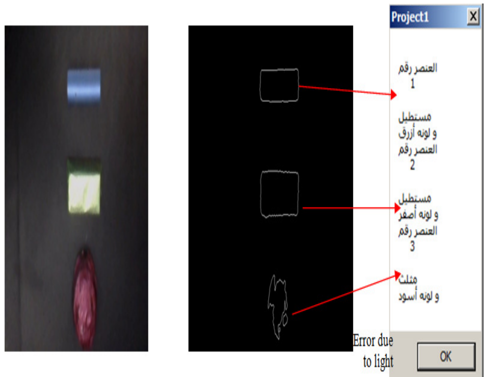


Fig.13 Achieved results of a real image

4. Future Work

Our future work will focus on: first, the robot will recognize the place of things, by adding a question in Arabic interface using the word "where" and currently it is implemented, however it is still under test. Second, the robot will determine a place for the user to put things, and also direct the user towards the place of the thing. The management process to move the robot towards the place where the item is there can be implemented in two phases. The first phase, which is already performed, is to create a direction which will guide the robot motion. Second, is to apply it to a real robot to check its validity which is currently running.

5- 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] H.J.C. Luijten,"*Basics of color based computer vision implemented in Matlab",*TechnischeUniversiteit Eindhoven, Department Mechanical Engineering, Dynamics and Control Technology Group, Eindhoven, June, 2005, pp 1-24.
[6] E. Menegatti, S. Behnke, C. Zhou," *Humanoid soccer robots*", Robotics and Autonomous Systems, contents lists available at Science Direct, journal homepage: www.elsevier.com/locate/robot, Robotics and Autonomous Systems 57 (2009) 759_760.
[7] J. Sinapov and Al. Stoytchev, "*Object Category Recognition by a Humanoid Robot Using Behavior-Grounded Relational Learning",* Developmental Robotics Laboratory, Iowa State University, {jsinapov, alexs}@iastate.edu,2011,pp 1-7.
[8] MathworksMatlab Image Processing function list, http://www.mathworks.com/products/image/functionlist.html, 2012.
[9] H. Holzapfel, D. Neubig, A.Waibel,"*A dialogue approach to learning object descriptions and semantic categories"*, Contents lists available at Science Direct, Robotics and Autonomous Systems 56 (2008) 1004_1013.

[10] J. Carbonell, Towards a self-extending parser, in: Annual Meeting of the Association for Computational Linguistics, 1979.
[11] 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.
[12] M. Khalifa, V. Liu," *KNOWLEDGE ACQUISITION THROUGH COMPUTERMEDIATEDDISCUSSIONS: POTENTIAL OF SEMANTIC NETWORK REPRESENTATIONS AND EFFECT OFCONCEPTUAL FACILITATION RESTRICTIVENESS", Twenty-Sixth International Conference on Information Systems,2005, pp 221-232.*
[13] P. Tanwar, T. V. Prasad, M. S. Aswal, "*Comparative Study of Three Declarative Knowledge Representation Techniques"*, PoonamTanwar et. al. / (IJCSE) International Journal on Computer Science and Engineering Vol. 02, No. 07, 2010, 2274-2281.

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 l 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.