Humanlike Robotic Vision System Using Artificial Neural Network

Kazi Abu Zubair*, Subha Singha, Sabrina Alam, Kazi Tanvir Ahmmed
Department of Applied Physics, Electronics and Communication Engineering
University of Chittagong
Chittagong, Bangladesh
*kzubair.bd@gmail.com, subha_cu10@hotmail.com, sabrina.apece@cu.ac.bd, tanvir@cu.ac.bd

Abstract— **Robotic manipulator, which is a kind of robot that interacts with its environment in order to manipulate objects, is the most extensively-employed kind of robot in practical field. Most of the robotic manipulators functioning in the industries and other fields are programmed to manipulate objects from fixed predefined points in space. However, this might create problems in real world situation where the location of the object is arbitrary and cannot be predefined. This problem can be solved if the robot is able to recognize the object by learning which one is the specific object and which one is not. The robot can then find the position where the object is located in order to drive the manipulator tool point towards that position. In this paper, an approach has been taken to design and analyze performance of a humanlike robotic vision system for robotic manipulator using artificial neural network along with image processing. Here, the robot captures image of the environment, analyzes it and determines the points in space where the object to be manipulated is locate.**

Keywords—Robotics; Artificial Neural Network; Artificial Intilligence; Object Recognition; Image Processing

I. INTRODUCTION

The robot vision system must be considered as part of a larger entity that interacts with the environment by analyzing images to produce descriptions of what is imaged [1]. The robot vision before the development of Artificial Neural Network relied mostly on image processing using pixel by pixel calculation. Gibson (1950) introduced first a concept on optical flow and based on his theory, mathematical models for optical flow computation on a pixel-by-pixel basis are developed [2-3]. Larry Roberts presented a research on the possibility of extracting 3D geometric information from 2D views in 1960 [4]. In such researches in [2-6] the success rate of object recognition was much less in comparison with the human brain. This is because human brain, unlike conventional robot vision mechanism, has an outstanding capability of analyzing real world non linear data. In order to be able to recognize different objects successfully in different illumination, background, view point and to some extent position of the workspace [7], the vision system of the robot should be similar to the human vision. The research on how human brain processes data using its cognitive skills and the development of such computer model which replicates the functionality of human brain is a wide research field over a few decades. Such computer model of human brain, which is known as Artificial Neural Network or ANN was first introduced by McCulloch and Pitts in 1940 [8]. In 1950's Rosenblatt developed a two layered network known as 'Perceptrons' [9]. 'Perceptrons' were capable of learning certain classifications by adjusting connection weights. These studies led to the current development of the neural network.

The practical vision technology was eagerly anticipated in industry, particularly in Japan, in the mid-1960s, typically at Hitachi's Central Research Laboratory [10]. In those attempts, the success rate was not high and the need for real AI based vision system had aroused. After the development of efficient learning algorithms like 'backpropagation' [11-13], during 1990's, machine vision starts becoming more common in manufacturing environments leading to creation of machine vision industry.

Modern-day robots can be carefully hand-programmed or "scripted" to carry out many complex manipulation tasks, ranging from using tools to assemble complex machinery, to balancing a spinning top on the edge of a sword [14-15]. But the difference between human and such robot are still apparent in because every time the robot must know the exact geometric description of the objects they are dealing with. In some recent work such as in [16-20] some robotic vision system is developed where the system tries to replicate the human vision system. In the research work [14], the goal was to make the robot grasp novel objects where the robot neither requires nor tries to build a 3D model of the object. A method is presented in [21] where the robot learns new objects from scratch, by simply observing the object in different poses and scales. The system used active tracking and depth data coming from a stereo system with 'vergence' control to segment the object from the background, and was tested with a hundred different household objects, showing around 80% recognition rate on average. The drawbacks of the system were that a certain amount of texture in the objects was required [20].

Indeed, in most cases of robotic vision, there was always an intention to analyze the geometrical shape of the object for the sake of recognition. However, the human brain, which can recognize different objects easily rely mostly on the learning process and less on geometrical description. The work presented in this paper focuses on the design and performance of a robotic vision system for manipulator arm which learns to recognize specific objects only from images of the objects, in other words, only by seeing it.

II. METHODOLOGY

Five distinct objects, three with different colors and shapes and two with same color but different shape are taken as the test objects. The objects that the robot learned to recognize are as follows-

- Object 1; Long Blue Object
- Object 2; Circular Blue Object
- Object 3; Rectangular Orange Object
- Object 4; Trapezoidal Green Object
- Object 5; Rectangular Charleston Green Object

The images of different objects are listed in TABLE I. The workspace in this project is a white surface of 25cm by 25cm in dimension. The objects to be recognized will be placed on this workspace. The manipulator is considered to be anchored at (0, 0) position. A camera is fixed over the workspace to take images of what is placed on the workspace. Fig 1 shows the view of the system with the workspace below and a camera fixed above the workspace. Graphical User Interface in PC shows the identified objects with their respective position.

A. *Design of the Neural Network:*

The Artificial Neural Network (Fig. 2) that has been used in this paper is a feed forward neural network with 150 hidden layers. It has 1200 input neurons and 6 output neurons. In most application of the object recognition using ANN, the RGB image is converted to Gray-Scale image and is applied at the input of the neural network. In that case, the efficiency of object recognition is reduced because the network treats the same objects with different color as identical objects.

In the system described in this paper, the approach was to make the robot recognize color objects by taking all RGB information as the input data. RGB image is a matrix of dimension m-by-n-by-3. Each color pixel contains three values for three primary colors; Red, Green and Blue. The network is designed to deal with images of 20 by 20 pixels. For that reason, the input layer must have 1200 neurons. The schematic of the neural network is shown in Fig. 2.This is a 6-class classifier neural network. Each of last five classes corresponds to one specific object and the first class corresponds to a Non-Object class. The corresponding output class will acquire the highest value for a certain combination of the inputs. If the image given to the network has no object in it, the Non-Object class will have the highest value.

B. *Data Preparation for Neural Network*

Preparing data is an important and critical step in neural network data analysis and it has an immense impact on the success of a wide variety of complex data analysis, such as data mining and knowledge discovery [22]. The main reason is that the quality of the input data into neural network models may strongly influence the results of the data analysis [23-24]. The training data that is required to perfectly train a neural network must be developed carefully for the optimum performance.

For preparing training data for the neural network in this paper, several images in different lighting conditions are taken by the webcam. Several of such images are combined and from the combined image, the some cropped images of 20x20 pixels of specific object-class and non-object class are taken as data image which are assigned with specific value.

The prepared data contains two matrixes, one for the image data and another for their corresponding class-value. In order to feed the data to the neural network, the second matrix is converted into a 6 by N matrix where N is the number of data taken. Each row of the matrix contains '0' for negative data and '1' for positive data for the corresponding class. Fig. 3 illustrates the data preparation procedure in a flow chart.

C. *Training The Neural Network*

The training data is divided into three parts before training the network with it, where-

- Percentage of the total data used for training: 70%
- Percentage of the total data used for validation: 15%
- Percentage of the total data used for testing: 15%

The network is trained using MATLAB's Neural Network Pattern Recognition tool (nprtool). The training data prepared in the data preparation stage is corrected, extended and edited several times before feeding it to the neural network. For the ease of understanding, the whole process is divided into three approaches.

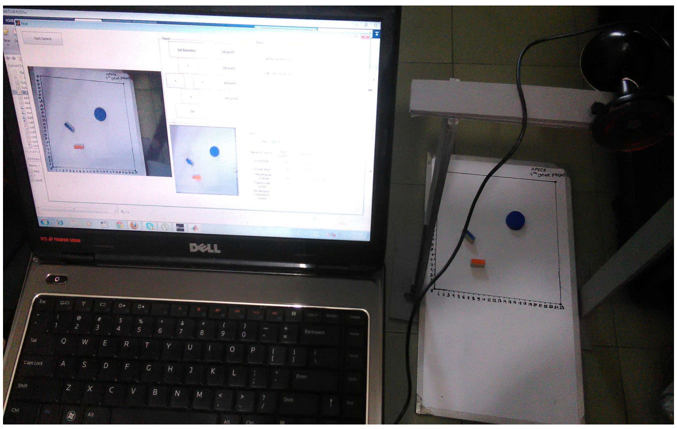


Fig. 1. The robotic vision system developed in this research work

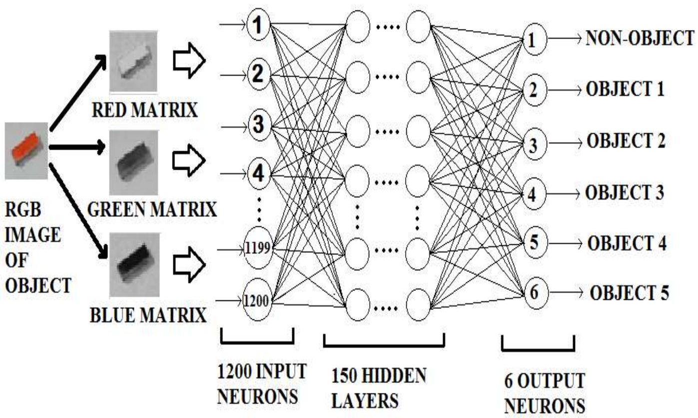


Fig. 2. Feed forward artificial neural network topology used in the vision system

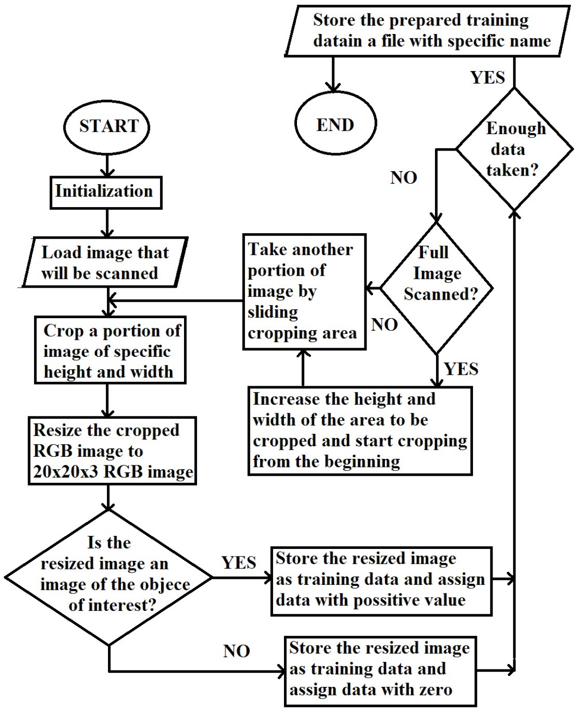


Fig. 3. Flow chart of the data preparation mechanism

1) First Approach: At first, the data taken for the ANN training contained minimum Non-Object class. The data for Object class is taken by keeping objects mostly in the center position and the Non-Object examples were taken as images of empty background image in different lighting condition.

TABLE I. DIFFERENT OBJECT AND NON OBJECT EXAMPLES IN THE FIRST APPROACH

Different Data for Object Class and Non-Object Class in the First Approach					
Object 1	*Object 2*	*Object 3*	*Object 4*	*Object 5*	*Non-Object*

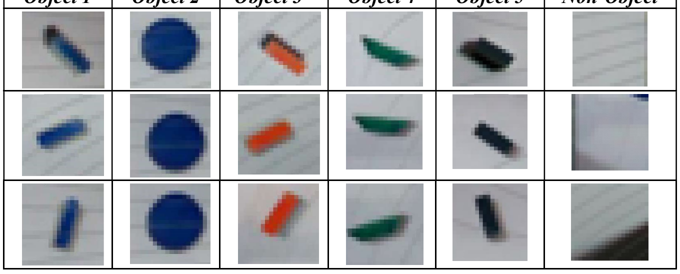


TABLE II. SOME NON-OBJECT EXAMPLES IN THE SECOND APPROACH

Some Non-Object Examples Taken in the Second Approach

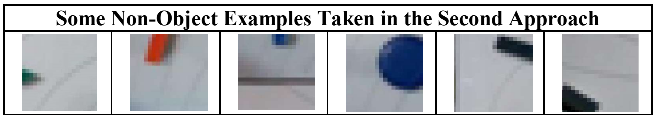
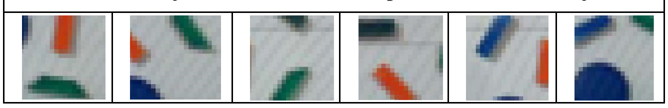


TABLE III. SOME NON-OBJECT DATA TAKEN BETWEEN THE SPACE OF TWO OBJECTS IN THE FINAL APPROACH

Some Non-Object Data between the Spaces of Different Objects



The data taken in this step is corrected several times and finally fed to the network to observe its performance. Some examples of data taken in this approach is tabulated in TABLE I.

2) Second Approach: The number of Non-Object class data was small compared to other Object classes in the first approach and the network seemed to assume some images around the object as incorrect Object-Class data which actually falls under Non-Object class. To solve this, In second approach, the number of Non Object Class has been increased to a significant number by taking images around every object which can be considered as Non-Object image(TABLE II).

3) Final Approach: With the progress achieved in the first and the second approach, the network was able to successfully recognize objects that are separated from other objects by a noticeable distance. But when several objects are close to each other, the network assumed the image taken between the spaces by which multiple objects are separated as wrong Object class. In order to make sure that the network predicts such space as Non-Object class, several Non-Object image data (TABLE III) are taken from the space between nearby objects.

D. *Recognition of Object of Interest*

The objects are recognized by taking 20x20 pixels of the workspace and feeding the cropped image to the neural network each time to see if the image selected contains the object of the interest. Fig. 4 illustrates the block diagram of object recognition procedure.

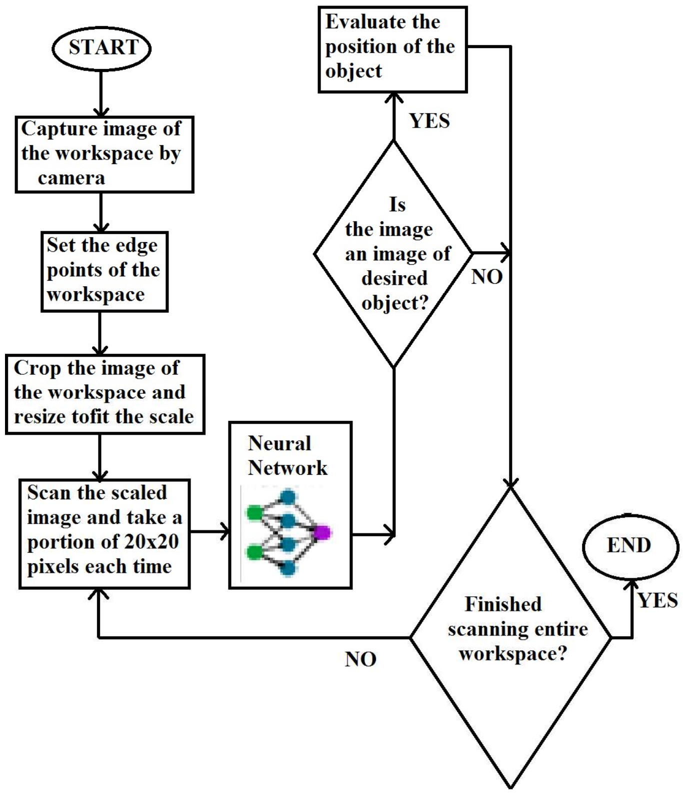


Fig. 4. Block diagram of object recognition and position evaluation

III. Performance Analysis

A. *Performance Analysis of Different Approaches:*

In the first approach, the amount of data taken was not massive. The focus of this approach was to take positive Object class data with accuracy. The performance plot of the network trained with the data taken in the first approach is shown in Fig. 5. The curve can be easily obtained using MATLAB's plotperform() function of Neural Network Toolbox. The Cross-Entropy of the best validation performance in this approach is $1.6523e^{-05}$ which is much smaller. It is clear from the plot that the network trained in this approach is much confident about the prediction. However, the characteristic of good classifier network is such that, its train, test and validation curve mostly decrease in a similar fashion. Though the cross entropy in this approach is minimum between all the approaches, it sometimes gives wrong prediction very confidently which is unacceptable.

In the second approach, an increased number of data is taken. The focus of this approach was to take more Non-Object class data to minimize errors that occurred when the network predicted some images around one object as another object. The cross-entropy for the best validation performance in the second approach is much higher than the network trained previously in the first approach. But the network shows significant improvement as the test and validation curve mostly move in a similar fashion unlike the curves in the first approach. However, the training curve falls much rapidly compared to the test and validation curve (Fig. 6).

In the final approach, the focus was to take more Non-Object class from the space between multiple objects. The performance plot of the final approach is shown in Fig. 7.The performance shows that the cross entropy for best validation performance is much greater than the network trained in the previous approaches, however, the three curves are now closer to each other than before and they converges to local minima with similar fashion.

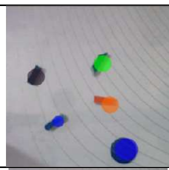
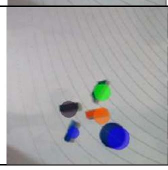
The Confusion plot of the third approach is shown in Fig. 8. There is an increased number of confusions in this approach. However, in recognizing new data, the network seemed to have better performance than the previous approaches. The confusion matrix shows that only 0.7% of the data are predicted incorrectly and the rest of the 99.3% data images are correctly recognized.

B. *Testing The Network With New Data:*

After the network has been trained successfully, the network is tested with several new images captured by the camera. The TABLE IV and TABLE V show some of these images along with the prediction of the network for these images. If the network successfully identifies an image, it draws a circle at the position where the object is located. The circles are drawn as per the following rules:

- Medium green circle for Trapezoidal Green Object.
- Small blue circle for Log Blue Object.
- Large blue circle for Circular Blue Object.
- Medium orange circle for Rectangular Orange Object.
- Medium dark circle for Charleston Green Object

TABLE IV. Test Result Of Specific Object Recognition From A New Image (Sussessful Cases)

Testing the Network With New Image (Successful Cases)		
Actual image	***Object recognition by the network***	***Observation***
		All of the five objects are identified successfully
		All of the objects are recognized successfully
		The Objects are brought closer and the network identifies each object successfully.

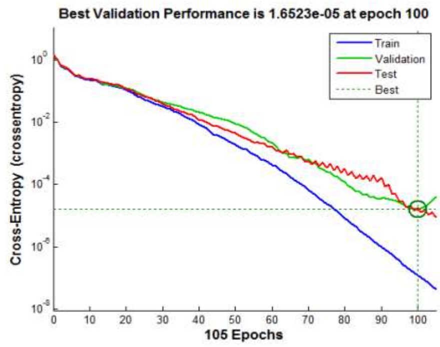


Fig. 5. Network training performance for the first approach

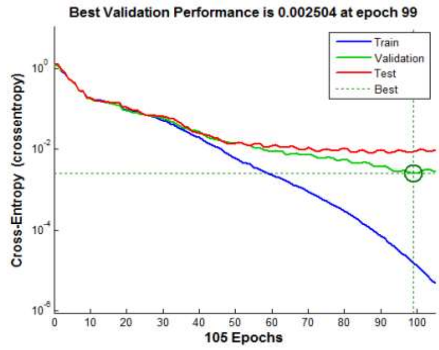


Fig. 6. Network training performance for the second approach

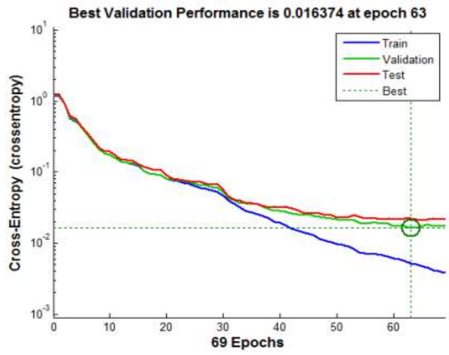


Fig. 7. Network training performance for the final approach

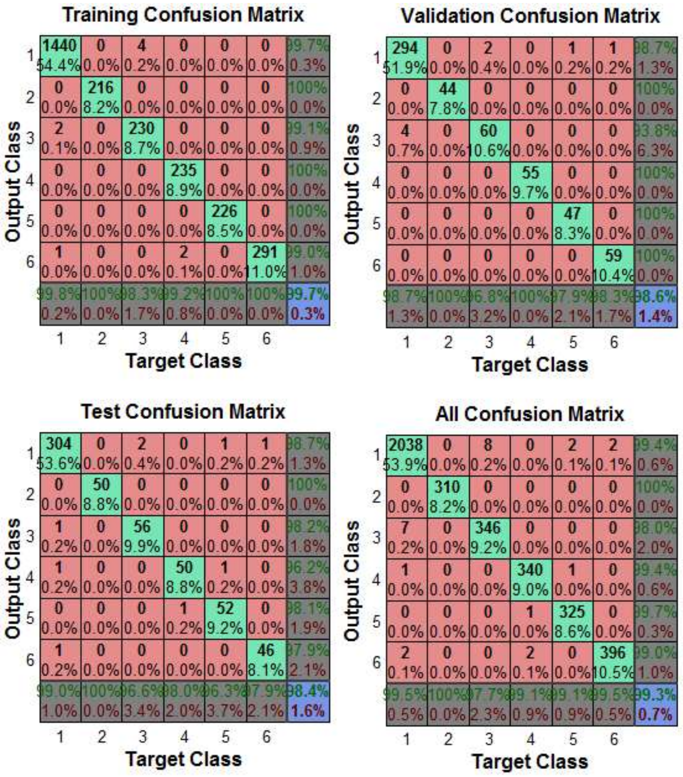
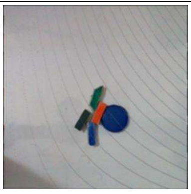


Fig. 8. Confusion matrix of the network after final approach

TABLE V. TEST RESULT OF SPECIFIC OBJECT RECOGNITION FROM A NEW IMAGE (UNSUCCESSFUL CASES)

Testing the Network with New Data (Unsuccessful Cases)		
Image captured by camera	***Object recognitions by neural network.***	***Observation***
		The objects are brought more close to each other and the network identifies all of them correctly, except an extra green object is predicted wrongly between two blue objects.
		It shows the congested condition and the network now predicts an extra Charleston green object incorrectly in the vicinity of Green object.

The result shows that the network recognizes mostly all of the objects when they are placed at a significant distant apart. Some error prediction occurs when the objects are much dense and at a congested condition. This problem might be solved if smaller area of the cropped image is fed to the network, however, this would take more time to scan the full image of the workspace.

TABLE VI, tabulates some position evaluation examples using the vision system discussed in this paper.

TABLE VI. SOME POSITION EVALUATION RESULT FOR VARIOUS OBJECTS

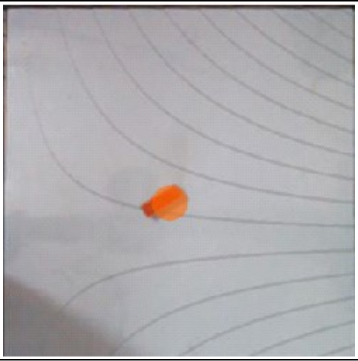
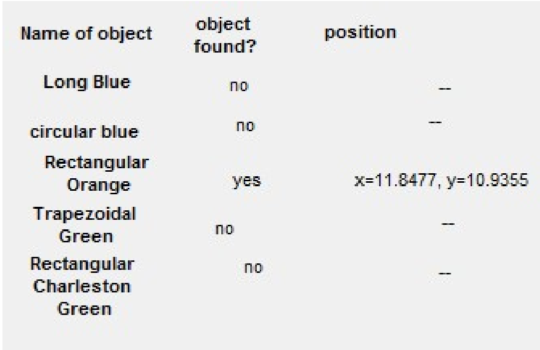
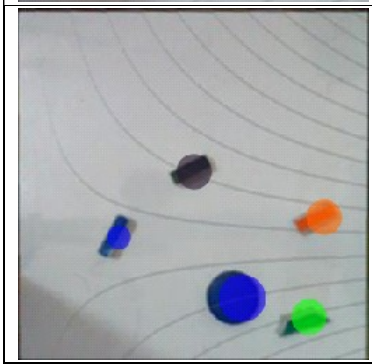
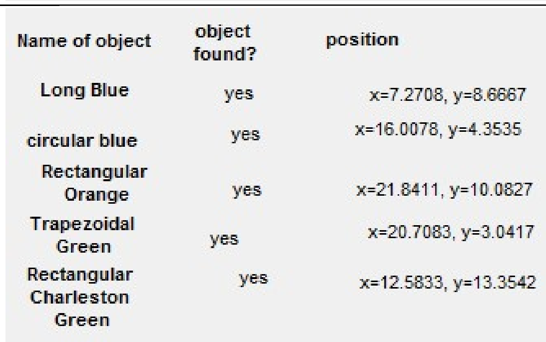
Some Example of Position Evaluation by the Vision System		
Image of the workspace with marked object	***Position evaluation by the vision system (snapshot from the graphical user interface of the system)***	***Observation***
		A rectangular orange object is given and the system successfully identifies the position of the object at (11.84,10.93)
		All of the objects are given and the system successfully identifies each object at position respectively (7.2,8.6), (16.0,4.3), (21.84,10.1), (20.7,3.04) and (12.58,13.4)

TABLE VII. RECOGNITION EFFICIENCY OF DIFFERENT OBJECT AND NON OBJECT CLASS

Recognition Efficiency of Different Object and Non-Object Class				
Object class	***No. of test image***	***Successful recognition***	***False recognition***	***Efficiency***
Object 1	310	310	0	100%
Object 2	354	345	8	97.774%
Object 3	343	340	3	99.125%
Object 4	328	325	3	99.085%
Object 5	398	396	2	99.497%
Non-Object	2048	2038	10	99.512%
Total	3781	3755	26	99.312%

The vision system in this work has a very low error or false recognition percentage. The maximum error occurs for Object 2 and is 2.2%. However, the error rate can be further reduced by careful selection of training data. Other errors are negligible. The performance of the system is tabulated in TABLE VII.

IV. CONCLUSION

In this work, a vision system for a robotic manipulator is designed so that any object of interest and its position on a scaled workspace can be recognized independently. The neural network is trained to recognize several objects and tested if the recognition system works in different lighting condition. The test result shows that the system is successful in almost all of the cases. The robotic vision system in this paper is developed using comparatively small amount of data and the network performance is evaluated in three different approaches. In the first two approaches, the network showed minimum confusion but is comparatively inefficient in recognizing new data. However, the final approach shows the best result where it is seen to be most successful in recognizing objects from new images. In some similar work, as in [25] the background is uniformly white by default and the target image information is passed to the neural network. This process gives high efficiency (99.99% at optimum lighting condition) for single object with uniform background according to the article [25]. However, the false recognition rate may increase if it must distinguish between several objects where one or more objects may lie in the vicinity of the target object. The work presented in this paper overcomes this problem and is 100% efficient for the recognition of one of the five test objects considered in this paper and the overall efficiency is 99.31% (Table VII).

REFERENCE

[1] B. Klaus & P. Horn, *Robot Vision*. Cambridge, MA: MIT Press, 1986.

[2] James J., Gibson, "The Perception of the Visual World," *The American Journal of Psychology,* vol. 64, no. 3, pp. 440-444, Jul. 1951.

[3] James J., Gibson, "The visual perception of objective motion and subjective movement," *Psychological Review,* vol. 101, no. 2, pp. p318-23, April 1954.

[4] L. G. Roberts, "Machine Perception of Three-Dimensional Solids," Ph. D. thesis, Massachusetts Institute of Technology. Dept. of Electrical Engineering, MA, 1963.

[5] Berthold K. P. Horn, & Brian G. Schunck "Determining optical flow," *Artificial Intelligence*, vol. 17, no. 1-3, pp. 185–203, Aug. 1981.

[6] Gordon, D. A., "Static and dynamic visual fields in humanspace perception," *Journal of the Optical Society of America*, vol. 55, no. 10, pp. 1296-1302, Oct. 1965.

[7] V. D. Hunt, *Industrial Robotics Handbook*. New York: Industrial Press Inc., 1983.

[8] W. S. McCulloch and W. Pitts , "A logical calculus of the ideas immanent in nervous activity," *The bulletin of mathematical biophysics*, vol. 5, no. 4, pp. 115–133, Dec. 1943.

[9] Rosenblatt, Frank, "The Perceptron: A Perceiving and Recognizing Automaton," Cornell Aeronautical Laboratory, New York. Rep. 85-460-1,1957.

[10] M. Ejiri, "Machine Vision in Early Days: Japan's Pioneering Contributions," in *Computer Vision – ACCV 2007 : Proceedings of the 8th Asian Conference on Computer Vision,* vol. 4843, Y. Yagi, S. B. Kang, I. S. Kweon, H. Zha, Eds. Berlin: Spinger, 2007. pp. 35-53.

[11] A. Atiya, "Learning Algorithms for Neural Networks," Ph. D. thesis, California Institute of Technology, CA, Sept. 2005.

[12] D. E. Rumelhart*, Geoffrey E. Hinton† & Ronald J. Williams*, "Learning representations by back-propagating errors," in *Nature,*vol. 323, no. 6088, pp. 533 – 536, Oct. 1986.

[13] Y. Chauvin, D. E. Rumelhart, *Backpropagation: Theory, Architectures, and Applications*. New York: Psychology Press, Feb. 1995.

[14] Ashutosh Saxena, Justin Driemeyer, Andrew Y. Ng, "Robotic Grasping of Novel Objects using Vision," in *International Journal of Robotics Research (IJRR),*vol. 27, no. 2, pp. 157-173 , Feb 2008.

[15] Torikai, Shinichi, Yasuko Sakagami en Satoshi Matsuoka, "Robots for the future," *Nipponia,*vol. 1, no. 13, pp. 12-17 ,2010.

[16] A. Teichman, J. Levinson, S. Thrun, "Towards 3D object recognition via classification of arbitrary object tracks," in *Proc. IEEE Int. Conf. on Robotics and Automation (ICRA)*, May. 2011, pp. 4034 - 4041.

[17] C. Kanan, "Fine-grained object recognition with Gnostic Fields," in *Proc. IEEE Int. Conf. on Robotics Applications of Computer Vision (WACV)*, March 2014, pp. 23 - 30.

[18] A. Teichman, "Practical object recognition in autonomous driving and beyond," *in Proc. IEEE Int. Conf. on Advanced Robotics and its Social Impacts (ARSO)*, Oct. 2011, pp. 35 – 38.

[19] T. Satonaka, T. Baba ,T. Otsuki, T. Chikamura,T. H. Meng, "Object recognition with luminance, rotation and location invariance," in *Proc. IEEE Int. Conf. on Image Processing*, Oct 1997,vol. 3, pp. 336 - 339.

[20] F. R. Aparici, A. R. Ayats, G. A. Ribas, X. T. Genis, "Object detection methods for robot grasping: Experimental assessment and tuning," in *Proc. 15th Catalan Conference on Artificial Intelligence (CCIA)*, 2012, pp. 123-132.

[21] H. Kim, "Semi-autonomous Learning of Objects," in *Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition Workshop(CVPRW)*, June 2006, pp. 145-150.

[22] X. Hu, "DB-HReduction: A data preprocessing algorithm for data mining applications," in *Applied Mathematics Letters* ,vol. 16, no. 6. Amsterdam, Netherlands : Elsevier, Aug. 2003. pp. 889–895.

[23] K. U. Sattler, E. Schallehn, "A data preparation framework based on a multidatabase language," in *Proc. IEEE Int. Symposium on Database Engineering and Applications*, Jul. 2001, pp. 219 - 228.

[24] Lean Yu, Shouyang Wang, Kin Keung Lai, "Foreign-Exchange-Rate Forecasting with Artificial Neural Networks," in *International Series in Operations Research & Management Science* ,vol. 107. Berlin: Spinger, 2007.

[25] J.P.N. Cruz, M.L. Dimaala, L.G.L. Francisco, E.J.S. Franco, A.A. Bandala, E.P. Dadios, "Object Recognition and Detection by Shape and Color Pattern Recognition utilizing Artificial Neural Network" in *IEEE International Conference of Information and Communication Technology (ICoICT),* March 2013, pp. 140-144.