An Approach to Estimate Object Depth from a Visual Projection Scene for Eye-in-Hand Robot Applications

Dr. Waladin K. Sa'id* **Dr. Ghassan A. Al-Kindi **** **Dr. Hussam K. Abdul-Ameer*****

Received on: 23/11/2005
Accepted on: 5/12/2006

Abstract:

In this paper, the depth of a monitored object using eye-in-hand configuration is estimated and verified by application of a proposed approach that is based on a pre-calibration process of the image and the physical area of the monitored object with respect to camera height. In addition, neural network approach is also used to manipulate the acquired vision data in order to obtain the required object depth. Both proposed approaches implemented using the eye-in-hand configuration and showed good results to improve the overall capability of the eye-in-hand configuration to perform intelligent robot tasks.

الخلاصة

يقدم البحث مقترحا لمنهجية تخمين ارتفاع الأجزاء باستخدام تقنية منظومة الرؤيا المحمولة علـى ذراع الإنـسان الآلي مع تطبيقها. تعتمد المنهجية على المعايرة المسبقة للصورة الحاسوبية للحصول على المساحة الحقيقية لمسقط الجزء و ترابطها مع المساحة الصورية وموقع منظومة الرؤيا (الكاميرة). كذلك هنالك مقترحا لمنهجية مـضافة لتخمـين ارتفـاع الأجزاء باستخدام الشبكة العصبية. أثبتت نتائج التطبيق لكلا التقنيتين الحصول على نتائج مقبولة لتخمين ارتفـاع الأجـزاء مما يحسن من الإمكانيات العامة لتطبيقات الإنسان الآلي ذو منظومة الرؤيا المحمولة على الذراع.

List of Symbols

a_i	Object image area (in ixels)
a_p	Object physical area (in m^2)
d	Object depth (in cm)
$f(x,y)$	Image function
h	Camera height (in cm)
h_a	Actual Camera height (in)
h_p	Perspective Camera height in cm)
m,n	Image size

*Control and Systems Engineering Dept.University of Technology Baghdad- Iraq
**Mechanical Engineering Dept.University of Technology Baghdad-Iraq
***Mechatronics Engineering Dept.Baghdad University - Baghdad Iraq

1. Introduction

Depth information is vital for enhancing the level of automation of present-day programmable manipulators [1]. The computation of the object depth is essential for the accurate execution of many robotic tasks, such as object handling and assembly operations [2].

Depth perception is needed to pick up a part from a random position and orientation within the work environment [1]. The early efforts in depth recovery concentrated upon stereo camera pairs with known geometries to derive an estimated object depth based on its projection in each image plane and triangulation [2]. Sinz et al [3] compared two approaches to the problem of estimating the depth of a point in space from observing its image position in two different cameras. Li and Duncan [4] stated that the most difficult part of stereo vision analyses is the correspondence problem, i.e., locating features in one image that correspond to given features in the other image. Other methods used what is commonly referred to as shape from "X", Where "X" may be shading, texture, or motion. These techniques Typically rely on lighting models and surfaces that posses a known quality [5].

Some of the promising results have been obtained with structured lighting [6]. While systems based on structured lighting principal are expensive and highly robust, their principal shortcoming is that they can be used only when the ambient light does not interfere excessively with the controlled illumination [1].

In this paper, a proposed method to estimate object depth is developed and verified. Since a number of previously published articles showed that neural network offers a potential route to the development of machine- vision algorithms that is capable of being independently adopted to a wide range of applications [7]. The neural network is employed in this paper to simulate the function of the developed method due to its promising potential in machine vision algorithms.

2 Methodology overview

The developed methods employ both the physical area and the image area of the monitored object to estimate the object depth. The object image area in the

Image (a_i) can be obtained by summing up the number of pixels that represents

The monitored object projection in the image scene. Whereas the object physical area (a_p) of the monitored object is computed using the object area in the image (a_i) and a preset calibration factor. The relation between image area, the physical area, and the camera height (h) is investigated and verified to obtain the depth of the monitored object. This relation is mapped using neural network method to activate the operation of the developed approach. Experimental setup was established to investigate the operation of eye-in-hand configuration, where a camera moving mechanism is build up and used to emulate the movements of the eye-in-hand configuration.

3 The Proposed Method

To explain and verify the proposed method, samples of different areas are used. Computer image sets of these samples are captured using an Epson Photo-500 digital camera. Where each set of these images contains projection views of that sample as captured from different camera height. Each captured image is then processed to acquire object area obtained from its corresponding image plane. Image

Preprocessing operations including threshold and noise pixel filtering are applied to identify the monitored object from image background and remove undesirable pixels that are related to noise from the image. Hence, the object area in the image (a_i) can be estimated by counting the number of white pixels in the image plane, which correspond to the monitored object. Mathematically, (a_i) can be represented as follows:

$$a_i = \sum_{i=1}^{m} \sum_{j=1}^{n} f(x, y) \qquad (1)$$

Where m and n are the image resolution and $f(x,y)$ is the image intensity function. Since the physical and image areas of the monitored object are known, the perspective camera height for this area could be derived. Thus by finding the absolute difference between the actual camera height (h_a) and the perspective camera height (h_p), depth of the monitored object (d) could be identified. Depth in this case is equal the difference between the actual and perspective camera heights. The mathematical model of the developed method can be expressed by:

$$d = \left| \; h_p - h_a \; \right| \qquad (2)$$

Where:

$$h_p = f(a_p, a_i, h) \qquad (3)$$

Hence object depth could be estimated if the type of the relationships between h_p and a_p, a_i, and h are known. However, the nature of this function is influenced by the camera properties, image resolution, lighting condition, and applied image preprocessing.

4- Neural Networks

Neural networks (NN)s have recently showed validity in many engineering applications [7]. The ability to learn and generalize from a set of training data is one of the most powerful aspects of NNs [8]. References [9] and [10] extensively illustrate NNs basics, structure, training rules, etc.

The developed method, to compute object depth is based on a pre-calibration process for series of images to enable the evaluation of equation (3). Images of objects of specified physical areas are acquired using known camera heights. The problem then is bounded to estimate the mathematical model of equation (3) which is difficult to compute using ordinary methods. Hence a multilayer feed-forward artificial neural network is adopted in this work. This neural network type can be trained using back-propagation (BP) algorithm, which is an effective tool for modeling nonlinear relations among massive input and output data when mathematical relations among these input and output parameters cannot be easily formulated [11].

The adopted neural network consists of two inputs, the first is dedicated to object image area in pixels while the other is dedicated to the object physical area in physical square units. The output of the neural network is the perspective camera height (h_p). This output is then subtracted from the physical height position of the camera (h_a) to acquire object height (d). Figure (1) represents the architectural of the adopted neural network.

The inputs and the outputs of the adopted neural network are scaled within a specific range, from (1) to (0), where the network training can be made more efficient if inputs and targets are normalized.

To map inputs to targets, BP error algorithm is used to train the adopted neural network, where BP is currently the most widely applied phenomena to neural networks to learn complicated multidimensional mapping [12]. This algorithm does not require any specific domain knowledge and provides a direct solution to nonlinear estimation problems [3] .When learning process is achieved; the network becomes ready to run with any input.

5- Experimental Setup Description

A hardware model is designed and built to emulate the function of the eye-in-hand configuration. The

mechanical mechanism is constructed so it can offer required movements. Figure (2) shows a photograph of the developed experimental setup. The end-effecter of the mechanism is a holder of the digital camera, which is linked with a personal computer to perform image processing of the captured images. A digital camera type Epson Photo *PC-500* is used to capture the images. The images were digitized at 640×480 pixel with 24 bit colors. The camera is gripped on the mechanism and connected to the PC to perform image transfer and then image processing.

6- Results and Discussion

The developed method is verified using the experimental setup, where plane square objects having areas of (1, 5, 10, 15, 20, and 25) cm^2 are used as samples, figure (3). Images of these objects are captured with different camera heights using height step of 3 cm. The images are processed to compute image area of the samples. Figure (4) presents the relation between camera height, image area, and physical area of the samples. This figure can be used as a calibration curve for this specific application. If the monitored object with a known physical area is captured, figure (5) shows a test-monitored object where this object has a 25 cm^2 area with 10 cm height made from wood. By computing the object area in image plane, the perspective camera height can be found as shown in figure (4). Since the real camera height is known, the depth of the monitored object presents the difference between these two heights.

Computer program was developed to achieve the learning process of the neural networks via BP algorithm and to test the trained network. Figure (6) shows the window of the developed program. The system can be divided into two parts, one for training the network while the other is used to test the trained network weights. Number of neurons in the input layer, output layer, first, and second hidden layers can be selected. The required accuracy for the network and the learning parameter (η) can be set by user. In the test part, network weights value can be saved and loaded. Textboxes are used to enter the inputs to get outputs with respect to load weights.

The training sets which are acquired from the experimental setup are proved to be inadequate to train the neural network and to give it generalizing property, so the mathematical model of each curve in figure (4) is used to generate more training sets, where the relation between the camera height and object image area is considered. This relation is presented as a polynomial of 4^{th}-order. These polynomials for each physical area are expressed mathematically as follows:

$$\begin{bmatrix} a_i^1 \\ a_i^5 \\ a_i^{10} \\ a_i^{15} \\ a_i^{20} \\ a_i^{25} \end{bmatrix} = \begin{bmatrix} 946364 & -14065 & 129319 & -0.45914 & 0.006453 \\ 5977 & -102803 & 867647 & -2.90299 & 0.037753 \\ 11693 & -19957 & 166663 & -5.55995 & 0.072329 \\ 163341 & -27428 & 229783 & -7.68104 & 0.101075 \\ 195086 & -31719 & 270895 & -9.19762 & 0.124022 \\ 29135 & -497902 & 411286 & -136028 & 0.175951 \end{bmatrix} \begin{bmatrix} 1 \\ z \\ z^2 \\ z^3 \\ z^4 \end{bmatrix}$$

Where the upper suffix in a_i presents the physical area and z presents the camera height in cm.

Figure (7) shows the samples that are used to estimate their depths based on the developed methods. Table (1) shows the results of both discussed methods. Both methods implement the discussed three parameters to estimate object depth. These parameters are camera height, physical area, and the image area of the monitored object. Based on the pre-calibration that is shown in figure (4), depths of the objects were estimated as presented in table (1). The results showed good accuracy of both methods compared with actual objects depth. Depth difference was less than 1 cm for all cases where the differences are more likely resulted due to the sampling limitation, thus affects the computation results of image area calculation and consequently affect the overall accuracy of the system. Depth computation using the neural network method showed to be marginally less accurate than the other investigated method. This is due to the insufficient training of the network, however, the neural network technique showed to be faster than the other method to estimate object depth, hence, could be beneficially used to establish approximate depth of the monitored object in applications where computational time is of significant. Otherwise, the other technique may establish more superior results.

6 Conclusions

In this paper, the two proposed techniques to estimate object depth showed to give acceptable results. Accuracy within 1 cm is obtained in all cases. Such accuracy may satisfy a number of pick and place robot applications. Hence, the proposed techniques may be employed with confidence in such applications. However better accuracy may be achieved using higher image resolution. The first proposed method showed to give more superior results when compared to the results obtained from using the neural network approach. This was due to the insufficient training of the neural network; however, the neural network based approach showed to give fast response, thus makes it more suitable to be implemented in application areas where time is the most significant factor.

References

[1] S. Y. Nof "Hand Book of Industrial Robotics" John Wiley & Sons-Inc., USA, 1999.

[2] C. E. Smith, S. A. Brandt and N. P. Papanikolopoulos "Eye-In-Hand Robotic Tasks in Un-calibrated Environments", IEEE Trans. On Robotics and Automation, Vol. 13, No. 6, Dec. 1997, PP. 903-914.

[3] F. Sinz., J. Q. Candela, G. H. Bakir, C. E. Rasmussen and M.O. Franz "Learning Depth from Stereo" Pattern Recognition. 26th DAGM Symposium. Proceedings (Lecture Notes in Comput. Sci. Vol.3175), 2004, p 245-252, 2004.

[4] L. Li and J. H. Duncan "3-D Translation Motion and Structure From Binocular Image Flow", IEEE Trans. On Pattern Analysis and Machine Intelligence, Vol. 15, No. 7, July 1993, PP. 657-667.

[5] K. Lee and C. Kuo "Shape From Shading With a linear Triangular Element Surface Model", IEEE Trans. On Pattern Analysis and Machine Intelligence, Vol. 15, No. 8, Aug. PP. 815-822, 1993.

[6] Z. Yang and Y. F. Wang "Error analysis of 3D shape construction from structured lighting" Pattern Recognition, v 29, n 2, Feb, p 189-206, 1996.

[7] J. Carusone and G. De-Leuterio "The 'Feature CMAC': A Neural-Network-Based Vision System for Robotic Control" IEEE International Conference on Robotics and Automation, May. 1998, PP. 2959-2964.

[8] P. Picton, " Neural Networks" Second Edition, Published by Palgrave, 2000.

[9] P. Van Der and S. Krose, "An introduction toNeural Networks", Eighth edition, University of Amsterdam, November 1996.

[10] E. J. Rzempoluck, "Neural Network Data Analysis Using Simulnet" Springer-Verlag, 1998.

[11] D. Xue and Z. Dong "An Intelligent Contra-flow Control Method for Real-Time Optimal Traffic Scheduling using Artificial Neural Network, Fuzzy Pattern Recognition, and Optimization " IEEE Control Systems Technology, Vol. 8, No. 1, Jan. 2000, PP. 183-191.

[12] Neural Networks Toolbox For use With MATLAB – Ver. 6 / Users Guide, 1998.

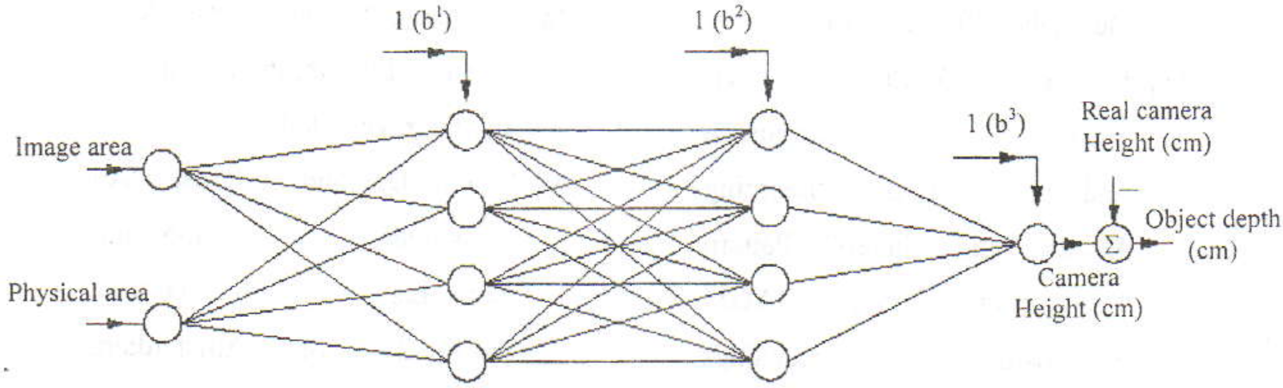
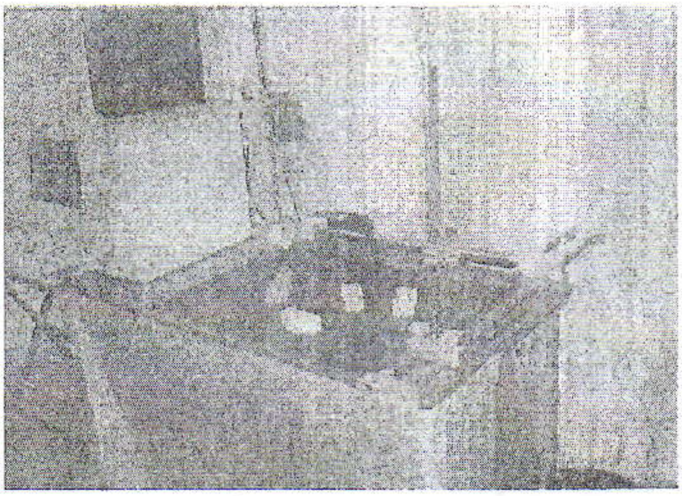
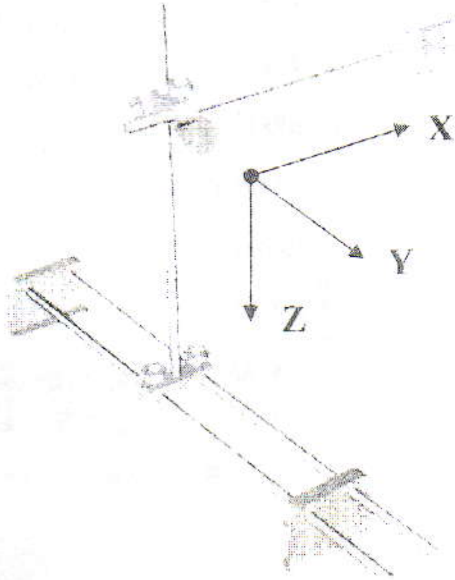


Figure (1) The adopted neural network to estimate object depth.

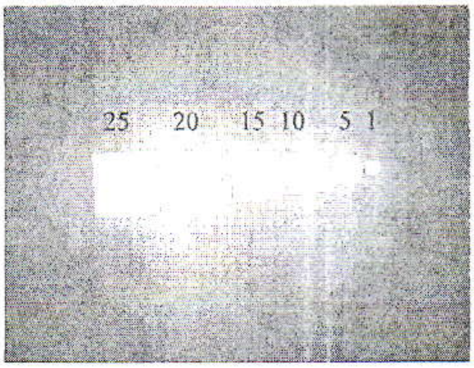


A Photograph of the Experimental Setup.



Schematic Diagram of the Developed Mechanism.

Figure (2) The Developed Experimental Setup.



Areas in cm^2

Figure (3) The samples used to calibrate the system.

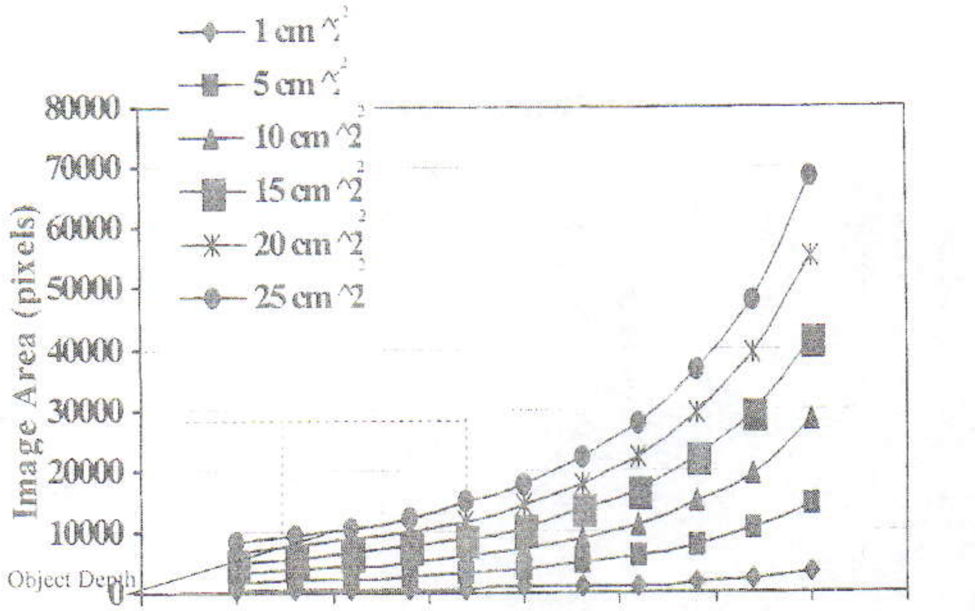


Figure (4) Image area and camera height relation.

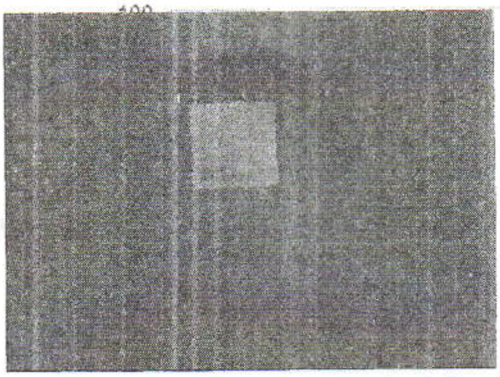


Figure (5) Test Object, camera height 10 cm.

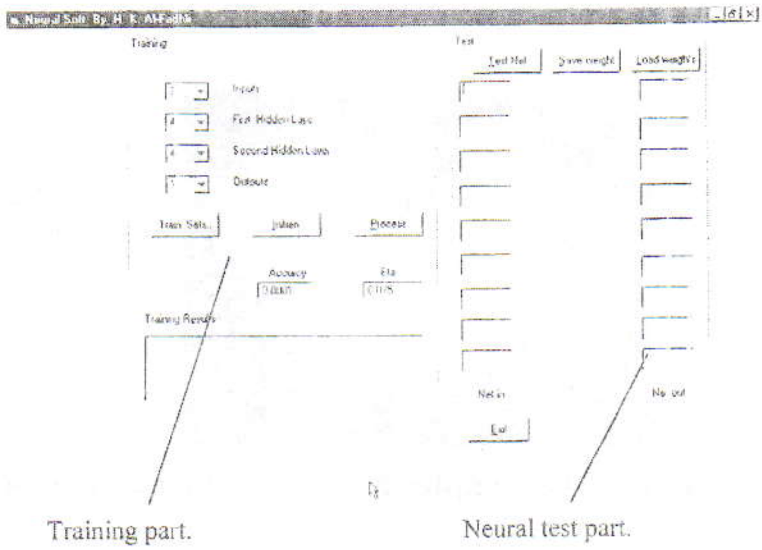


Figure (6) The Developed Neural Network Program Window.

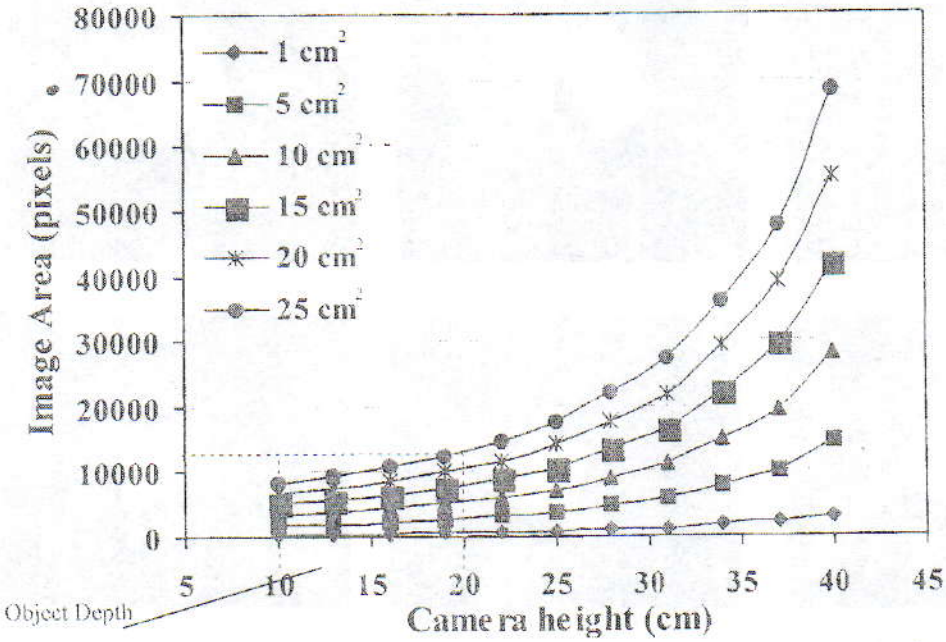


Figure (4) Image area and camera height relation.

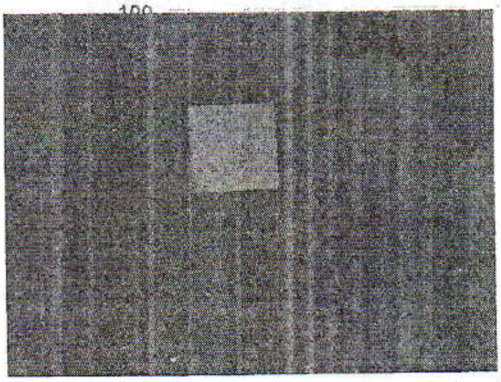


Figure (5) Test Object, camera height 10 cm.

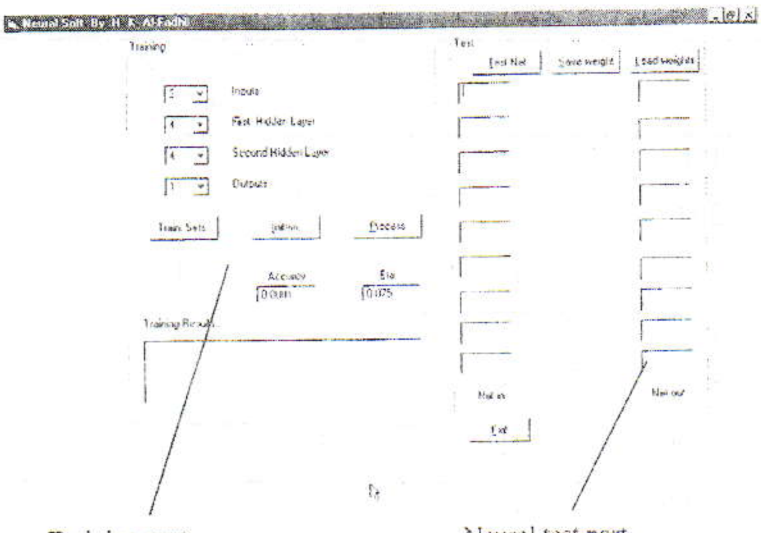
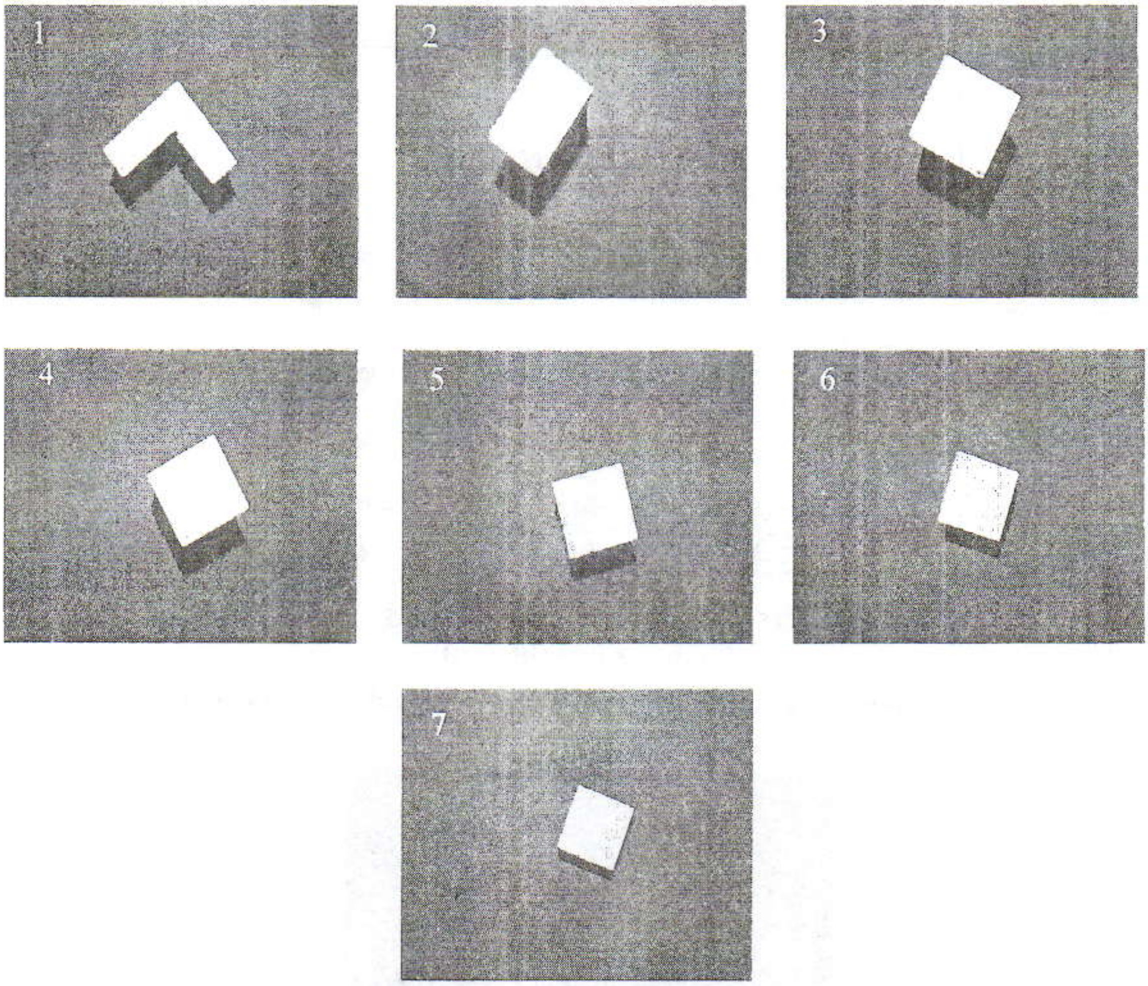


Figure (6) The Developed Neural Network Program Window.



Figure(7) The samples which are used in object depth estimation.

Table (1) Results of object depth estimation using the developed methods.

Figure	Area (cm^2)	Area (pixels)	Camera Height (cm)	Object height (Prop. method) cm	Object height (N. N.) cm	True Depth (cm)
1	20	17671	20	8.2	8.13	8.5
2	16.5	20494	25	8	7.37	8.5
3	25	23327	15	14	13.33	13.5
4	25	19201	15	11.15	10.33	11
5	25	16630	15	8.75	7.89	8.5
6	25	13997	15	5.75	5	6
7	25	12065	15	3.2	2.45	3.5