AN OPTIMISED CAMERA-OBJECT SETUP FOR 3D OBJECT RECOGNITION SYSTEM

M. Khusairi Osman, M. Yusoff Mashor, M. Rizal Arshad

Center for Electronic Intelligent System (CELIS), School of Electrical & Electronic Engineering,
Universiti Sains Malaysia, Engineering Campus,
14300 Nibong Tebal, Seberang Perai Selatan, Pulau Pinang, Malaysia.
Tel: +604-5937788 ext. 5742, Fax: +604-5941023, E-mail: *khusairi@eng.usm.my*

Abstract- **This paper proposes an effective method for recognition and classification of 3D objects using multiple views technique and neuro-fuzzy system. First, a proper condition for camera-object setup is investigated to select an optimal number of views and viewpoint condition. 2D intensity images taken from multiple sensors are used to model the 3D objects. In the processing stage, we proposed to use moment invariants as the features for modeling 3D objects. Moments have been commonly used for 2D pattern recognition. However, we have proved that with some adaptation to multiple views technique, 2D moments are sufficient to model 3D objects. In addition, the simplicity of 2D moments calculation reduces the processing time for feature extraction, hence increases the system efficiency. In the recognition stage, we proposed to use a neuro-fuzzy classifier called Multiple Adaptive Network based Fuzzy Inference System (MANFIS) for matching and classification. This proposed method has been tested using two groups of object, polyhedral and free-form objects. Computer simulation results showed that the proposed method can be successfully applied to 3D object recognition.**

Keywords- **3D object recognition, neuro-fuzzy system, moment invariants.**

I. INTRODUCTION

A model-based object recognition system finds a correspondence between certain features of the input image and comparable features of the model [1]. Such a process involves extracting features from images, and comparing them with a stored representation of the object. Recognizing 3D objects from images is a difficult problem, primarily because of the inherent loss of information in the projection from 3D to 2D [2]. In addition, the image of the 3D objects depend on various factors such as camera viewpoint and the viewing geometry since handling of 3D scenes allows additional degrees of freedom (DOF) for the orientation of the object in space [3].

Most model based 3D object recognition system considers the problem of recognizing objects from the image of a single view [4][5][6][7]. A single view may not be sufficient to recognize an object unambiguously since only one side of an object can be seen from any given viewpoint [3]. Sometimes, two objects may have all views in common with respect to a given feature set, and maybe distinguished only through a sequence of views [2]. In addition, since the single dependency on image view, this method requires complex features to represent the object.

To overcome this problem, modeling 3D objects using multiple views technique in a recognition task was proposed by some researchers. This paper continues our previous work [9] on developing a system that applied multiple view technique for 3D object recognition task. In this work, modification to the camera-object setup has been made. A suitable camera-object setup for multiple views was proposed for better recognition rate. In general, objects can be recognized not only by their shape, but also based on other visual cues, such as color, texture, characteristic motion, their location relatives to other objects in the scene, context information and expectation [29]. Our work will focus on the recognition of the isolated objects using shape information. Some researches on 3D object recognition limit their object to only polyhedral objects [17][18][19][20][21] since its shape simplicity compare to free-form objects. However, our proposed system was not limit to only polyhedral objects but also consider objects with free-form shape.

Due to the inherent loss of information in the 3D to 2D imaging process, an effective representation of properties of 3D object should be considered. We choose 2D moments as features for 3D object modeling. Although moments are commonly applied to 2D object or pattern recognition, an adaptation with multiple views technique enables this technique to be used in 3D object modeling. The simplicity of 2D moment calculation will reduce processing time, hence increases the system efficiency.

Recently, most researchers are focusing on applying neural network for 3D object recognition. Compared with the conventional 3D object recognition, neural networks provides a more general and parallel implementation paradigm [8]. In this work, we proposed to use a neuro-fuzzy classifier called Multiple Adaptive Network based Fuzzy Inference System (MANFIS). A neuro-fuzzy classifier combines fuzzy reasoning system and neural networks into an integrated functional model [33]. The integrated system will possess the advantages of both neural networks and fuzzy system. In recognition stage, using moments as inputs, MANFIS recognize an input object by matching input and model features.

II. RELATED RESEARCH

Most 3D object recognition systems used a model-based approach [8]. In this section some related works in 3D object recognition based on a model-based approach are briefly discussed.

Earlier works on 3D object recognition were influenced by Marr's philosophy [10]. Marr claims that to recognize 3D objects, one must have enough 3D information about the object to be recognized. Based on this, most early approaches attempt to describe the full 3D shape before performing recognition task. Some examples are wire-frame and surface-edge-vertex (SEV) representation [4]. A wire-frame model consists of object edges. It represents an object using possible edge junction. Jong and Buurman [13] used stereo vision system to acquire 3D wire-frames of the polyhedral objects consisting of straight lines. The system acquires images of objects in certain stable condition using stereo vision, and combined the two observations in the learning stage. This system does not require exact knowledge about poses. However, recognition is limited to polyhedral objects only. The SEV representation is a large data structure which contains a list of the edges and vertices of an object and some form of topological relationship [11].

Flynn and Jain [7] developed a system called BONSAI which identifies and localizes 3D objects in range images of one or more parts that that have been designed on a computer-aided-design (CAD) system. Recognition is performed by constrained search of the interpretation tree using unary and binary constraints derived automatically form the CAD models to prune the search space.

3D part orientation was developed for recognition and locating 3D objects in range images [22]. This algorithm uses a CAD model with simple features such as type and size. The CAD model describes edges, surfaces, vertices and their relationship. The limitation of this algorithm is that the CAD model requires complex data structure and user intervention.

In contrast to methods that rely on predefined geometry model for recognition, view-based method has been proposed by some researchers. In view-based technique, 3D object is described using a set of 2D characteristic view or aspects. Paggio and Edelman [23] showed that 3D objects can be recognized from the raw intensity values in 2D images, using a network or generalized radial basis functions. They demonstrate that full 3D structure of an object can be estimated if enough 2D views of the object are provided. Murase and Nayar [24] develop a parametric eigenspace method to recognize 3D objects directly from their appearance. Eigenvectors are computed from set of images in which the object appears in different poses. An important advantage of this method is the ability to handle the combined effects of shape, pose, reflection properties and illumination.

One of the main disadvantages of view-based technique is the inherent loss of information in the projection from 3D object to 2D image. Furthermore, the image of a 3D object depends on such factor such as the camera viewpoint and the viewing geometry. A single view-based approach may not be applicable for 3D object recognition since only one side of an object can be seen from any given viewpoint [3]. To overcome this, multiple-view technique have been proposed by several researchers [17][25][26][28][29].

Recently, neural networks have been used to solve 3D object recognition problem. Yuan and Niemann [27] develop an appearance based neural image processing algorithm for recognizing 3D objects with arbitrary pose in 2D image. Wavelet transform is used to extract compact feature for object representation and a feed-forward network using resilient backpropogation training algorithm is used to train extracted features. Kawaguchi and Setoguchi [28] proposed a new algorithm for 3D object recognition based on Hopfield nets and multiple view approach. The algorithm computes the surface matching score between the input image and the object model. Object with smallest matching error is considered as the best matched model. Ham and Park [8] proposed a hybrid hidden-Markov model and neural network

(HMM-NN) to recognize 3D objects from range image. 3D features such as surface type, moments, surface area and line length are extracted in image processing step. Features are trained using HMM and neural network is used as post-processing step to increase the recognition rate.

III. CAMERA-OBJECT SETUP

In this section, a proposed methodology for camera-object setup will be discussed. Each object to be recognized must be placed in its stable condition at the centre of the turntable. The turntable is a circular horizontal platform that can be rotated 360 degree. A, B, C and D represent the coordinate where the cameras to be placed around the turntable. A, B and C are located on the same horizontal, but differ 45^0 from each other. Point D is perpendicular to the turntable. Figure 1 shows the location of the points and object. Since all points have the same distance from the centre of the turntable, all cameras must have the same focal lengths. For features stability[1], cameras at point A, B and C are proposed to be fixed at 45^0 from perpendicular view. Camera at point D is fixed at the top of the object. Figure 2 shows how all these cameras are fixed.

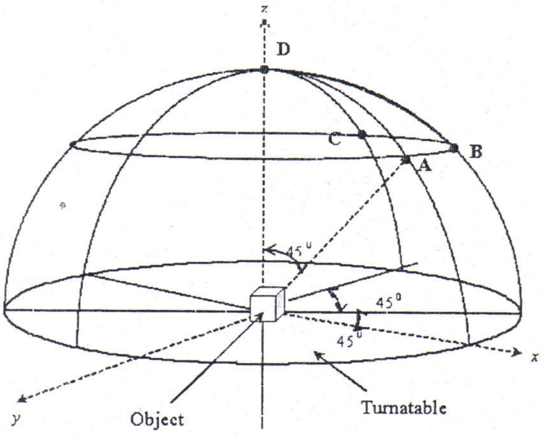

Figure 1: Image acquisition set-up

Since we have four cameras at different location, we consider eight possibilities for camera-object setup condition. Each condition is investigated to find the most effective condition for our system. Table 1 describes these conditions.

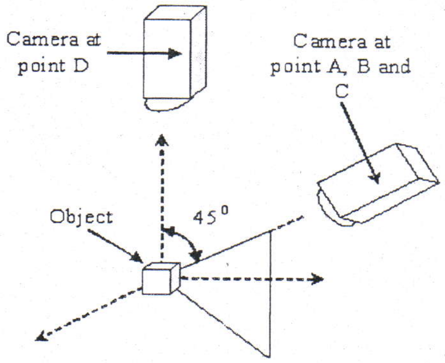

Figure 2: Camera position for point A, B, C and D

Table 1: Possibility condition for camera setup

Condition	Camera position (at point)	No. of camera used
1	A	1
2	A & B	2
3	A & C	2
4	A & D	2
5	A, B & C	3
6	A, B & D	3
7	A, C & D	3
8	A, B, C & D	4

After an object of interest is placed at the centre of the turntable, the object's images are further acquire. Then, the object will be rotated 5^0 and the same process is repeated. Each rotation will rotate the object 5^0 and so on until 360^0 is complete. Hence, for each object, we have 72 image sets. These images are partitioned into two groups, 32 image sets for training data and 32 image sets for testing data. For training data, we considered images with 0^0, 10^0, 20^0,.......350^0 condition, and the rest of the images (image with 5^0, 15^0, 25^0.....355^0 condition) are used for testing. The training data set is used to build the 3D object model in the recognition stage.

IV. FEATURES EXTRACTION

Captured images are then digitized by the DT3155 framegrabber from Data Translation Inc. and sent to the pre-processing and feature extraction stage. In the pre-processing stage, images are threshold automatically using iterative thresholding method [13][14]. This method leads to a good separation between object and background in several applications [30]. In feature extraction stage, we choose Hu's moments [15] as features for 3D modeling. Although Hu's moments are

[1] Small change in shape should produce small change in description. This will simplifies the problem of comparing shapes.

commonly and widely used for 2D object recognition, we prove that some adaptation with multiple views technique, 2D moments are sufficient to model 3D objects.

In order to understand how to utilize moment invariant method, let $f(i,j)$ be a digital image with $i = 1,2,3....M$ and $j = 1,2,3.....N$. Two-dimensional moments and central moments of order $(p+q)$ of $f(i,j)$ are defined as:

$$m_{pq} = \sum_{i=1}^{M}\sum_{j=1}^{N} i^p j^q f(i,j) \qquad (1.1)$$

$$U_{pq} = \sum_{i=1}^{M}\sum_{j=1}^{N} (i-\bar{i})^p (j-\bar{j})^q f(i,j) \qquad (1.2)$$

where

$$\bar{i} = \frac{m_{10}}{m_{00}} \quad \text{and} \quad \bar{j} = \frac{m_{01}}{m_{00}} \qquad (2)$$

From the second and third order moments, a set of seven invariant moments which is invariants to translation, rotation and scale derived by Hu are as follow:

$$\varphi_1 = \vartheta_{20} + \vartheta_{02} \qquad (3.1)$$

$$\varphi_2 = (\vartheta_{20} - \vartheta_{02})^2 + 4\vartheta_{11}^2 \qquad (3.2)$$

$$\varphi_3 = (\vartheta_{30} - 3\vartheta_{12})^2 + (3\vartheta_{21} - \vartheta_{03})^2 \qquad (3.3)$$

$$\varphi_4 = (\vartheta_{30} + \vartheta_{312})^2 + (\vartheta_{21} + \vartheta_{03})^2 \qquad (3.4)$$

$$\varphi_5 = (\vartheta_{30} - 3\vartheta_{12})(\vartheta_{30} + \vartheta_{12})[(\vartheta_{30} + \vartheta_{12})^2 - 3(\vartheta_{21} + \vartheta_{03})^2] + (3\vartheta_{21} - \vartheta_{03})(\vartheta_{21} + \vartheta_{03})[3(\vartheta_{30} + \vartheta_{12})^2 - (\vartheta_{21} + \vartheta_{03})^2] \qquad (3.5)$$

$$\varphi_6 = (\vartheta_{20} - \vartheta_{02})[(\vartheta_{30} + \vartheta_{12})^2 - (\vartheta_{21} + \vartheta_{03})^2] + 4\vartheta_{11}(4\vartheta_{11}(\vartheta_{30} + \vartheta_{12})(\vartheta_{21} + \vartheta_{03}) \qquad (3.6)$$

$$\varphi_7 = (3\vartheta_{21} - \vartheta_{03})(\vartheta_{30} + \vartheta_{12})[(\vartheta_{30} + \vartheta_{12})^2 - 3(\vartheta_{21} + \vartheta_{03})^2] - (\vartheta_{30} - 3\vartheta_{12})(\vartheta_{21} + \vartheta_{03})[3(\vartheta_{30} + \vartheta_{12})^2 - (\vartheta_{21} + \vartheta_{03})^2] \qquad (3.7)$$

where ϑ_{pq} are the normalized central moments defined by

$$\vartheta_{pq} = \frac{U_{pq}}{U_{00}^{r}} \qquad (4.1)$$

$$r = [(p+q)/2] + 1, \; p+q = 2,3,4.... \qquad (4.2)$$

V. RECOGNITION

In the recognition stage, we proposed to use a neuro-fuzzy classifier named MANFIS (multiple adaptive networks based fuzzy inference system). MANFIS contains a number of ANFIS networks [16] which are arranged in parallel combination to produce a network with multiple outputs since ANFIS is a single output network. Figure 3 shows an example MANFIS network with three inputs, x_1, x_2, x_3 , and eleven outputs, f_1, f_2, f_3.....f_{11}. For our recognition purpose, the number of input depends on the number of cameras used (see Table 1) while the number of outputs depends on the number of objects to be recognized. A hybrid learning algorithm which combines gradient descent and least square estimator is used to for learning procedure. In the recognition step, if any output node has the largest value greater that 0.5, that node is determined as 1. Otherwise, the node is considered as 0. Details discussion on algorithm and learning of ANFIS can be found in [16][31] and a brief discussion on MANFIS can be found in our previous work in [9].

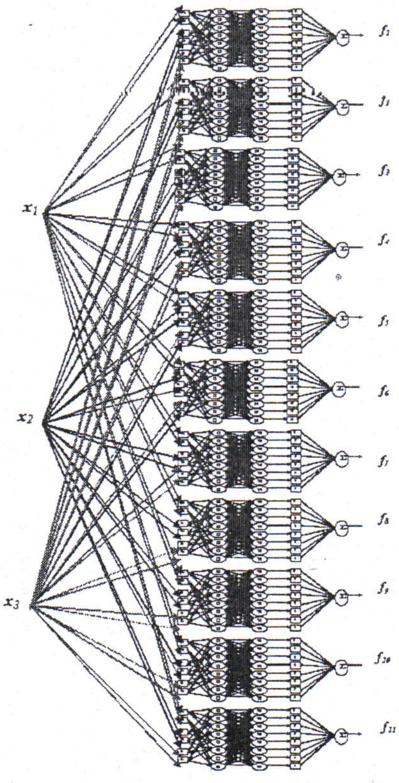

Figure 3: MANFIS network with 11 output

VI. RESULTS AND DISCUSSION

We chose two types of objects in order to analyze our system's performance. Each type consists of eleven 3D objects. The first type, Type 1 object, contains simple 3D shape like cylinder, box, trapezoid, sphere etc. The second type, Type 2 object contains free-form objects. Figure 4 and 5 show these types of object.

In order to find the best camera-object position for our system, we have conducted eight analyses using different camera positions as given in Table 1. In these cases, all MANFIS parameters are set to default values (No. of membership function, MF=2, initial step size=0.1). Hu's first moment, φ_1 (Equation 3.1) is used as feature. We choose the first type object for this analysis. Table 2 summarizes the performance of the system for each condition.

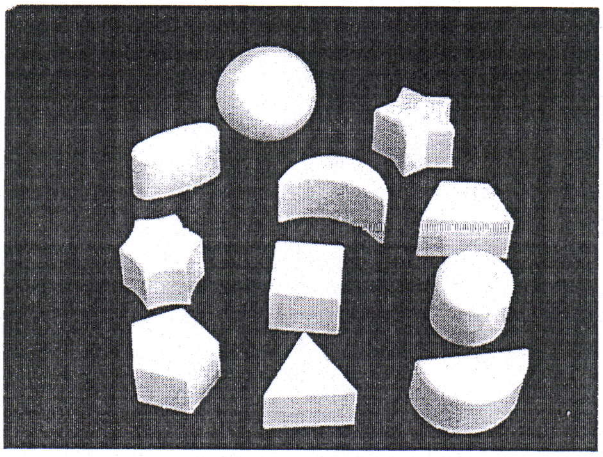

Figure 4: Type 1- simple 3D shape

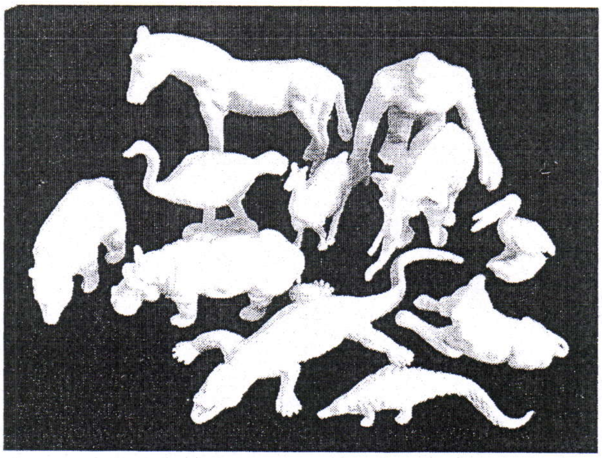

Figure 5: Type 2 - free-form object

As we can see from the table, condition 1 gave the poorest recognition rates compared to other conditions. In this case, recognition has performed through a single view. Since the features are only dependent on the single view, this method was insufficient to model 3D objects. Generally, by adding the number of views, it will increase the recognition rate to a certain rate. However, too many views, for example, condition 8 will decrease the recognition rate. So, optimum number of views with suitable camera position are required to achieve the best recognition rate. System with camera condition 7 gives the highest recognition rate with 95.71% for training and 95.20% for testing. Since condition 7 gives the best performance, we choose this method for further analysis.

Table 2: System performance under different camera setup condition

Condition	Accuracy (%)	
	Training	Testing
1	42.68	40.40
2	66.16	66.92
3	64.14	64.65
4	90.91	90.91
5	88.64	86.87
6	90.91	91.41
7	95.71	95.20
8	88.64	88.13

Table 3 and Table 4 show the system performance using simple object (Type 1) and free-form objects (Type 2) for different order of moment after some refinement (by selecting appropriate initial step size and number of MF). Our results show that better recognition rate achieved when using lower order moment compare to higher order moments. Generally, higher order moments are more sensitive to noise compare to lower order [32]. As a result, the features stability will decrease and this factor reduces the recognition rate. For Type 1 object, maximum recognition rate is 100% and for Type 2 object is 98.99%, both using first Hu's moment. Since the free-form objects have a complex shape, recognition rate is lower than polyhedral object. However, for overall performance, recognition rate achieved using this method is better compared to our previous work in [9].

Table 3: System performance for object type 1 using different Hu's moments

Hu's moment	No. of MF	Initial step size	Accuracy (%)	
			Training	Testing
φ_1	2	0.01	100.00	100.00
φ_2	2	0.20	91.92	88.38
φ_3	3	0.10	99.75	99.49
φ_4	2	0.10	57.83	50.76
φ_5	2	0.10	70.96	59.60
φ_6	2	0.10	71.46	59.85
φ_7	2	0.10	60.35	46.97

Table 4: System performance for object type 2 using different Hu's moments

Hu's moment	No. of MF	Initial step size	Accuracy (%)	
			Training	Testing
φ_1	3	0.10	99.75	98.99
φ_2	3	0.30	98.23	96.72
φ_3	2	0.20	89.14	87.37
φ_4	2	0.10	76.52	74.75
φ_5	2	0.10	74.24	72.22
φ_6	2	0.10	73.99	69.70
φ_7	2	0.10	51.20	45.96

VII. CONCLUSION AND FUTURE WORK

This paper proposes an efficient 3D object recognition system using multiple views technique and neuro-fuzzy system. A new method for camera-object position is proposed to increase the recognition rate. In feature extraction stage, Hu's moments are used to model the 3D objects. Our proposed method also shows that some adaptation with multiple-views technique, Hu's moments are able to model the 3D object well. The simplicity of moments calculation proved that our proposed system did not required complex features for 3D representation, hence reduce processing time in feature extraction stage. In recognition stage, we propose to use a neuro-fuzzy called MANFIS. Computer simulation results for two types of object show that this method can be successfully applied to 3D object recognition.

Currently, we are applying Hu's moment for the features. In the future, we will try to use other variation of moments such as Zernike and Legendre moments for comparison. Future work also tends to compare the MANFIS performance with other neural networks type such as multiple layer perceptron (MLP) and radial basis fuction (RBF) network.

REFERENCES

[1] Pope, A. R. (1994). Model Based Object Recognition: A Survey of Recent Research. *Technical Report TR-94-04*. University of British Columbia.

[2] Roy, S. D. Chaudhury, S. & Banerjee, S. (2003) Active Recognition Through Next View Planning: A Survey. *Pattern Recognition* (Accepted for Publication).

[3] Büker, U. and Hartmann, G. (1996) Knowledge-Based View Control of Neural 3D Object Recognition System. *Proceeding of International Conference on Pattern Recognition*. D:24-29.

[4] Besl, P. J. and Jain, A. C. (1985) Three-dimensional Object Recognition, *ACM Computer Survey*. 17:76-145

[5] Elsen, I.; Kraiss, K. –F.; and Krumbeigel, D. (1997) Pixel Based 3D Object Recognition with Bidirectional Associative Memories. *International Conference on Neural Netwoks*. 3:1679-1684, 1997.

[6] Roh, K. S.; You, B. J. and Kweon, I. S. (1998) 3D Object Recognition Using Projective Invariant Relationship by Single View. *Proceedings. of the IEEE International Conference on Robotic and Automation*. 3394-3399

[7] Flynn, P. J. & Jain, A. K. (1991) BONSAI: 3D Object Recognition Using Constraint Search. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 13(10): 1066-1075.

[8] Ham, Y. K. and Park, R. –H. (1999) 3D Object Recognition In Range Images Using Hidden Markov Models And Neural Networks. *Pattern Recognition*. 32:729-742.

[9] Osman, M. K. Mashor, M. Y. & Arshad, M. R. (2003) Multi-View Technique For 3-D Object Recognition Using Neuro-Fuzzy System. AIAI 2003, 24th - 25th.June 2003, Kuala Lumpur. Malaysia, paper no.6 [Invited session].

[10] Marr, D. (1982) *Vision.* San Francisco: W. H. Freeman.

[11] Haralick, R. & Shapiro, L. (1993) *Computer and Robot Vision.* Vol I, II. MA: Addison-Wesley.

[12] Jong, J. J. & Buurman, J. (1992) Learning 3D Object Descriptions from a Set of Stereo Vision Observations. *International Conference on Pattern Recognition ICPR'92.* I: 768-771.

[13] Riddler, T. W. and Calvard, S. (1978) Picture Thresholding Using an Iterative Selection Method. *IEEE Transactions on Systems, Man and Cybernetics.* 8: 630-632.

[14] Trussell, H. J. (1979) Comments on Picture Thresholding using an Iterative Selection Method. *IEEE Transactions on Systems, Man and Cybernetics.* 9(5): 311.

[15] Hu, M. K. (1962) Visual Pattern Recognition By Moment Invariants. *IRE Transactions on Information Theory.* 8(2):179-187.

[16] Jang, J. –S. R. (1993) ANFIS: Adaptive-Network-Based Fuzzy Inference System. *IEEE Transactions On Systems, Man and Cybernetics.* 23(3):665-685.

[17] Farias, M. F. S. & de Carvalho, J. M. (1999) Multi-view Technique For 3 dimensi Polyhedral Object Recognition Using Surface Representation. *Revista Controle & Automacao.* 10(2): 107-117.

[18] Park, K & Cannon, D. J. (1996) Recognition and Localization of a 3D Polyhedral Object using Neural Network. *Proceedings of the 1996 IEEE International. Conference on Robotics and Automation, ICRA 1996.* 3613-3618.

[19] Kawaguchi, T. & Baba, T. (1996) 3D Object Recognition Using a Genetic Algorithm. *Circuits and Systems, ISCAS'96.* 3:321-324.

[20] Wang, P. S. P. (1997) Parallel Matching of 3D Articulated Object Recognition. *International Journal Of Pattern Recognition And Artificial Intelligence.* 13(4): 431-444.

[21] Procter, S (1998) Model Based Polyhedral Object Recognition using Edge-Triple Features. *Phd Thesis.* UK: University of Surrey.

[22] Bolles, R. C. & Horaud, P. (1986) 3DPO: A Three Dimensional Part Orientation System. *International Journal of Robotics Research.* 5(3): 3-26.

[23] Paggio, T. & Edelman, S. (1990) A Network That Learns to Recognize 3D Objects. *Nature 343*: 263-266.

[24] Murase, H. & Nayar, S. K. (1995) Visual Learning and Recognition of 3D Objects from Appearance. *International Journal of Computer Vision.* 14: 5-24.

[25] Abbasi, S. & Mokhtarian, F. (2001) Affine-Similar Shape Retrieval: Application to Multiview 3D Object Recognition. *IEEE Transactions on Image Processing.* 10(1): 131-139.

[26] Seibert, M. & Waxman, A. M. (1992) Adaptive 3D Object from Multiple Views. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* 4(2): 107-124.

[27] Yuan, C. & Niemann, H. (2000) An Appearance Based Neural Image Processing Algorithm for 3D Object Recognition. *International Conference on Image Processing, ICIP2000.* 344-347.

[28] Kawaguchi, T. & Setoguchi, T. (1994) A Neural Network Approach for 3-D Object Recognition. *IEEE International Symposium on Circuits and Systems, ISCAS 1994.* 6: 315-318.

[29] Ullman, S. (1998) Three-Dimensional Object Recognition Based on the Combination of Views. *Cognition.* 67: 21-44.

[30] Klette, R. & Zamperoni, P. (1996) *Handbook of Image Processing Operators.* England: John Wily & Sons.

[31] Jang, J. S. R. Sun, C. T. & Mizutani, E. (1997) *Neuro-Fuzzy And Soft Computing: A Computational Approach to Learning and Machine Intelligence.* New Jersey: Prentice Hall.

[32] Prokop, R. J. & Reeves, A. P. (1992) A Survey of Moment-Based Techniques for UnOccluded Object Representation and Recognition. *CVGIP: Graphics Models and Image Processing.* 54(5): 438-460.

[33] Lin, C. T. & Lee, C. S. G. (1996) *Neuro-fuzzy Systems.* Prentice Hall.