Proceedings of International Conference on Man-Machine Systems 2006
September 15-16 2006, Langkawi, Malaysia

3D Object Recognition Using Multiple Views and Neural Networks

[1]M. Y. Mashor **[2]M. K. Osman** **[2]M. R. Arshad**

[1]Electronic & Biomedical Intelligent Systems (EBItS) Research Group, School of Mechatronic Engineering
Kolej Universiti Kejuruteraan Utara Malaysia, 02600 Jejawi, Arau, Perlis, MALAYSIA
E-mail: yusoff@kukum.edu.my

[2]Control and ELectronic Intelligent Systems (CELIS) Research Group, School of Electrical & Electronic Engineering,
Universiti Sains Malaysia, Engineering Campus, 14300 Nibong Tebal, Pulau Pinang, MALAYSIA.

Abstract

This paper proposes a method for recognition and classification of 3D objects. The method is based on 2D moments and neural networks. The 2D moments are calculated based on 2D intensity images taken from multiple cameras that have been arranged using multiple views technique. 2D moments are commonly used for 2D pattern recognition. However, the current study proves that with some adaptation to multiple views technique, 2D moments are sufficient to model 3D objects. In addition, the simplicity of 2D moment's calculation reduces the processing time for feature extraction, thus decreases the recognition time. The 2D moments were then fed into a neural network for classification of the 3D objects. In the current study, two neural network models were used to perform the classification, namely multilayered perceptron (MLP) network and hybrid multi-layered perceptron (HMLP) network. Two distinct groups of objects that are polyhedral and free-form objects were used to access the performance of the proposed method. The recognition results show that the proposed method has successfully classified the 3D objects with the accuracy of up to 100%.

Keyword

3D object recognition, neural network, moment invariants.

1. Introduction

3D objects recognition has drawn the attention of many computer vision researchers. It is the necessary step for the development of an effective vision system that is capable to operate in a variety of applications such as the automation of manufacturing process [1]. Model based vision system is the most widely used approach for shape or object recognition. In this approach, extracted features from the objects to be recognized would be matched against the previously stored features of object models [2]. Earlier researches in 3D object recognition attempt to recover full 3D shape information before performing the recognition task. This method is known as object based representation. View-based method that does not rely on predefined geometry model for recognition has been proposed by some researchers as an alternative to the conventional methods. Instead of using object models, this approach uses 2D model views. In view-based technique, 3D object is described using a set of 2D characteristic views. Paggio and Edelman [3] showed that 3D objects can be recognized from the raw intensity values in 2D images using a generalized radial basis functions. They demonstrated that full 3D structure of an object can be estimated if enough 2D views of the object are provided. Murase and Nayar [4] developed a parametric eigenspace method to recognize 3D objects directly from their appearance. Eigenvectors are computed from set of images from the object appearances in different poses.

Main disadvantage of view-based technique is the inherent loss of information in the projection from 3D object into 2D image [5]. Moreover, the 2D image of a 3D object depends on factors such as the camera viewpoint and the viewing geometry. A single 2D view-based approach may not be appropriate for 3D object recognition since only one side of an object can be seen from any given viewpoint [6]. One solution to this problem is to use several 2D views of the object. There are several researches that are based on active object recognition system [5][7], where the camera is moved around the object to gather additional multiple 2D views until enough features are gathered to sufficiently classify the 3D objects. However, this approach requires a complicated and expensive setup that is difficult to be realized [8]. A better alternative is to obtain the features from several 2D views from a few static cameras as suggested in [2][9].

In the current study, multiple views technique from static cameras is proposed to obtain the features of 3D objects. This study focuses on the recognition of the isolated objects using shape information. Due to the inherent loss of information in the 3D to 2D image projection process, an effective representation of 3D object properties using 2D images should be considered. 2D moments are used in the current study 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.

Recently, neural network becomes a popular choice for 3D object recognition. Compare to conventional 3D object recognition approaches, neural network normally provides a better generalization, robustness and parallel implementation paradigm properties [11]. Multilayered perceptron (MLP) network and hybrid multilayered

perceptron (HMLP) network [16] have been selected to perform the recognition task in the current study.

2. Image Acquisition and Features Extraction

Three cameras are used in the current study to obtain three 2D images of the 3D objects. The proposed camera-object setup is shown in Figure 1. The three cameras are placed at points A, B and C. A and B are located on the same horizontal, but differ 90^0 from each other. Point C is perpendicular to the turntable. Each object to be recognized must be placed in its stable condition at the centre of circular turntable, which can be rotated 360 degree. Illumination using controlled lighting condition is provided to have an object without shadow and reflection. 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, cameras at point A and B are proposed to be fixed at 45^0 from perpendicular view rather than at the *x*-*y* plane. This position is proposed to minimize the change of shape's description while the object is rotated. Camera at point C is fixed at the top of the object. Figure 2 shows how these three cameras are fixed.

After an object of interest is placed at the centre of the turntable, the 2D images of the object are acquired. Then, the object will be rotated 5^0 at a time and the three 2D images will be acquired again. Each time the object will be rotated at 5^0 until 360^0 is completed. Hence, for each object 72 2D image sets are obtained. These images are divided into two groups, 36 image sets for training data and 36 image sets for testing data. The acquired images at 0^0, 10^0, 20^0,…, 350^0 were used as the training set and the rest of the images (image at 5^0, 15^0, 25^0,…, 355^0) were used as the testing set. The training data set is used to build the 3D object model in the recognition stage.

The 2D captured images are then digitized and sent to the pre-processing and feature extraction stage. In the pre-processing stage, images will be automatically thresholded using iterative thresholding method [12][13]. Thresholding provides a good separation between object and background in several applications [14]. In feature extraction stage, Hu's moments [15] were used as the features for 3D modeling.

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:

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

$$U_{pq} = \sum_{i=1}^{M}\sum_{j=1}^{N} (i-\bar{i})^p (j-\bar{j})^q f(i,j) \tag{2}$$

where

$$\bar{i} = \frac{m_{10}}{m_{00}} \quad \text{and} \quad \bar{j} = \frac{m_{01}}{m_{00}} \tag{3}$$

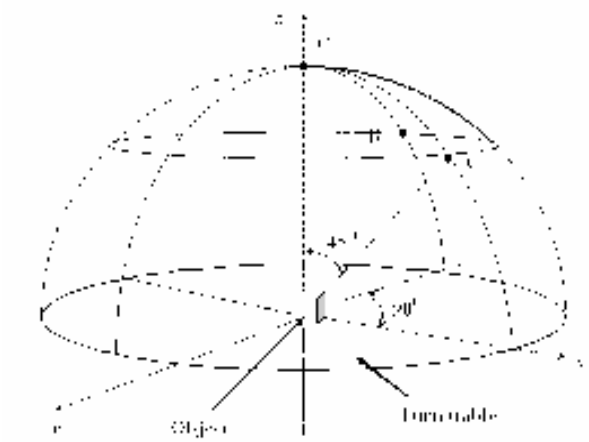

Figure 1- Image Acquisition Set-Up

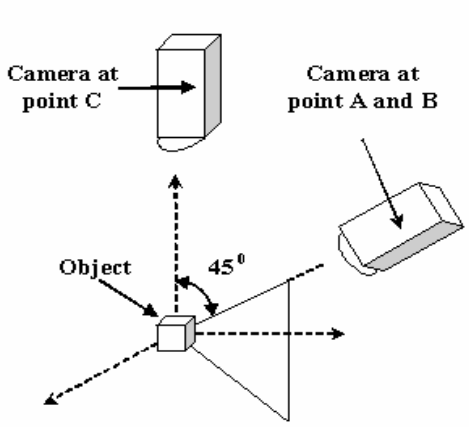

Figure 2. Camera Position For Point A, B And C

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:

$$\varphi_1 = \vartheta_{20} + \vartheta_{02} \tag{4}$$

$$\varphi_2 = (\vartheta_{20} - \vartheta_{02})^2 + 4\vartheta_{11}^2 \tag{5}$$

$$\varphi_3 = (\vartheta_{30} - 3\vartheta_{12})^2 + (3\vartheta_{21} - \vartheta_{03})^2 \tag{6}$$

$$\varphi_4 = (\vartheta_{30} + \vartheta_{312})^2 + (\vartheta_{21} + \vartheta_{03})^2 \tag{7}$$

$$\varphi_5 = (\vartheta_{30} - 3\vartheta_{12})(\vartheta_{30} + \vartheta_{12})[(\vartheta_{30} + \vartheta_{12})^2 - 3(\vartheta_{21} + \vartheta_{03})^2] + (3\vartheta_{21} - \vartheta_{03})(\vartheta_{21} + \vartheta_{03})[3(\vartheta_{30} + \vartheta_{12})^2 - (\vartheta_{21} + \vartheta_{03})^2] \tag{8}$$

$$\varphi_6 = (\vartheta_{20} - \vartheta_{02})[(\vartheta_{30} + \vartheta_{12})^2 - (\vartheta_{21} + \vartheta_{03})^2] + 4\vartheta_{11}(\vartheta_{30} + \vartheta_{12})(\vartheta_{21} + \vartheta_{03}) \tag{9}$$

$$\varphi_7 = (3\vartheta_{21} - \vartheta_{03})(\vartheta_{30} + \vartheta_{12})[(\vartheta_{30} + \vartheta_{12})^2 - 3(\vartheta_{21} + \vartheta_{03})^2] - (\vartheta_{30} - 3\vartheta_{12})(\vartheta_{21} + \vartheta_{03})\,[3(\vartheta_{30} + \vartheta_{12})^2 - (\vartheta_{21} + \vartheta_{03})^2] \tag{10}$$

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

$$\vartheta_{pq} = \frac{U_{pq}}{U_{00}^r} \tag{11}$$

and

$$r = [(p+q)/2]+1,\ \ p+q = 2,3,4.... \tag{12}$$

3. Recognition

The current study investigates the capability of Multilayered Perceptron MLP network and Hybrid Multilayerred Perceptron (HMLP) network for 3D object recognition. HMLP network is a MLP network with linear direct connections between input and output nodes. HMLP network with one hidden layer is shown in Figure 3. HMLP network with one hidden layer can be expressed by the following equation [16]:

$$\hat{y}_k(t) = \sum_{j=1}^{n_h} w_{jk}^2 F\left(\sum_{i=1}^{n_i} w_{ij}^1 v_i^0(t) + b_j^1\right) + \sum_{i=0}^{n_i} w_{ik}^\ell v_i^0(t); \quad \text{for } 1 \le k \le m \tag{13}$$

where w_{ij}^1, w_{jk}^2, w_{ik}^ℓ denote the weights between input and hidden layer, weights between hidden and output layer, and weights between input and output layer respectively. b_j^1 and v_i^0 denote the thresholds in hidden nodes and inputs that are supplied to the input layer respectively; n_i, m and n_h are the number of input nodes, output nodes and hidden nodes respectively. $F(\bullet)$ is an activation function that is normally be selected as sigmoidal function. In this paper, sigmoidal function was used for the activation function for both MLP and HMLP network.

The weights w_{jk}^2, w_{ik}^ℓ, w_{ij}^1 and thresholds b_j^1 are unknown and should be selected to minimise the prediction error defined as:

$$\varepsilon_k(t) = y_k(t) - \hat{y}_k(t) \tag{14}$$

where $y_k(t)$ and $\hat{y}_k(t)$ are the desired outputs and network outputs respectively.

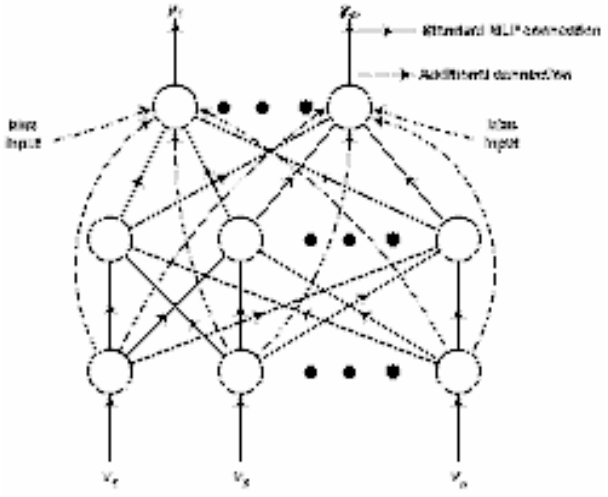

Figure 3. One-hidden layer HMLP network

In this study, the number of input nodes depends on the number of cameras used (3 cameras) while the number of outputs depends on the number of objects to be recognized. In the recognition step, output node which has the largest value is determined as 1. Otherwise, the node is considered as 0. In the current study, MLP network has been trained using Levenberg-Marquardt [18] and HMLP network using modified recursive prediction error (MRPE) algorithm [16]. Levenberg-Marquardt (LM) algorithm appears to be the fastest method for training moderate-sized MLP neural network and also has a very efficient implementation [19]. While HMLP network trained using MRPE algorithm has been proved to be better than MLP network for system identification applications [16]. Based on these arguments, these two network models are investigated in the current study.

4. Results and Discussion

Two types of objects have been used to test the performance of the proposed approach. Each type consists of eleven 3D objects. The first type, will be referred as Type 1 object, contains simple 3D shape like cylinder, box, trapezoid, sphere etc. The second type, will be referred as Type 2 object contains free-form objects. Figure 4 and 5 show these types of objects.

Based on some analysis on MLP and HMLP networks the following parameters were found to be the optimum values for MLP and HMLP networks, respectively. Both networks have the same number of input and output nodes. Both networks have 11 output nodes to represent 11 objects for both types of objects. Inputs to the networks were assigned as in Table 1 and Table 2 respectively. The optimum number of hidden nodes found to be 13 and 15 for HMLP and MLP networks, respectively. The LM algorithm was assigned to have training time step as t = 0.01. The designing parameters for MRPE were selected to be their typical values as $\alpha_m(0) = 0.00001$, $\alpha(t)_g = \alpha_m(t)(1-\alpha_m(t))$, $a = 0.01$, $b = 0.9$, $\lambda_0 = 0.99$, $\lambda(0) = 0.95$ and $\mathbf{P}(0) = 10000\,\mathbf{I}$. Matrix P(0) was updated using:

$$\mathbf{P}(i) = \mathbf{P}(i\text{-}1) + \mathbf{P}(i\text{-}1)/i \tag{14}$$

after every training epoch, where i is the number of current training epoch. Please refer to Mashor [16] for the definitions of these parameters.

Table 1 and Table 2 show the recognition performance of the proposed method for simple objects (Type 1) and free-form objects (Type 2), respectively. The inputs of the both networks for both cases were Hu's moment and the results were produced after 200 training epochs. Generally, the networks that used lower order moments achieved better recognition rate compared to the ones that used higher order moments. Higher order moments change rapidly for each rotation and normally more sensitive to noise compare to lower order moments [17]. Consequently, the features stability will decrease, thus reduce the recognition rate. Better recognition rate could be achieved by combining Hu's moments. Both network models

achieved 100% accuracy for both training and testing data sets when the first three Hu moments were used to train the network models for type 1 of 3D objects. For type 2 objects only HMLP network could achieve 100% accuracy for both training and testing data sets. However, the performances of both networks just differ slightly.

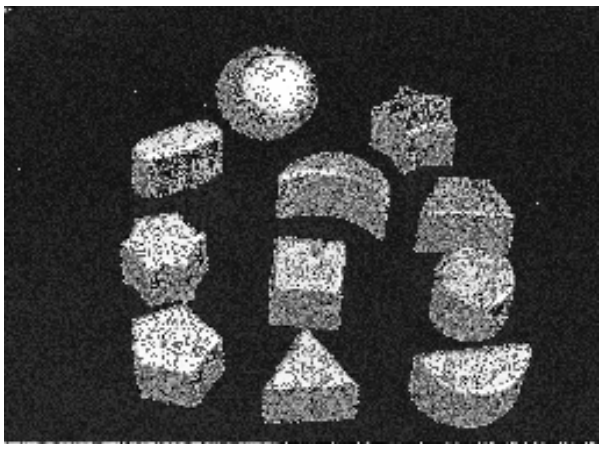

Figure 4. Type 1- simple 3D shape

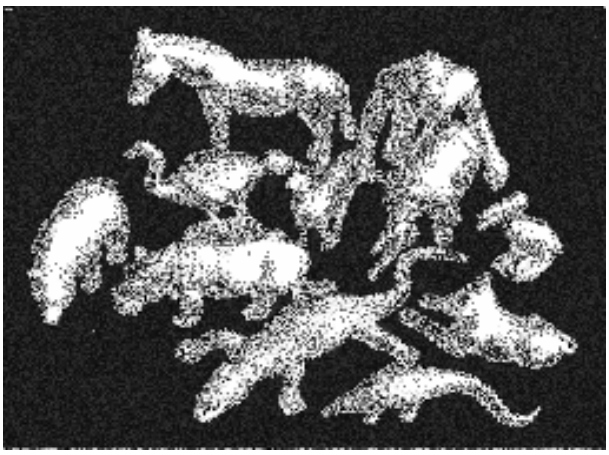

Figure 5. Type 2 - Free-Form Object

5. Conclusion

A 3D object recognition method is proposed using 2D multiple views technique and neural networks. MLP network trained using LM algorithm and HMLP network trained using MRPE algorithm were employed for recognition. The recognition results show that with some adaptation to multiple-views technique, Hu's moments are adequate to model the 3D objects. By using 2D moments the proposed method do not require complex features calculation as for 3D representation, thus reduces processing time in feature extraction stage. In addition, since Hu's moments are global features, it can be applied arbitrarily to any 3D objects. Both network models produce excellent recognition rate. 100% recognition rates were obtained for the type 1 objects for both training and testing data sets. For type 2 objects only HMLP network could achieve 100% accuracy for both training and testing data sets. Both networks produced approximately the same performance. However, HMLP network is slightly more efficient than the MLP network where for both types of 3D objects the network requires less hidden nodes.

Table 1: Recognition performance for 3D object type 1 using Hu's moment

Hu's moment	MLP-LM		HMLP-MRPE	
	Training Accuracy (%)	*Testing Accuracy (%)*	*Training Accuracy (%)*	*Testing Accuracy (%)*
φ_1	100	100	100	98.99
φ_2	90.15	89.39	94.70	90.15
φ_3	99.75	97.73	99.75	98.23
φ_4	57.83	58.08	72.22	59.34
φ_5	66.67	62.88	81.31	69.70
φ_6	72.98	70.71	82.07	68.43
φ_7	56.82	55.30	71.97	56.06
$\varphi_1 + \varphi_2$	100.00	100.00	99.24	98.99
$\varphi_1 + \varphi_2 + \varphi_3$	100.00	100.00	100	100

Table 2: Recognition performance for 3D object type 2 using Hu's moment

Hu's moment	MLP-LM		HMLP-MRPE	
	Training Accuracy (%)	Testing Accuracy (%)	Training Accuracy (%)	Testing Accuracy (%)
φ_1	95.71	95.71	95.96	96.96
φ_2	90.91	91.41	93.29	92.93
φ_3	84.60	83.59	93.18	89.39
φ_4	78.79	74.75	84.45	79.80
φ_5	75.00	74.75	84.09	80.05
φ_6	73.99	70.96	86.36	79.29
φ_7	45.71	41.67	74.49	60.10
$\varphi_1 + \varphi_2$	99.49	99.75	100	100
$\varphi_1 + \varphi_2 + \varphi_3$	100	99.75	100	100

References

[1] Shirai Y., 1987, *Three-Dimensional Computer Vision*, New York: Springer-Verlag,.

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

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

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

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

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

[7] Schiele B., and Crowley J. L., 1998, Transinformation for Active Object Recognition, *InProceeding of the 6th International Conference on Computer Vision* 249-254.

[8] Selinger A., and Nelson R. C., 2001, Appearance-Based Object Recognition Using Multiple Views, *In Proceedings of the 2001 IEEE Conference on Computer Vision and Pattern Recognition,* 1: 905-911.

[9] Mao J., Flynn P. J., and Jain A. K., 1995, Integration of Multiple Feature Groups and Multiple Views into a 3D Object Recognition System, *Computer Vision and Image Understanding,* 62(3): 309-325.

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

[11] 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.

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

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

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

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

[16] Mashor M. Y., 2000, Hybrid Multilayered Perceptron Networks, *International Journal of System and Science*, 31(6):171-185.

[17] Prokop R. J., and 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.

[18] Hagan M. T., and Menhaj M., 1994, Training Feedforward Networks with the Marquardt Algorithm. *IEEE Transactions on Neural Networks*. 5(6): 989-993.

[19] Demuth H., and Beale M., 2001, Neural Network Toolbox for Use with Matlab User's Guide Ver. 4. MA: The MathWorks, Inc.