International Journal of Advanced Culture Technology Vol.6 No.3 142-150 (2018)
DOI https://doi.org /10.17703//IJACT2018.6.3.142

IJACT 18-9-16

Object Recognition Using the Edge Orientation Histogram and Improved Multi-Layer Neural Network

Myung-A Kang

Division of Computer Science Engineering, Gwangju University, Korea
makang@gwangju.ac.kr

Abstract

This paper describes the algorithm that lowers the dimension, maintains the object recognition and significantly reduces the eigenspace configuration time by combining the edge orientation histogram and principle component analysis. By using the detected object region as a recognition input image, in this paper the object recognition method combined with principle component analysis and the multi-layer network which is one of the intelligent classification was suggested and its performance was evaluated. As a pre-processing algorithm of input object image, this method computes the eigenspace through principle component analysis and expresses the training images with it as a fundamental vector. Each image takes the set of weights for the fundamental vector as a feature vector and it reduces the dimension of image at the same time, and then the object recognition is performed by inputting the multi-layer neural network.

Keywords: *PCA, MLNN, EOH, Eigenspace, Singular Value Decomposition*

1. Introduction

In the environment that the industry is developing rapidly, the development of automated object recognition system has occupied an important position to replace the work force in various fields including production line, medical and military. Especially, the study to recognize 3-D objects such as parts assembly and inspection and military equipment has been separated into the field called Computer Vision or Robot Vision, so by determining what objects are detected on the computer and providing the ability to perform appropriate tasks as programmed using the determined results, many researches to grant the visual ability that person can do to computer have been proceeding [1]. Generally, to recognize the object using computer, the method to recognize the object based on saving many information about the object is used. As a recognition method based on the model, this method recognizes the type, location and position of object existed in the image by comparing features extracted from input images and model features by saving them in the database through extracting the physical features of model that is to recognize and features that express its geometric correlation [2][3]. The object recognition can be divided into 2D object recognition and 3D object recognition largely depending on the target object. Since 2D object recognition is only possible to recognize the section of object, the objects are mostly with flat objects [4][5]. In contrast, 3D object recognition is the method, which is possible to recognize every views of object. Since it needs to recognize every view, this is a problem which is not solved well compared to 2D object recognition and it is approached in various method [6].

Manuscript received: July 13, 2018 / revised: July 25, 2018 / Accepted: August 3, 2018
Corresponding Author: makang@gwangju.ac.kr
Author's affiliation
Division of Computer Science Engineering, Gwangju University, Korea

In this paper, after converting input images into silhouettes images through preprocessing, input images are projected to the lower dimensional vector space, that is parametric eigenspace which can express the features of object appearance by a statistical technique called PCA (Principal Component Analysis) for features extracted through EOH(Edge Orientation Histogram). Each object is presented as a trajectory of consecutive points sequentially, and the object is recognized by comparing the trajectory pre-learned and trajectory of input image. In this paper, features using silhouette image and EOH were used to improve configuration speed of eigenspace. In addition, in PCA object recognition the point to point method that compared the components and distance of representing values resulted from reflecting the experimental image to eigenspace after making the representing values of each object inside the model image was used a lot. But since this method calculated the simple distance only, when recognizing in real time, the decline of recognition was happened due to lots of errors. Therefore, in this paper the recognition method that combines with the multi-layer neural network by computing the small number of eigen object feature vector through PCA is suggested to improve the recognition rate.

2. Preprocessing

2.1 Background Removal

The image sequence obtained through camera is the one that obtained in a simple background and in the image obtained from general environment, it includes a lot of object background that doesn't need for object recognition. However, since the object region front view is needed for object recognition, it is necessary to separate the background and body areas first and to do this the background model should be created first. However, because the brightness of the light is not constant and changes often, all are not equal and it is difficult to obtain stable background model even if the same background is taken with the same camera in a certain period of time.

In this paper, after obtaining the background image for a certain period of time by measuring the changes in the brightness of the background due to light changes and considering the time factor , pixel value was determined when the light was brightest and pixel value was determined when the light was darkest by analyzing each pixel existing in the following image region . The difference between these two pixel value is the threshold of brightness that can appear with the change of light. By using these 3 factors, the Background Model will be configured. Such information is shown in equations (1-4) [7].

$$BM = \{P_{\max}(x), P_{\min}(x), D(x)\}_{x \in R} \tag{1}$$

$$P_{\max}(x) = MaxI_t(x), (1 \le t \le T_i) \tag{2}$$

$$P_{\max}(x) = MinI_t(x), (1 \le t \le T_i) \tag{3}$$

$$D(x) = P_{\max}(x) - P_{\min}(x) \tag{4}$$

Once the background model is made, the binary image $B(x)$ will have the maximum value 255 if the difference value which is obtained from differential operation between the brightest pixel value of $p_{\max}(x)$ input image $I(x)$ and the darkest pixel value $P_{\min}(x)$ is larger than threshold $D(x)$ and otherwise it will have the minimum value 0.

$$B(x) = \begin{cases} 255 & if \mid P_{\max}(x) - I(x) \mid or \mid P_{\min}(x) - I(x) \mid > D(x) \\ 0 & otherwise \end{cases} \tag{5}$$

The equation (5) becomes the basis for separating the region that has the difference in the motion of object's pose changes ignoring the brightness differences that can be caused by the light. In the binary image obtained from the result of equation (2), due to changes in lighting which is off from the threshold of brightness value set from background model, it is classified as foreground region even though it is background, so small point of 1 pixel can be included. Therefore, to remove this noisy, Morphological operation was used. A single erosion operation was performed and the dilation operator was used to restore because the reduction of object was occurred at this time.

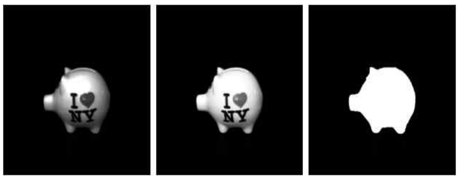

Figure 1. (a) The minimum brightness of the image (b) The maximum brightness of the image

(c) Silhouette image extracted using the background model parameters of input image.

2.2 EOH Creation

If the eigenspace is configured using normalized images obtained through the method described in the previous section, computations for 76,800 dimensions are needed because the size of object image is 320 * 240. In this paper, by using the edge orientation histogram to reduce the size of dimension, a method to reduce the feature vector size, that is the size of dimension into 36 was proposed. Edge has to be extracted first to create EOH. Edge is sought using equation(6) in this thesis. If EOH is created using edge direction having extracted edge from image, this can be used as characteristic information of image.

$$G_x = Sobel(I_{ROI}), G_y = Sobel(I_{ROI})$$
$$m_{i,j} = \sqrt{G_x(i,j)^2 + G_y(i,j)^2} \quad (6)$$

Edge image obtained with critical value is created to EOH with direction of 36 levels (equation 7).

$$\theta_{i,j} = \arctan(G_x(i,j) / G_y(i,j)) \quad (7)$$

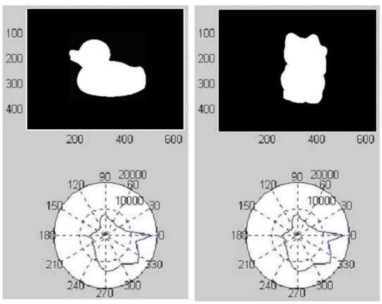

Figure 1. Images which create orientation histogram

3. Object Recognition

There are many difficulties to recognize as a same form if the direction of object is rotating when using a single camera. In this paper, to recognize the object as the same form even though the direction of object is rotating, the solution for object orientation problem is suggested. To recognize the rotating object as a same form, after obtaining the image by rotating the object to 5° each and then using EOH obtained through the method explained in section 2 the object model database was configured as shown in Figure 3.

Figure 3. Extraction of EOH in binary image and DB configuration.

In Figure 3, x_N in $x_N^{(R)}$ represents the number of objects and is the rotation, and $EH_N^{(R)}$ is a set of EOH.

3.1 Space Generation using PCA

Principal Component Analysis can reduce the high dimensional input data set to low dimensional meaningful data set. In case of object image data, since one object has lots of frame numbers which configure the rotating motion, and is relatively difficult to extract the features, it is necessary to apply the methods that have fast recognition speed and are possible to have effective feature extraction. Therefore, the method to express as low dimensional vector using EOH of consecutive rotating motion extracted from section 2 will be described.

Let the feature vector obtained from previous section as and it can be expressed as an equation (8). To calculate the eigenspace of this vector, first determine the average vector of all the feature vectors and then determine the difference of each feature vector. The average vector and new feature set can be expressed as equations (9-10).

$$x = \{x_1, x_2, x_3, ..., x_N\}^T \tag{8}$$

$$c = (\frac{1}{N}) \sum_{i=1}^{N} x_i \tag{9}$$

$$X = [x_1 - c, x_2 - c, ..., x_N - c]^T \tag{10}$$

Next, to determine the eigenspace that satisfies the equation (11), the eigenvalue λ and eigenvector e for covariance matrix Q can be obtained.

$$Q \stackrel{\Delta}{=} XX^T \quad , \quad \lambda_i e_i = Qe_i \tag{11}$$

To determine the eigenvector, by using Singular Value Decomposition instead of using eigenvalue decomposition the eigenvector for covariance matrix of feature set X can be obtained easily. The feature set X which subtracts the average vector c was projected to the eigenspace obtained like this using an equation (12).

$$f_j = [e_1, e_2, e_3, ..., e_k]^T (x_n - c) \tag{12}$$

The low dimension vector space obtained like this is called eigenspace. As described earlier, the main purpose of PCA is to reduce the dimension and summarize the data through inducing several principal component vectors. Therefore, in order to explain the entire object images adequately, the number of Principal Component that needs to have should be determined. The equation (13) is the method to select the number that we need.

$$\left(\sum_{i=1}^{k}\lambda_i / p\right)\times 100 \geq 80 \qquad (13)$$

Where, let is λ_i is ith eigenvalue and p is total number of eigenvalues, among total variance the rate that the principal component C_i that can be explained is λ_i / p . In other words, the number will be selected when the cumulative ratio that can be explained by k number of principal components is more than 87%. Figure 4. shows the cumulative contribution of principal component depending on the number of eigenvalue obtained from feature set of objects used in experiment.

Eigenvalue	Rate(%)
1	44.5
2	59.6
3	71.1
4	77.6
5	82
6	84.1
7	85.1
8	86
9	86.8
10	87.3
11	87.8
12	88.1
13	88.7
14	89
15	89.3
16	89.7

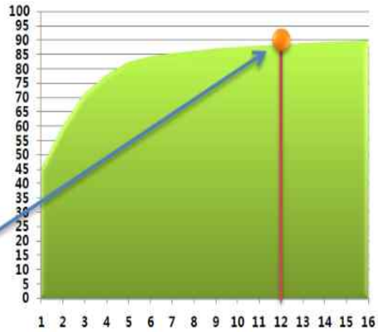

Figure 4. Cumulative contribution depending on the number of eigenvalue

3.2 Multi-layer Neural Network using the Error Back-Propagation

In the object recognition method, the feature vector of object image input is performed by finding the object inside the model image having the most close feature vector of Euclidean distance. But, even though the real object image in the object image space succeeded in the matching, there were incorrect matching that recognized the different object image was occurred. To solve these problems, instead of utilizing the existing Euclidean distance the classified method by utilizing the neural network is suggested. The weight of eigenspace obtained through PCA is used as a multi-layer neural network, and in this paper by learning the weight of neural network using the error back-propagation algorithm, it performs as the classifier. The structure of neural network designed in this study is shown in Figure 5.

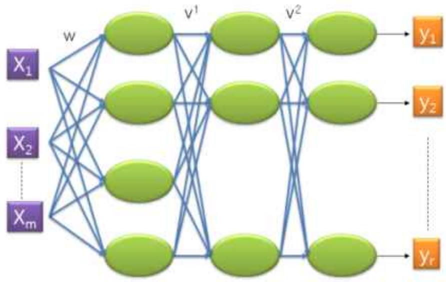

Figure 5. The structure of multi-layer neural network

The weight learning between input layer and the first hidden layer in Figure 5. can be found through the repetitive learning by equation (14-20) for the following error back-propagation algorithm [8].

$$w^{k+1} = w^k - \eta \frac{\partial E^k}{\partial w_{ij}} \tag{14}$$

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial h_i} \frac{\partial h_i}{\partial w_{ij}} \tag{15}$$

$$\frac{\partial E}{\partial h_i} = \frac{\partial E}{\partial s_i} \frac{\partial s_i}{\partial h_i} \tag{16}$$

$$E = \frac{1}{2} \sum_{i=1}^{r} (t_l - y_l)^2 \tag{17}$$

$$s_i = \frac{1}{1 + e^{(-\lambda h_i)}} \tag{18}$$

$$h_i = \sum_{l=1}^{q} w_{il} x_l + w_{iq_{+1}} \tag{19}$$

$$y_i = \sum_{l=1}^{r} v_{il} s_l + v_{ir_{+1}} \tag{20}$$

In here, k is the repetition number of learning, i, j are the number of input layer and hidden layer node, respectively, η is the convergence constant, s is the activation function, h is intermediate total, y is input value, t is the target value, m is the number of input node, r is the output node number, and v_{ij} is the learning weight between hidden layer and input layer. The weight learning between hidden layer and output layer can be obtained by equation (21-22).

$$v^{k+1} = v^k - \eta \frac{\partial E^k}{\partial v_{ij}} \tag{21}$$

$$\frac{\partial E}{\partial v_{ij}} = \frac{\partial E}{\partial y_i} \frac{\partial y_i}{\partial v_{ij}} \tag{22}$$

From two learning weight equation (14) when the optimum value is obtained with the equation (22) by repetitive learning, the learning is done and it is used as a classifier by using it.

4. Experiment Result

The algorithm of object recognition system suggested in this paper extracted the object using the sihouette technique from the background image obtained through the camera, and then performed the object recognition by applying the multi-layer neural network after calculating the proper vector through PCA after normalization of image size and noise removal. Fig.1 shows the algorithm flow chart suggested in this paper

4.1 Recognition result using MLNN

The number of objects used in the experiment was 25 and one object was obtained by camera from the image rotating 5° each as shown in Figure 6. A set of rotated objects by 0° to 360° for each object is called an object of one image. The main problem in recognition using principal component analysis is that it has some difficulties to apply in real-time because the time that configures the eigenspace and speed that performs the object recognition are not fast. For that reason, to configure the eigenspace for object recognition, the object image taken by 640 * 480 was converted into 320 * 240 through size normalization.

And, in this study silhouette image was created without using every object image data. Only with silhouette image of object, it has much of the object features so these features were used as input data for EOH. Also, the object silhouette image had 320 * 240 = 76,800 data and only 36 feature data among these data were extracted using EOH. In other words, in the preprocess 76,800 data was reduced to 36 data.

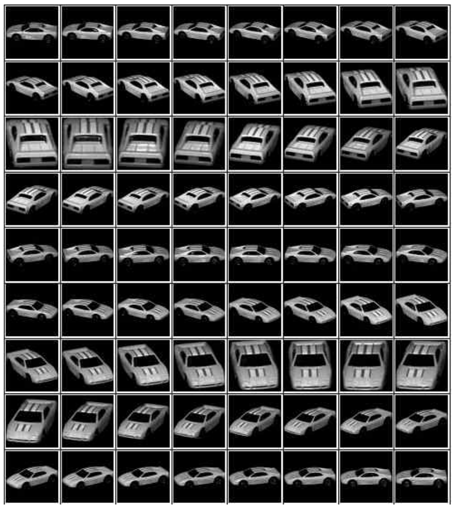

Figure 6. Set image of rotation model of a car object

To verify the superiority of method suggested in this paper, after learning the weight of neural network by inputting the weight from the PCA result into the neural network, the results of applying the recognition performance to Euclidean distance measure method and Mahananobis distance measure were shown in Table 1. For the suggested neural network, the number of input node was 400 and 10 hidden layers were used. The initial weight was 0.01 and the allowable limit of error was 0.005. When using the suggested method as shown in Table 1, it showed quite improved result in matching failure than the existing suggested Euclidean and Mahananobis methods and it also showed the superiority in the incorrect matching than the existing method.

Table 1. Matching success rate by matching methods

Matching Method	Matching Failure	Incorrect Matching	Matching Success
Euclidean	5.5%	10.0%	84.5%
Mahananobis	4.5%	7.5%	88.0%
Proposed MLNN	1.5%	2.0%	96.5%

4.2 Recognition rate by learning rate

In this paper, to know the difference of recognition rate according to the optimum learning rate, the experiment was carried out as shown in Table 2. In the experimental environment 1 in Table 2 the recognition for 200 object images with 500 object image learning model was studied. In the experimental environment 2, the object image learning model was obtained in the same way as in the experimental environment 1 but not like in the experimental 1 the recognition with the learning model that was not obtained was studied in 100 images. In this experiment, the critical value of output error was used as 0.025.

Table 2. The Experimental environment by the number of object image used in learning

Number of image / Environment	Learning Image	Recognition Image
Experimental environment 1	500	200
Experimental environment 2	500	100

The following Table 3. is the result of each experiment for Table 1.

Table 3. Recognition rate by learning rate in the experimental environment

Experimental Environment 1			Experimental Environment 2		
Learning rate	The number of learning	recognition rate (%)	Learning rate	The number of learning	recognition rate (%)
0.01	3360	95.32	0.01	3360	90.68
0.1	372	93.54	0.1	372	90.68
0.2	149	97.20	0.2	149	90.68
0.3	101	95.35	0.3	101	90.68
0.4	68	95.36	0.4	68	90.68
0.5	70	93.50	0.5	70	90.68
0.6	47	93.50	0.6	47	90.68
0.7	59	93.52	0.7	59	90.68
0.8	43	91.65	0.8	43	90.68
0.9	27	91.65	0.9	27	90.68
1.0	28	93.50	1.0	28	90.68

As the learning rate increases for every experiment, the number of learning is reduced. In the recognition rate by the learning rate in Table 2, when the learning rate is learned by 0.2, the recognition rate was the best and in the Table 3 there are no differences in the recognition rate by learning rate. In addition, the recognition rate is better when the object image used in the learning was included in the input image.

5. Conclusion

In this paper, by combining EOH and PCA the feature data obtained through EOH from silhouette image which is the information of object form instead of using principal component analysis using existing object image was extracted and by utilizing this data in PCA, the method that lowers the dimension and maintains the recognition rate up to 90% and reduces the eigenspace configuration time significantly was suggested. Combining the PCA, MLNN algorithms that were used for object recognition individually the performance was compared with the existing PCA matching method. With the PCA as a pre-processing, by calculating the feature vector and utilizing this as input data in the neural network, the method that could lower the dimension and enhance the performance was suggested. As a result, the method suggested by comparing with the existing method showed the improvement in the recognition rate as 96.5%. In addition, by studying the changes of recognition rate according to the learning rate in various environments, the most optimum value of learning rate was calculated. It is necessary to study more on the optimization process for pre-processing and parameters of neural network in the future.

6. Acknowledgement

This Study was conducted by research funds from Gwangju University in 2015.

References

[1] P. J. Besl and R. C. Jain. Three-Dimensional Object Recognition. ACM computing surveys, vol. 17, no. 1, pp. 75-145, March 1985.

[2] Jong-Min Kim, and Maeng-Kyu Song, Three Dimensional Gesture Recognition Using PCA of Stereo Images and Modified Matching Algorithm. IEEE Fuzzy Systems and Knowledge Discovery Vol.4, pp.116-120, Oct. 2008.

[3] Rafael C. Gonzalez, Richard E. Woods. Digital Image Processing. Addison Wesley, 1992.

[4] Rafael C.Gonzalez, Richard E.Woods. Digital Image Processing 2/E. Prentice Hall, pp.519-532, 2001.

[5] JY. Tak, E. Hwang. Indexing and Matching Scheme for Recognizing 3D Objects from Single 2D Image. International Conference on Internet and Multimedia Systems and Applications (IMSA'09), pp.60-67, 2009.

[6] Ragael C. Gonzalez and Richard E. Woods. Digital Image Processing. Prentice Hall, 2002.

[7] Ismail Haritaoglu, David Harwood and Larry S. Davis. W4: Who? When? Where? What? A Real-time System for Detecting and Tracking People. Third Face and Gesture Recognition Conference, pp.222-227, 1998.

[8] A. D. Kulkarni. Computer Vision and Fuzzy-Neural Systems. Prentice-Hall, 2001.