3-D Object Recognition Using 2-D Poses Processed by CNNs and a GRNN

Övünç Polat and Vedat Tavşanoğlu

Electronics and Communications Engineering Department
Yıldız Technical University
Besiktas, Istanbul 34349, Turkey
{opolat, tavsanav}@yildiz.edu.tr

Abstract. This paper presents a novel approach to automatically recognize objects. The system used is a new model that contains two blocks; one for extracting direction and pixel features from object images using Cellular Neural Networks (CNN), and the other for classification of objects using a General Regression Neural Network (GRNN). A data set consisting of different properties of 10 different objects is prepared by CNN.

1 Introduction

Among the artificial neural network(ANN) approaches developed for pixel-based and feature-based object recognition are feed-forward, Hopfield, and fuzzy ANNs. In general, the types of features that are used for object recognition differ from the features used by the neural-based segmentation approaches [1].

In this work, object recognition is achieved by comparing extracted features of the object with the features available in a reference set. To this end, each object is placed on a turntable which is rotated through 360 degrees and poses are taken with a fixed camera. Images of the objects are taken at pose intervals of 5 degrees. This corresponds to 71 poses per object, in total 710 poses for 10 objects. By using a CNN, a feature space consisting of poses of the image taken from 71 directions is constituted and this procedure is repeated for all object images. Apart of the constituted feature vectors are used to simulate the GRNN, and the remaining are used as test patterns.

1.1 The Cellular Neural Network (CNN)

The CNN used in this paper is defined by the following equations:

$$\dot{x}_{ij} = -x_{ij} + \sum_{C(k,l)\in S_r(i,j)} A(i,j;k,l)\,y_{kl} + \sum_{C(k,l)\in S_r(i,j)} B(i,j;k,l)\,u_{kl} + z_{ij} \tag{1}$$

$$y_{ij} = f(x_{ij}) = \frac{1}{2}\left(\left|x_{ij}+1\right| - \left|x_{ij}-1\right|\right) \tag{2}$$

F.A. Savacı (Ed.): TAINN 2005, LNAI 3949, pp. 219 – 226, 2006.

where:
x_{ij} is the state of cell C(i,j)
u_{ij} is the input of cell C(i,j)
y_{ij} is the output
A is the feedback template
B is the control or input template
z is the threshold of cell C(i,j)

In this paper the A and B templates used are of dimension 3x3 [2] and here eqn.(1) is solved using the rastering algorithm [3].

In the literature there are a number of papers that used feature extraction for character recognition. In [4] CNN-Gabor filters were used to extract orientation information from hand-written characters. In [5] various templates were used and the decision was taken on a basis of how many pixels remain black after processing and also the position of these black pixels were taken into account.

1.2 Artificial Neural Network

In this work, both multilayer perceptron(MLP) and radial basis function neural networks (RBFNN) are employed. Among all results, the best results are obtained by using the general regression neural network(GRNN) which is a kind of RBFNN. The GRNN model developed by Specht [6] is a powerful regression tool with a dynamic network structure whose network training speed is extremely fast. Due to the simplicity of the network structure and its implementation, it has been widely applied to a variety of fields including image processing. Specht [7] addressed the basic concept of inclusion of clustering techniques in the GRNN model.

2 The Realization of the System for Object Recognition

Fig. 1 shows the system for object recognition using a CNN and a GRNN. The system used is a new model that contains two blocks: one for extracting direction and pixel features from object images using a CNN, and the other for classification of objects using a GRNN.

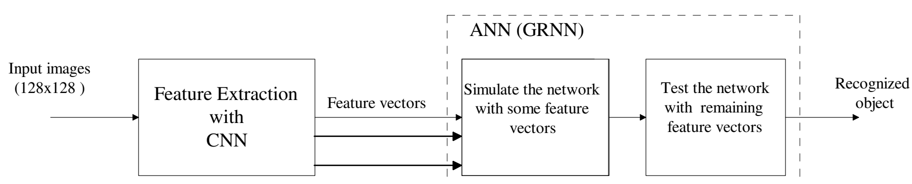

Fig. 1. System for object recognition using a CNN and a GRNN

2.1 Feature Extraction from the Object Images

In this work, a CNN is used for feature extraction from the input images obtained from the object to be recognized. Edge detection templates are applied to the input images which are converted to gray level and the number of pixels between the levels 0.4-0.8 of the image obtained is determined as a feature. Then the “Horizontal skeleton from the right” template is applied and this time the number of pixels equal to 1 and between 0.4-0.8 are determined as two different features. Also “Horizontal skeleton from the left”, ”Vertical skeleton from the bottom” and “Vertical skeleton from the top” templates are applied to the edge detected image and the number of pixels between 0.4-0.8 are determined as three different features. Thus, a feature vector consisting of six different features of the input images is obtained. Now the application of horizontal and vertical templates yields the components of the object images at that direction. The templates used for feature extraction from the input images are given as follows :

$$A=\begin{bmatrix}0&0&0\\0&1&0\\0&0&0\end{bmatrix},B=\begin{bmatrix}-1&-1&-1\\-1&8&-1\\-1&-1&-1\end{bmatrix},\text{ z=-1,}\qquad\text{(Edge detection)}$$

$$A=\begin{bmatrix}0&0&0\\0&3&0\\0&0&0\end{bmatrix},B=\begin{bmatrix}0.5&0&0.125\\-0.5&0.5&-0.5\\0.5&0&0.125\end{bmatrix},\text{ z=-1}\qquad\text{(Horizontal skeleton from the left)}$$

$$A=\begin{bmatrix}0&0&0\\0&3&0\\0&0&0\end{bmatrix},B=\begin{bmatrix}0.125&0&0.5\\-0.5&0.5&-0.5\\0.125&0&0.5.\end{bmatrix},\text{ z=-1}\qquad\text{(Horizontal Skeleton from the right)}$$

$$A=\begin{bmatrix}0&0&0\\0&3&0\\0&0&0\end{bmatrix},B=\begin{bmatrix}0.5&-0.5&0.5\\0&0.5&0\\0.125&-0.5&0.125\end{bmatrix},\text{ z=-1}\qquad\text{(Vertical skeleton from the top)}$$

$$A=\begin{bmatrix}0&0&0\\0&3&0\\0&0&0\end{bmatrix},B=\begin{bmatrix}0.125&-0.5&0.125\\0&0.5&0\\0.5&-0.5&0.5\end{bmatrix},\text{ z=-1}\qquad\text{(Vertical skeleton from the bottom)}$$

2.2 Recognition with GRNN

The feature vector, obtained by extracting features from ten object images by CNNs, is used to simulate and test the GRNN. In the simulation of GRNN, different number of poses from the original data set with different rotation intervals are selected as the reference poses. The remaining images in the data set are used to test the network. This procedure is shown in Figure 2.

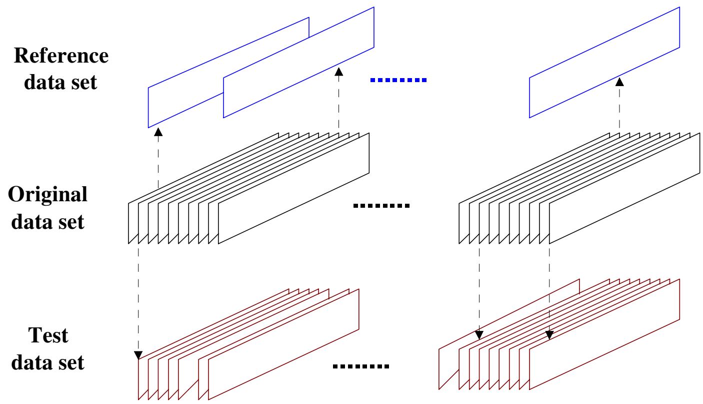

Fig. 2 Grouping of the database

3 Simulation and Results

For the experiments we are currently using images of the Columbia image database. Columbia Object Image Library (COIL-100) is a database of color images of 100 objects. We selected the 10 objects from the data set shown in Figure 3. The objects

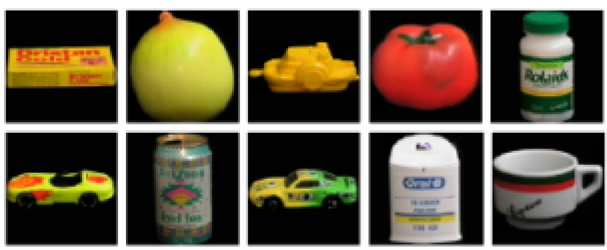

Fig. 3. 10 objects used the simulate recognition system

Fig. 4. The image sequence of object-6 in database with rotations 0^o to 25^o

were placed on a motorized turntable against a black background. The turntable was rotated through 360 degrees to vary object pose with respect to a fixed color camera. Images of the objects were taken at pose intervals of 5 degrees. This corresponds to 72 poses per object. The images were size normalized.[8]. We take image sequences of 71 images of each of the selected objects. Figure 4 shows the frames with rotations 0^o to 25^o in the object-6 data set from COIL100.

The images obtained by applying "edge detection", "horizontal skeleton from the right" and "vertical skeleton from the bottom" templates to the pose of the object taken at 0^o rotation is given in Figure 5.

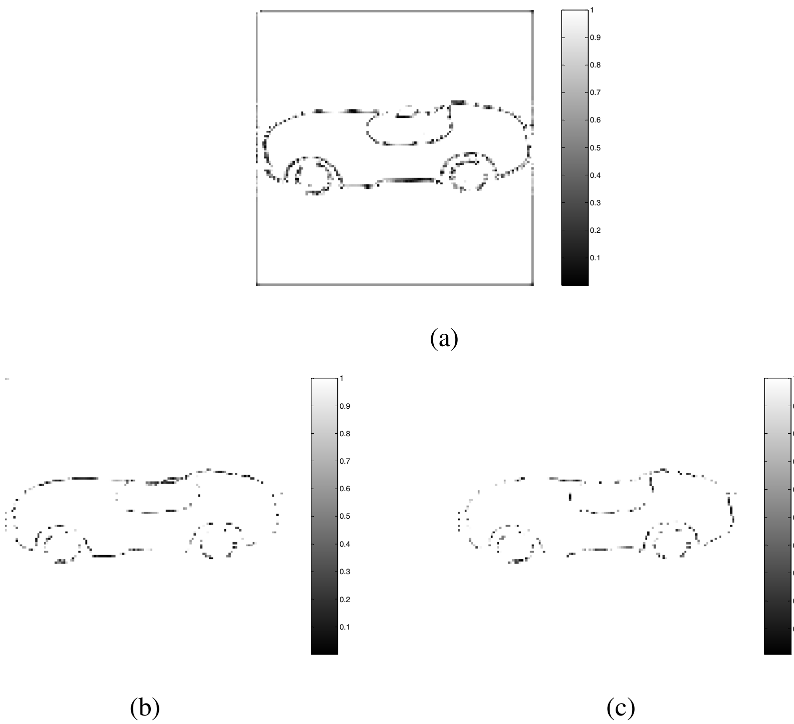

Fig. 5. Extracting features from object-6 with CNN a) Edge detection, b) Horizontal Skeleton from the right, c)Vertical skeleton from the bottom

As described above, different number of poses from the original data set with different rotation intervals are selected as reference poses. In this step, a code for every object is given to the GRNN target values. In Table 1, 10,18 and 35 poses from the original data set with 35^o, 20^o and 10^o rotation intervals, respectively are selected as the reference poses. The remaining images in the data set are used to test the network.

In Table 2, the accuracy of the test results obtained by using 35 poses taken with 10^o rotation intervals for the ten objects considered are given.

Table 1. Number of reference images and corresponding recognition rates

Number of reference images	Recognition rate of the test set
10 poses from original data set with 35° rotation intervals are selected as the reference poses. The remaining 61 images in the data set are used to test the network.	% 82,13
18 poses from original data set with 20° rotation intervals are selected as the reference poses. The remaining 53 images in the data set are used to test the network.	% 86,42
35 poses from original data set with 10° rotation intervals are selected as the reference poses. The remaining 36 images in the data set are used to test the network.	% 89,17

Table 2. Recognition rates of objects

	Obj1	Obj2	Obj3	Obj4	Obj5	Obj6	Obj7	Obj8	Obj9	Obj10	Average
Recog. rate	% 97,2	% 100	% 97,2	% 100	%80,6	%72,2	% 100	% 55,5	% 97,2	% 91,6	% 89,17

As can be seen from Table 2, recognition rate is low only for the ones that are similar. Because of the similarities between object 6 and 8, recognition rate for these objects decrease with the decrease in the number of reference poses. In order to increase the recognition rate, new features are added to the feature vector and the GRNN is again simulated and tested. In the processing with CNN, gray level images are converted to binary images and edge detection process is carried out.

Table 3. Number of reference images and corresponding recognition rates after adding new features

Number of reference images	Recognition rate of the test set
10 poses from original data set with 35° rotation intervals are selected as the reference poses. The remaining 61 images in the data set are used to test the network.	% 87,21
18 poses from original data set with 20° rotation intervals are selected as the reference poses. The remaining 53 images in the data set are used to test the network.	% 91,32
35 poses from original data set with 10° rotation intervals are selected as the reference poses. The remaining 36 images in the data set are used to test the network.	% 94,17

Then the “Horizontal skeleton from the right” and “Vertical skeleton from the bottom” templates are applied to the edge detected images and this time the number of pixels between 0.1-0.3 are determined as two different features. Thus, feature vector consisting eight different features of the input images is obtained. The test results of the simulated network with new features vector is given in Table 3.

Also in Table 4 the accuracy of the test results obtained by using 35 poses taken with 10^o rotation intervals for the ten objects considered are given.

Table 4. Recognition rates of objects after adding new features

	Obj1	Obj2	Obj3	Obj4	Obj5	Obj6	Obj7	Obj8	Obj9	Obj10	Average
Recog. rate	%97,2	%100	%100	%100	%100	% 80,56	%100	% 75	%91,67	% 97,2	% 94,17

As can be seen from Tables 3 and 4 the recognition rate in the test process increases after adding new features to the features vector. Also recognition rate in the test process varies with different spread(*width of the kernel in GRNN)* values. Depending on the number of reference poses, spread value can be determined in order to increase recognition rate.

4 Conclusion

In this work, a model is designed for 3D object recognition using 2D poses processed by CNNs where the recognition is achieved by a GRNN. The application is carried out for 10 objects and high recognition rate is obtained. When the number of objects is increased or similar objects are chosen for recognition, extraction of diagonal or convex components with CNNs procedure or choosing pixels between different intervals as input can help to increase the recognition rate.

References

1. M. Egmont-Petersen, D. de Ridder, H. Handels. : Image processing with neural networks - a review. The Journal of the Pattern Recognition Society, Vol. 35, No. 10, pp. 2279-2301, (2002).
2. Chua, L.O., Roska,T.: Cellular Neural Networks & Visual Computing. Cambridge University Press (2002).
3. Saatci,E., Tavsanoglu,V. : On the optimal choice of integration time-step for raster simulation of a CNN for gray level image processing. Circuits and Systems, ISCAS (2002). IEEE International Symposium on, Volume:1, Pages:I-625 - I-628, (26-29 May 2002).
4. Saatci, E., Tavsanoglu, V.: Multiscale handwritten character recognition using CNN image filters. Neural Networks, 2002. IJCNN '02. Proceedings of the 2002 International Joint Conference on , Volume: 3 , Pages:2044 – 2048,(12-17 May 2002).
5. Salermo, M., Sargeni, F., Bonaiuto, V., Favero, F.M. : Multifont character recognition by 9/spl times/9 DPCNN board. Circuits and Systems, 1997. Proceedings of the 40th Midwest Symposium on , Volume: 2 , Pages:1338 – 1341,(3-6 Aug. 1997).

6. D. F. Specht : A general regression neural network. IEEE Trans. Neural Networks, vol. 2, pp. 568–576, (Nov. 1991).
7. D. F. Specht,: Enhancements to probabilistic neural network. in Proc.Int. Joint Conf. Neural Network, vol. 1, pp. 761–768, (1991).
8. Sameer A. Nene , Shree K. Nayar , Hiroshi Murase : Columbia Object Image Library (COIL-100) . Technical Report No. CUCS-006-96, Department of Computer Science Columbia University New York, N.Y. 10027.