Object Recognition based on Aspect Vision using Two-Dimensional Space Spectrum Analysis and Neural Network

メタデータ	言語: eng 出版者: 公開日: 2010-04-06 キーワード (Ja): キーワード (En): 作成者: Izumi, Masao, Kato, Satoshi, Kawakami, Hiroshi, Fukunaga, Kunio メールアドレス: 所属:
URL	https://doi.org/10.24729/00008414

Object Recognition based on Aspect Vision using Two-Dimensional Space Spectrum Analysis and Neural Network

Masao IZUMI*, Satoshi KATO**, Hiroki KAWAKAMI***
and Kunio FUKUNAGA*

(Received June 15, 1991)

In this paper, we propose a method for object recognition using two−dimensional space spectrum analysis and neural network on the basis of aspect vision approach. In order to avoid mismatching caused by rotation and translation, we use two-dimensional space spectrum as training data of multilayer feedforward neural network, as well as edge images from the viewpoints partitioned uniformly. The multilayer feedforward neural network works as the matching processor of the aspect graph and its two-dimensional space spectrum, and recognizes the objects in the scene taken from unlearned viewpoints. With the experiments, we show the effectiveness of recognition of some kinds of chairs.

1. Introduciton

Object recognition in two dimensional images is a major area of research in computer vision. The key problem in recognition lies in two-dimensional (2−D) representation of three-dimensional (3−D) objects, or in matching between 2−D input image and 2−D representations. The approach of multiple 2−D views of the objects from different viewpoints is effective to the former problem. The aspect graph approach is recently studied aggressively which is based on topologically distinct views of an object to represent its shape. [1]− [3] The latter problem (2−D to 2−D matching problem) is usually discussed as a topological matching between an input data and database in the aspect graph approaches. But if an input image taken from tv camera contains noise that breaks a line into two or more line segments, it becomes difficult to maintain a similarity of topology between the input image and the database. In this paper, we propose a new method of object recognition, using multilayer feedforward neural network to be trained with aspect graphs of the objects of the uniformly partitioned view space and their two-dimensional space spectrum. The outline of our algorithm is as follows. (Fig.1)

* Department of Electrical Engineering, College of Engineering

** Student, Department of Electrical Engineering, College of Engineering, presently Matsushita Electric Industry Co., Ltd.

*** Graduate student, Department of Electrical Engineering, College of Engineering

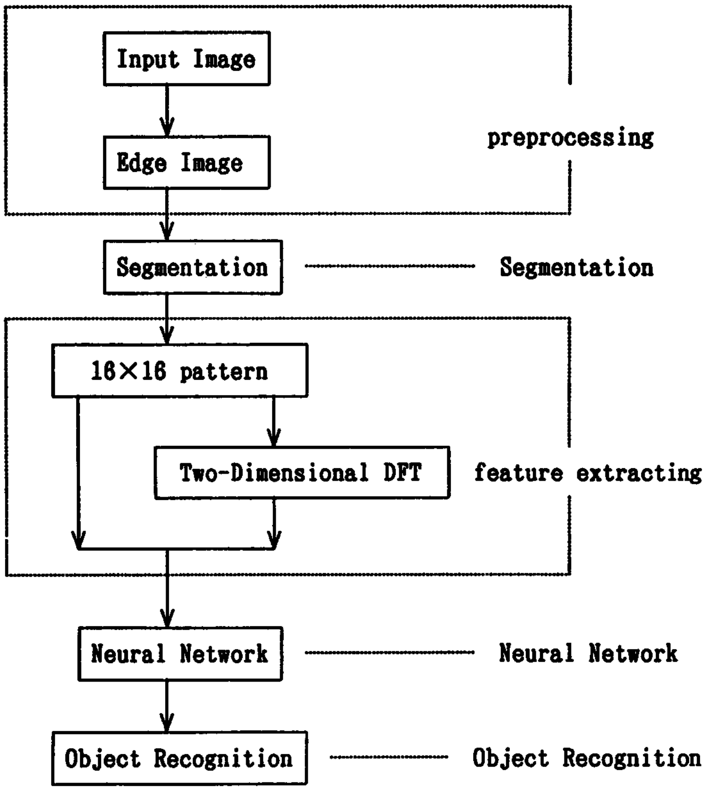


Fig. 1 Flow diagram of an algorithm

1) Generate line-drawing aspect images of the objects to be recognized from uniformly partitioned view spaces using surface data of objects.
2) Refine these images to training images which are suitable to the network.
3) Calculate two-dimensional space power spectrum of refined images.
4) Train the network with the data of both 2) and 3).
5) Input a gray scale image of the objects from CCD camera, detect edges and thin them, and calculate two-dimensional space power spectrum. Input them to the Input−layer of the network.

The remainder of this paper is organized as follows. Section 2 describes aspect graphs as training data of the network. Section 3 describes training data refinement to recognize the object from a viewing angle little different from the trained data. Section 4 describes an implementation of the proposed algorithm and shows some experimental results in the case of some kinds of chairs as the object world. Finally, section 5 discusses the major advantages and limitaion of the proposed method, and includes directions for the future work.

2. Aspect Graph

In the multiple 2−D view approach to 3−D object representation, there are two approaches to partitioning the space of viewpoints: 1) the uniformly partitioning approach that the viewing space is partitioned into a uniform manner by projecting a tessellated regular polyhedron on to the sphere, and 2) the aspect graph approach that the viewing space is partitioned according to the qualitative structure of the view. [1] In our approach we use a multilayer feedforward neural network for object matching, the former approach is better to create the training data of the network. Because so many graphs are generated in the case of complex objects, or less aspects are generated in the case of simple objects in the aspect graph approach. On the other hand, the uniform partitioning approach generates, of course, the uniform partitioned views so that every viewing angles to the object to be recognized are treated equally to the network for recognition. The aspect graph approach takes advantage of a topological matching to examine matching between an input data and database. But it becomes important to get high stability of topological structure of the input data to be recognized in order to avoid mismatching. It is not easy to keep the input data free from noise to get high stability of topological structure. So we take a pattern (not topological) matching using a simple three-layer feedforward neural network with an error back propagation learning, and two-dimensional space spectrum pattern is also used as a training data to avoid mismatching caused by rotation and translation. In the next sections, it is described about training data and network struture.

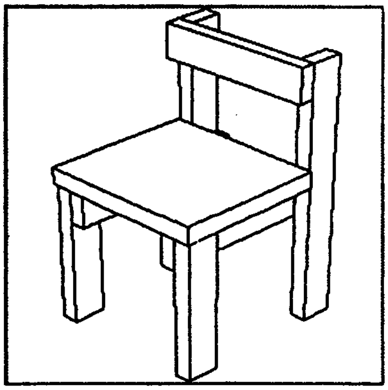
3. Input Data of Neural Network

In this section, it is described about the network structure and data refinement for training the neural network. It is a difficult problem to decide the number of units of the network. For the three-layer feedforward neural network, especially to an image data which needs large number of units of Input-layer, it needs reduction of the number of raw input data (image data) in order to reduce a network scale and learning time. In order to reduce of the number of units, we take an approach to refine training data to the network as follows.

3. 1 Edge Image Generation and Data Reduction.

Using the surface data of the objects, uniformly partitioned edge images are generated. (shown in Fig.2) Size of these images are 512×512 in our experiments.

In the next, we reduce the size of edge images to $n \times n$ ($n=16$, for example) images for decreasing the number of units of Input-layer of the network. We consider the number of edge points appeared in each $n \times n$ mesh.



(a) 512×512 image

.	.	.	.	.	.	26	32	64	.	.	.	.	.	.	.
.	.	.	.	.	.	32	1	34	64	58	30	.	32	5	.
.	.	.	.	.	.	10	70	34	5	6	34	64	53	32	.
.	.	.	.	.	.	.	64	32	28	32	32	51	17	32	.
.	.	.	.	.	.	14	85	32	.	.	1	80	8	30	.
.	.	.	.	32	32	18	.	42	43	16	.	64	16	16	.
.	10	44	24	.	.	.	.	.	.	16	32	83	30	2	.
.	14	78	36	24	.	.	.	.	.	14	32	90	32	.	.
.	.	64	60	64	40	32	.	28	24	50	42	72	32	.	.
.	.	74	75	.	24	46	60	51	46	32	22	64	32	.	.
.	.	64	12	20	.	32	68	34	18	.	.	64	32	.	.
.	.	64	.	32	.	14	66	32	.	.	.	64	32	.	.
.	.	64	4	30	.	.	64	32	.	.	.	64	32	.	.
.	.	30	28	.	.	.	64	32	.	.	.	39	5	.	.
.	.	.	.	.	.	.	64	32	.	.	.	.	.	.	.
.	.	.	.	.	.	.	64	38	.	.	.	.	.	.	.

(b) n×n image

Fig.2 Size reduction of images

3. 2 Median Filter

The reduction of the number of input data mentioned above is effective to reduce the number of whole units of the network. But a small difference in the edge images sized 512×512 caused by translation induces a large difference in n×n mesh pattern, because n×n mesh pattern is partitioned uniformly if an edge line lies across a boundary of mesh or not. In the case in Fig.3, there is a little difference between (a) and (b), but there is much difference between (c) and (d). In order to reduce this difference, we use a median filter to the n×n patterns. In Fig.3, there is a small difference between filtered patterns (e) and (f).

3. 3 Two-Dimensional Space Power Spectrum

And we calculate DFT (discrete Fourier transform) of resized edge images obtained above. Calculation from equation (2) derives only power component of two-dimensional space spectrum so that we can extract characteristics insensitive to translation.

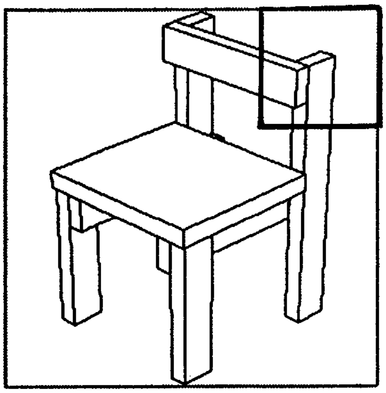
$$F(u,v)=\sum_{x=0}^{n-1}\sum_{y=0}^{n-1} f(x,y)\exp\left(j\cdot 2\pi\left(\frac{ux}{n}+\frac{vy}{n}\right)\right) \tag{1}$$

where f(x,y) is a value of position(x,y) of the reduced edge image.

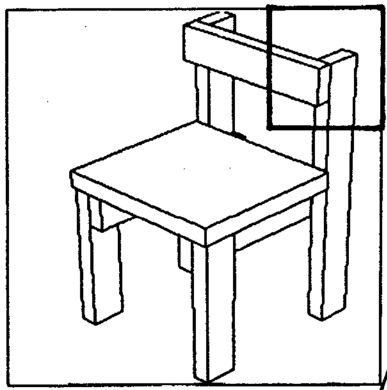
Power component of space spectrum is deduced by the equation as follows.

$$|F(u,v)|=\sqrt{\mathrm{Re}\{F(u,v)\}^2+\mathrm{Im}\ \{F(u,v)\}^2} \tag{2}$$

These two features, the numbers of edge points and power space spectrum of them, are used as input data to the network. The number of input data is 2×n×n.



(a) edge images (512×512)



(b) edge images (512×512)

.	.	.	.	.
30	.	32	5	.
34	64	53	32	.
32	51	17	32	.
1	80	8	30	.

(c) no filtered

.	.	.	.	.
50	20	18	27	.
14	44	100	51	.
32	26	68	32	.
.	20	62	32	.

(d) no filtered

.	.	.	.	.
6	30	5	.	.
32	32	32	17	.
32	34	32	17	.
32	17	32	16	.

(e) filtered

.	.	.	.	.
14	18	20	.	.
32	32	44	27	.
20	32	44	32	.
26	32	32	32	.

(f) filtered

Fig. 3 Median filter

4. Experiments and Results

4. 1 Network Structure and Learning

We use a three-layer feedforward neural network to recognize the object. (Fig. 4) The number of Input-layer units is $2\times n\times n$ ($n=16$. in this experiment), the number of Hidden-layer units is 30 and the number of Output-layer units is 5. The number of Hidden-layer units is decided by pre-experiment under condition of $n=16$. Results of this pre-experiment (Fig.6 (a)) show that recognition rate is the best value when the number of Hidden-layer units is 30. Learning algorithm is an error back propagation (Eq.(3)). In the following equations, $\Delta w^{K-1}{}_i{}^K{}_j(t)$ is the t-th weight difference from unit i in layer $k-1$ to unit j in layer k, $i^k{}_j$ is the sum of input to unit j in layer k, $o^k{}_j$ is an output of unit j in layer k, y_j is a training data of unit j, $f'(x)$ is a first derivative of sigmoid function $f(x)$.

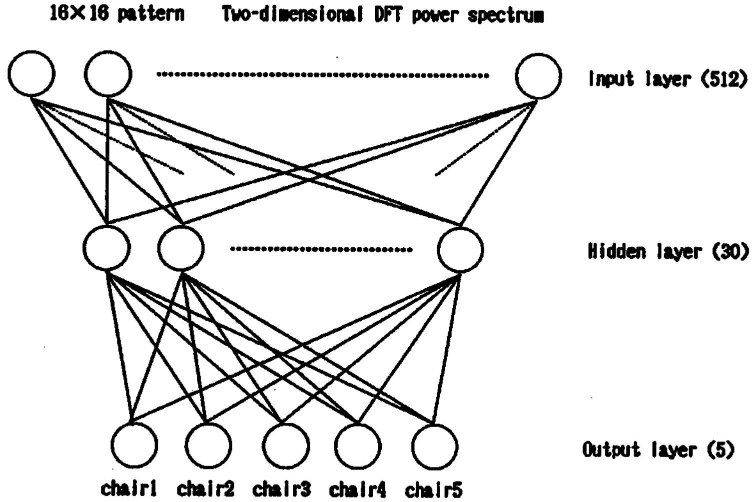


Fig. 4 Three-Layer Feedforward Network

$$\Delta w^{k-1}{}_i{}^k{}_j(t+1) = -\varepsilon d^k{}_j o^{k-1}{}_i + \alpha \Delta w^{k-1}{}_i{}^k{}_j(t) \tag{3}$$

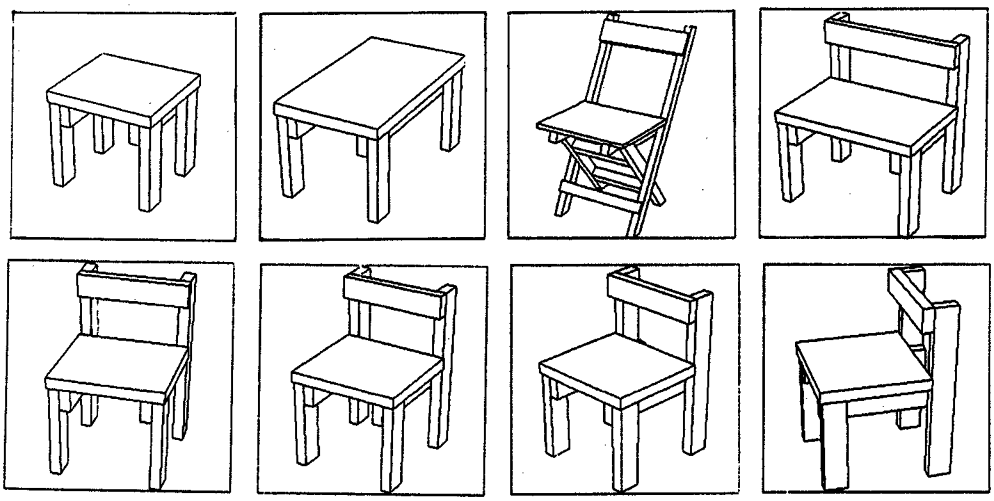
$$d^k{}_j = \begin{cases} (o^k{}_j - y_j) f'(i^k{}_j) & (k=m) \\ (\Sigma w^k{}_j{}^{k+1}{}_l d^{k+1}{}_l) f'(i^k{}_j) & (k\neq m) \end{cases} \tag{4}$$

$$f(x) = \frac{1}{1+\exp(-x)} \tag{5}$$

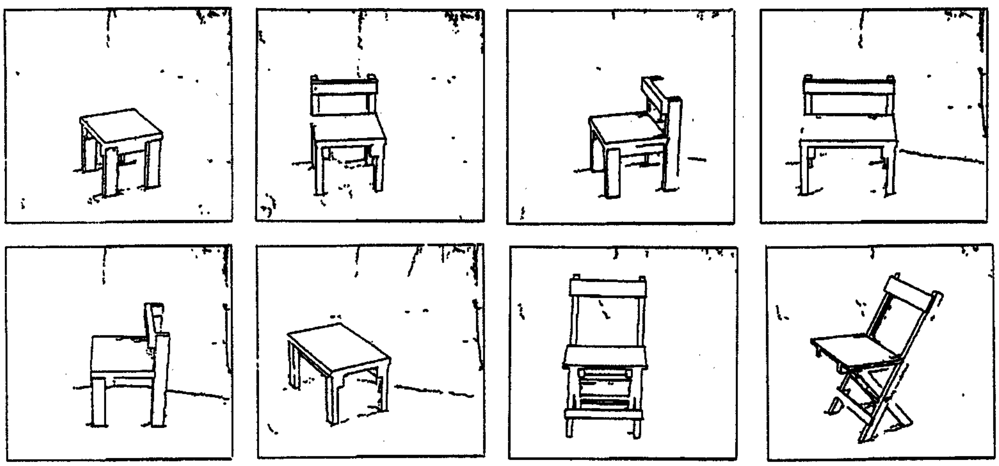
In our experiments, the parameter of momentum (α in Eq.(3)) is 0.9, and the learning rate (ε in Eq.(3)) is 0.25.

4. 2 Objects to be Recognized

Fig.5 (a) shows objects to be recognized. There are five kinds of chairs constructed by polyhedral units. First we generate edge images of all the objects from 37



(a) generated (training) data



(b) input data (from CCD camere, edge images)

Fig. 5 Object images

different directions (from 0 degree to 180 degree rotationg 5 degree at a step) under the same depression angle (about 30 degree) of using surface model data of the objects. The size of these edge images is 512×512.

4. 3 Training Network

Training data is prepared like as a following list. Experiment 1) comparison between training with median filtered edge images and training the edge images without median filtered. Experiment 2) comparison among training only using edge images, training only using two-dimensional space power spectrum and training using both data. For eah case, we train the network five times because the weights of all the units of the network are initially setup at random and the error back propagation decreases the MSE (mean square error) until it becomes less than threshold value (0.0001). After training, we calculate a recognition rate as the average of results from experiments five times.

4. 4 Object Recognition

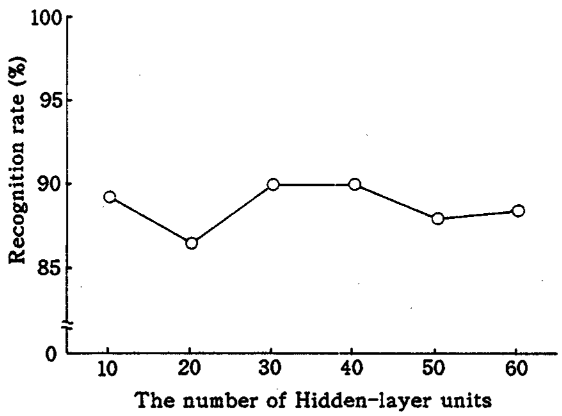
After training network, we present 60 patterns (12 patterns for each object), which are partly shown in Fig.5 (b), into Input-layer of the network taken from various directions under the same depression angle as the training data. These images can be obtained after transformation from gray scale images taken from CCD camera using Sobel operation and thinning operation. And these edge images are also transformed to two-dimensional space power spectrum using Eq.(2).

4. 5 Experimental Results

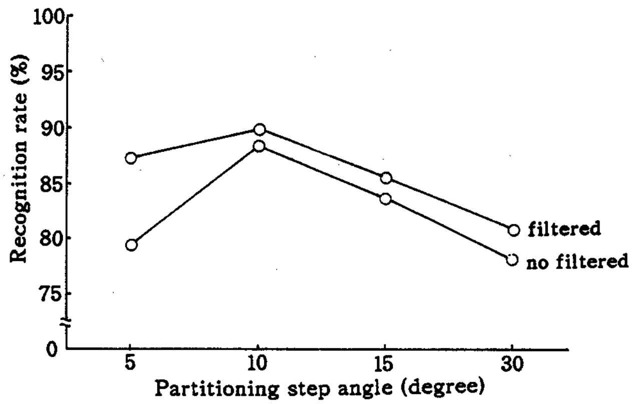
Figs.6 (b) and (c) show recognition rates on the various state of network and training data format. Fig.6 (b) shows results in the case of comparison between training with median filtered edge images and training with the edge images without median filtered. In all partitioning step angles, recognition rates trained with median filtered data are higher than without median filtered. Concerning about partitioning step angle, recongnition rate is best in the case that partitioning step angle is 10 degree. Fig.6 (c) shows results of Experiment 2) of 4.3. Recognition rates trained with both edge patterns and power spectrum patterns are higher than any others. In this experiment, recognition rate is best also in the case that partitioning step angle is 10 degree.

5. Conclusions

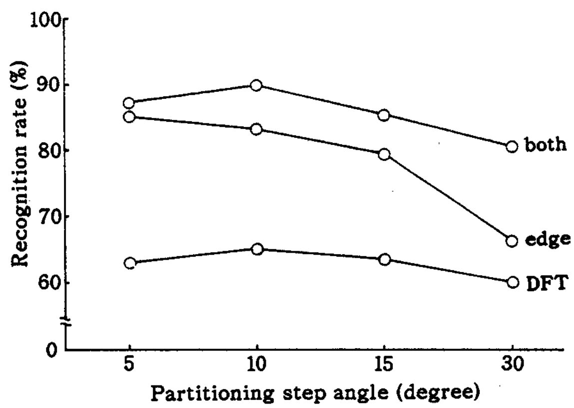
This paper proposes a new approach to object recognition using three-layer feedforward neural network under the idea of aspect graphs. The aspect graphs approach makes the recognition problem of 3−D objects into 2−D matching. But this matching process needs topological stability in the process before matching. In our approach, we consider an edge-based pattern matching problem using neural network instead of a topological matching problem. Adding this, because of training



(a) pre-experiment (various numbers of Hidden-Layer units)



(b) comparison between training with median filtered images and without filtered images



(c) comparison among training with only edge patterns, with only DFT patterns, and with both patterns

Fig. 6 Experimental results (recognition rate) the network with both edge data and two-dimensional space power spectrum, it can avoid mismatch between the input images and database caused by translation of rotation of input images. The limitation of this approach is a necessity of large number of training data in the case that input images of the objects to be recognized are taken from various directions under various depression angles. It is expensive to execute training the network with large amount of data from more various directions and of more various kinds of objects. Future directions of our work will focus on the idea of combination system with proposed approach in this paper and topological (aspect) approach. In the case of various depression angles, two step approach that it recognizes object under the estimated depression angle in the first step using global topological matching of the input images and database of aspects under various depression angles. This database in the first step is constructed with the aspects of major features that induce differences under various depression angles in the object world. In the case of such an object world like some kinds of chairs, this major feature will be a pedestal part of the chair.

Acknowledgements

The authous would like to thank Professor Shojiro Yoneda for his advice and encouragement.

References

[1] Z. Gigus and J. Malik, "Computing the Aspect Graph for Line Drawing of Polyhedral Objects", IEEE Trans. Pattern Analysis and Machine Intelligence, PAMI−12, 2, pp. 113−122, Feb. 1990.

[2] K. Ikeuchi and T. Kanade, "Automatic Generation of Object Recognition Programs", Proc. of IEEE, Vol. 76, No.8, pp. 1016−1035, Aug. 1988.

[3] C. Hansen and T. Henderson, "Towards the Automatic Generation of Recognition Strategies", Proc. of 2nd International Conference on Computer Vision, pp. 275−279, Dec. 1988

[4] M. Izumi, S. Kato and K. Fukunaga, "Object Recognition based on Aspect Vision using Neural Network", in Proc. of The 7th Scandinavian Conference on Image Analysis, Aalborg (Denmark), Aug. 1991.