2006 IEEE International Conference on
Systems, Man, and Cybernetics
October 8-11, 2006, Taipei, Taiwan

Structure Evaluation of Receptive Field Layer in TAM Network

Isao Hayashi and Toshiyuki Maeda

***Abstract*— TAM (Topographic Attentive Mapping) network is a biologically-motivated neural network with receptive fields of Gabor function. However, the receptive field's layer is monotype, and there is a lack of performance for rotating visual images. In this paper, we formulate four TAM networks with multilayer structure of extensive receptive fields. We discuss their performance and show the usefulness of TAM network using some examples of character recognition.**

I. INTRODUCTION

In the human visual system, the visual modalities, e.g., image's orientation is detected at the primary visual cortex (V1) through the lateral geniculate nucleus (LGN). Object's rotation and symmetry are detected using the orientation selectivity in the higher level visual cortex[1]. The source of orientation selectivity is captured in retina using receptive fields of simple cell, complex cell and hypercomplex cell for detecting contours of visual objects. As a model representing the receptive field, Marčelja[2] has proposed Gabor function using a complex sinusoidal plane wave of some frequency and orientation. Daugman[3] extended the Gabor function to two-dimensional model. Other many receptive field models [4]-[6] have been also proposed.

On the other hand, many powerful neural networks [7]-[12] based on Hubel-Wiesel's architecture have been proposed. We have also already formulated TAM (Topographic Attentive Mapping) network, which incorporates Gabor function type receptive field in the input layer. The TAM network has been proposed by Williamson[11], which has four layers, i.e., the feature layer, the basis layer, the category layer and the class layer, and represents the input-output relationship between the distributed input data and the categorized output classes. We have also proposed advanced TAM network, which has six layers including the retina and the LGN layers adding to the original TAM network[13]-[17]. After LGN layer, the feature map of TAM network is compound from density values of the sixteen orientations, and the input-output relationship of data sets is acquired by two procedures, the bottom-up procedure from feature map to the class layer, and the top-down procedure to category layer from class layer. However, the performance for rotating objects is not so well since the layer structure of receptive fields is monotype, and a mechanism getting orientation selectivity covering wider visual area doesn't work.

In this paper, we propose double layer architecture of receptive fields in the retina layer, which are a filtering detector for object's orientation selectivity at the first layer, and a rotation detector at the second layer, respectively. We propose here four types of receptive field architectures. Four receptive field types are as follows. The first type, the receptive field structure covering all vision area, CVA, has a receptive filed architecture which is covering whole vision area using the max size receptive field in the second layer. The second type, the receptive field structure with double layers, DL, contains the same side receptive filed in the second layer. The third type, the receptive field structure with orientation vector, OV, calculates the vector of orientation selectivity in the second layer with the max size of receptive field for multiplying the weights of visual images. The forth type, the receptive field structure with double orientation vector, DOV, is also calculating the vector of orientation selectivity in the second layer, but for multiplying the weights of the first layer's images.

Four types' receptive field architecture are mimic simple cell and complex cell, and thus the TAM network should be a biologically-motivated neural network. We discuss here the mechanism of four receptive field architectures, and show the usefulness and performance of the TAM network through some examples of character recognition.

II. RECEPTIVE FIELD STRUCTURE IN TAM NETWORK

The structure of TAM network with receptive field is shown in Figure 1. Sixteen orientation selectivities of visual images are detected in the retina layer by Gabor filtering. The feature map is compound from density values of these sixteen orientations. The procedure for detecting orientation selectivity in the retina and LGN layer is explained in Figure 2. Figure 3 shows the learning mechanism at the upper feature maps for recognizing the visual object images.

In the retina layer, the two-dimensional Gabor function type receptive field, $G(x, y)$, is represented as follows:

$$\begin{aligned} G(x,y) &= Ke^{-\frac{1}{2}\left(\frac{(x-\mu x)^2}{\sigma_x^2}+\frac{(y-\mu y)^2}{\sigma_y^2}\right)} \\ &\quad \times \sin\left(2\pi f_x x\cos\theta + 2\pi f_y y\sin\theta + \phi\right) \end{aligned}$$

where, K is amplitude, (μ_x, μ_y) is the center coordinate of Gabor function, σ_x and σ_y are standard deviations of axis X and Y, respectively, and f_x and f_y are frequencies of axis X and Y, respectively. Figure 4 shows an example of Gabor function.

This is partially supported by the Ministry of Education, Culture, Sports, Science, and Technology of Japan under Grant-in-Aid for Scientific Research number 18500181.

I. Hayashi is with Faculty of Informatics, Kansai University, 2-1-1, Ryozenji-cho, Takatsuki, Osaka 569-1095, Japan ihaya@kcn.res.kutc.kansai-u.ac.jp

T. Maeda is with the Faculty of Information Management, Hannan University, 5-4-33, Amami-higashi, Matsubara, Osaka 580-8502, Japan maechan@hannan-u.ac.jp

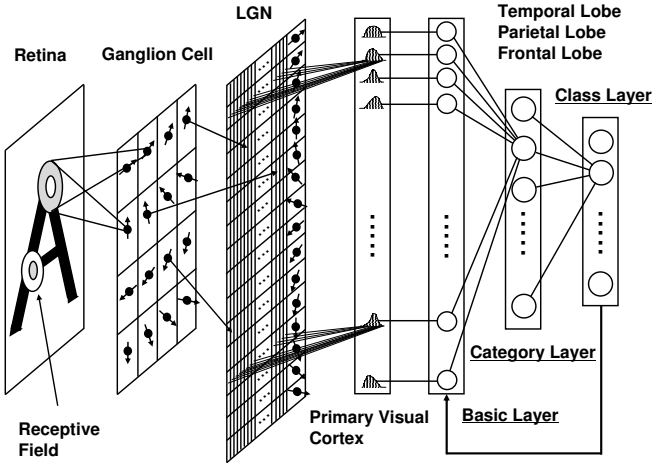

Fig. 1. TAM Network with Receptive Field

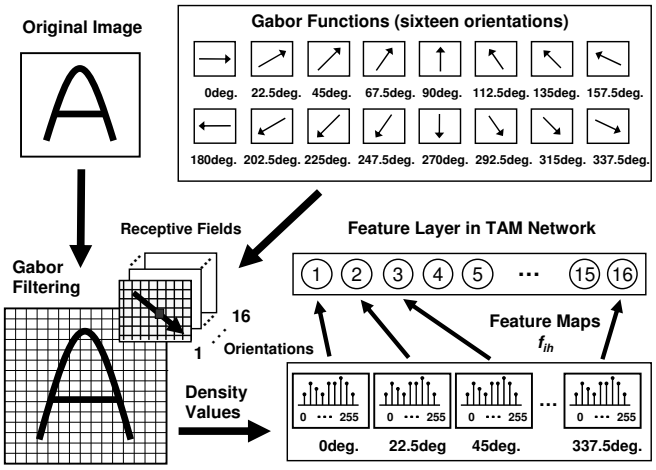

Fig. 2. Gabor Filtering Process

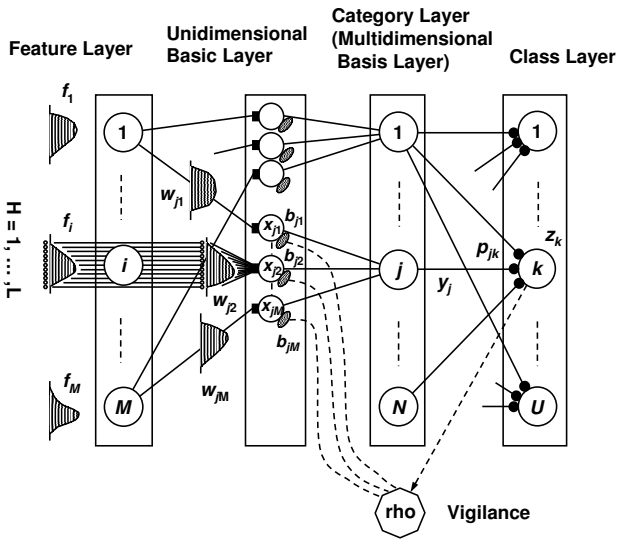

Fig. 3. Structure at the Upper Feature Maps

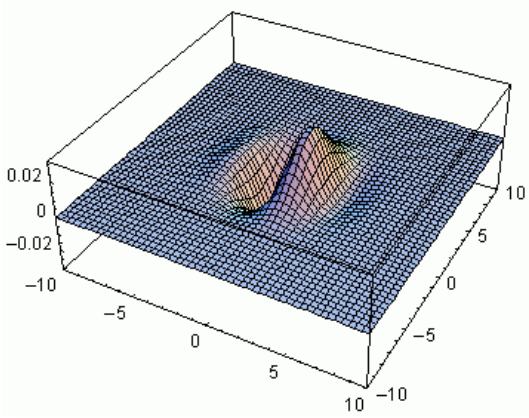

Fig. 4. Gabor Function

Now, the horizontal and vertical scale of the visual images in the retina layer denote as R_H $pixels$ and R_V $pixels$, respectively. The luminance of the visual images in the coordinate, (p, q), is denoted by $O(p, q)$, $0 \leq p \leq R_H$, $0 \leq q \leq R_V$. In the ganglion cell part of the retina layer, the i-th orientation selectivity of contour image, $C_i(x, y)$, $i = 1, 2, \cdots 16$, at the coordinate (x, y) is calculated by the following convolution process:

$$C_i(x, y) = \sum_{q=1}^{R_V} \sum_{p=1}^{R_H} G_i(x - p, y - q) \times O(p, q)$$

The i-th orientation selectivity of a visual image, $C_i^{all}(x, y)$, is calculated by the following union process:

$$C_i^{all} = \bigcup_X \bigcup_Y C_i(x, y), \qquad x \in X, \quad y \in Y.$$

However, the performance of the TAM network for rotating objects is not so well if the receptive field structure is applied, because the layer structure is monotype, and a mechanism getting orientation selectivity for covering wider visual area doesn't work. In this paper, we propose a double layer structure for overcoming rotation image problems. In these proposed methods, the i-th orientation selectivity, $C_i^2(x, y)$, at the second layer is captured using the i-th orientation selectivity, $C_i^1(x, y)$, which is calculated at the first layer. we propose here the following four types of receptive field architecures.

1) Receptive Field Structure Covering All Vision Area (CVA):

$$C_i^2(\frac{R_V}{2}, \frac{R_H}{2}) = \sum_{q=1}^{R_V} \sum_{p=1}^{R_H} G_i^{max}(\frac{R_V}{2} - p, \frac{R_H}{2} - q) \times C_i^1(p, q)$$

2) Receptive Field Structure with Double Layers (DL):

$$C_i^2(x,y) = \sum_{q=1}^{R_V}\sum_{p=1}^{R_H} G_i(x-p, y-q) \times C_i^1(p,q)$$

3) Receptive Field Structure with Orientation Vector (OV):

$$C_i^2(x,y) = \frac{C_i^1(x,y) \times \sum_{q=1}^{R_V}\sum_{p=1}^{R_H} G_j^{max}(\frac{R_V}{2}-p, \frac{R_H}{2}-q) \times O(p,q)}{\sum_{j=1}^{16}\sum_{q=1}^{R_V}\sum_{p=1}^{R_H} G_j^{max}(\frac{R_V}{2}-p, \frac{R_H}{2}-q) \times O(p,q)}$$

4) Receptive Field Structure with Double Orientation Vector (DOV):

$$C_i^2(x,y) = \frac{C_i^1(x,y) \times \sum_{q=1}^{R_V}\sum_{p=1}^{R_H} G_j^{max}(\frac{R_V}{2}-p, \frac{R_H}{2}-q) \times C_i^1(p,q)}{\sum_{j=1}^{16}\sum_{q=1}^{R_V}\sum_{p=1}^{R_H} G_j^{max}(\frac{R_V}{2}-p, \frac{R_H}{2}-q) \times C_i^1(p,q)}$$

where, $C_i^{max}(x,y)$, is the Gabor function at the coordinate, (x,y), with max size of receptive field for covering whole visual area.

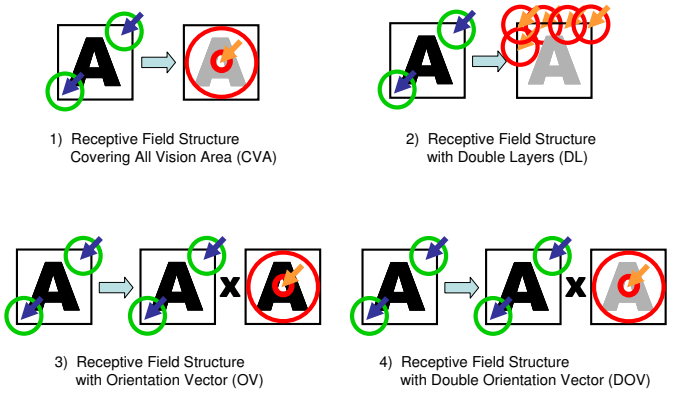

Fig. 5. Four Receptive Field Types

Figure 5 shows four types of receptive field architectures. The first receptive field type, CVA, has a receptive filed structure in the second layer which is covering whole visual area with the max size. The second receptive field type, DL, contains the same receptive filed structure in the second layer as same as the first layer structure. The third receptive field type, OV, calculates the orientation vector of the original visual images filtered by the max size of receptive field, and multiply the vector as a weight value by the first layer's orientation selectivity. The forth receptive field type, DOV, calculates also the orientation vector, but for the first layer's images filtered by the max size of receptive field, and multiply the vector by the first layer's orientation selectivity.

III. Algorithm of TAM Network

The i-th feature map, f_{ih}, at the feature layer of TAM network is calculated by the following normalization:

$$f_{ih} = \frac{\sum_{\{x,y|O_i(x,y)=h\}} C_i^1(x,y)}{\sum_{y=1}^{R_V}\sum_{x=1}^{R_H} C_i^1(x,y)}, \quad h = 1, 2, \cdots, 256.$$

We define here the rotation selectivity for estimating the rotation angle of visual images. Now, we prepare two kinds

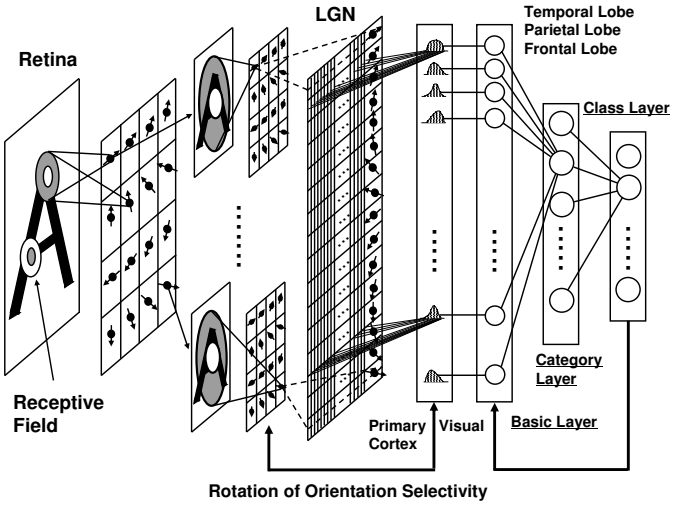

Fig. 6. TAM Network with Double Receptive Fields

of images, the training image for learning TAM network, and the checking image for estimating the accuracy of the learning recognition. After learning TAM network using the training image, the orientation selectivity of the checking image, $C_{i,CHD}^2(x,y)$, is calculated by four proposed methods, and the feature map of the checking image, f_{ih}, is exchanged for minimizing the following rotation angle, r, comparing with the orientation selectivity of the training image $C_{I,TRD}^2(x,y)$.

$$r = \min_i \left(\frac{\sum_x\sum_y C_{i,CHD}^2(x,y)}{R_H \times R_V} - \frac{\sum_x\sum_y C_{I,TRD}^2(x,y)^2}{R_H \times R_V}\right)$$

At the feature layers of TAM network, the activities, x_{ji}, $i = 1, 2, \cdots, M$, $j = 1, 2, \cdots, N$, are calculated in the basis layer by the following equation with feature maps, f_{ih}, and node's weights, w_{jih}, $h = 1, 2, \cdots, L$. The output signal of the category layer, y_j, is then calculated by products with activities x_{ji}.

$$x_{ji} = \frac{\sum_{h=1}^{L} f_{ih} w_{jih}}{1 + \rho^2 b_{ji}}$$

$$y_j = \prod_{i=1}^{M} x_{ji}$$

where, ρ represents the vigilance parameter, and b_{ji} are inhibitory weights.

The output prediction, K, at the class layer is calculated as follows:

$$z_k = \sum_{j=1}^{N} y_j p_{jk}$$

$$K = \{k | \max_k z_k\}$$

where, z_k are outputs of the k-th class layer, and p_{jk} are weighted connections.

TAM network has a mechanism of adding a new node to the category layer when there is some deviation between the output of class layer and the correct supervised output. Let K^* denote the index of the "correct" supervised output class. In the case of $K \neq K^*$, the vigilance parameter, ρ, is rising from the initial value, ρ^{init}, until either $\rho \geq \rho^{max}$ or $z_{K^*}/z_K \geq OC$, where OC is the maximal vigilance level. When the vigilance parameter reaches the maximal level, ρ^{max}, a node is added to the category layer.

On the other hand, when the subject of $z_{K^*}/z_K \geq OC$ is satisfied, the following feedback signal y_j^* is calculated, and the learning parameters, w_{jih}, p_{jk}, and b_{ji}, are renewed.

$$z_k^* = \begin{cases} 1 & ; \text{ if } k = K^* \\ 0 & ; \text{ otherwise} \end{cases}$$

$$\begin{aligned} y_j^* &= \frac{\prod_{i=1}^{M} x_{ji} \times \sum_{k=1}^{U} z_k^* p_{jk}}{\sum_{j'=1}^{N} \prod_{i=1}^{M} x_{j'i} \times \sum_{k=1}^{U} z_k^* p_{j'k}} \\ \triangle w_{jih} &= \frac{\alpha y_j^* (1-\lambda^{1/M})(f_{ih} - w_{jih})}{(\alpha-1)\lambda^{1/M} + n_j}, \quad \lambda \in (0,1) \\ \triangle p_{jk} &= \frac{\alpha y_j^* (z_k^* - p_{jk})}{\alpha + n_j} \\ \triangle b_{ji} &= b_j^{(rate)} y_j^* (x_{ji} - b_{ji}) \\ \triangle n_j &= \alpha y_j^* (1 - n_j) \end{aligned}$$

where, α, λ are parameters, and $b_j^{(rate)}$ are learning parameters.

IV. Character Recognitions

In order to show the performance and usefulness of TAM network, some examples of character recognition are illustrated here. As an input image, all 26 alphabets are filled in the electronic pad whose size is 15 $pixels \times$ 15 $pixels$. The following three kinds of experiments are discussed here.

1) Calculating recognition rates of all 26 kinds of alphabets using the original TAM network with single receptive filed layer.
2) Calculating recognition rates of 26 alphabets using the original TAM network when these alphabets are rotated and declined.
3) Calculating recognition rates of the rotation characters of 'A' and 'B' using four types of receptive field architectures.

The learning parameters are set as follows:

L	=	255	ρ^{init}	=	0.0
OC	=	0.8	ρ^{step}	=	0.1
α	=	0.01	ρ^{max}	=	100.0
λ	=	0.33	$b_j^{(rate)}$	=	0.0000001
μ_x	=	0.0	μ_y	=	0.0
σ_x	=	1.99	σ_y	=	1.92
f_x	=	0.127	f_y	=	0.127
ϕ	=	90.0	K	=	1.0.

First, we show the performance of the original TAM network. The training images are shown in Figure 7. The contour images of the left upper side of alphabet 'A' in Figure 7 are shown in Figure 8. In Figure 8, the left-upper side of images indicates the orientation selectivity at 0°, the next right side image shows the orientation selectivity at 22.5°, and the next image shows at 45°, 67.5°, $\cdots$, 337.5°, respectively. From those orientation images, we should notice that Gabor filtering certainly detects the orientation selectivity.

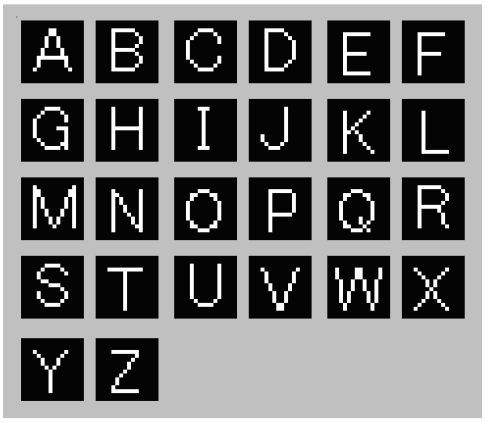

Fig. 7. Training Images

Fig. 8. Orientation Selectivity after Gabor Filtering

The resultant recognition rates of all 26 checking alphabets are shown in Figure 9. The recognition rates of both cases in two epochs and five epochs are compared as the average of 30 trials. In the case of the five epocks, the almost recognition rate is over the sixty percentages.

Next, we discuss recognition rates when the character images are rotated and declined. The training images are shown in Figure 7, and the checking images are shown in Figure 10. In Figure 10, three types of rotating charecters are prepared by rotating 90°, 180°, and 270°, and four types of declining characters are prepared. The resultant recognition rates are shown in Table I. The almost recognition rates are zero percentage, i.e., TAM network couldn't recognize the rotating and declining characters. However, the alphabets, 'I' and 'L', have relatively robustness for rotating 180° and 270°. The alphabets, 'N', 'O', 'X', and 'Z', are strongly robust for rotating 180°. Since these alphabets consist of straight lines and simple connections, the results should be understandable. We suppose that the result comes from the mono-layer as the receptive field, and thus there is a lack of mechanism for rotation.

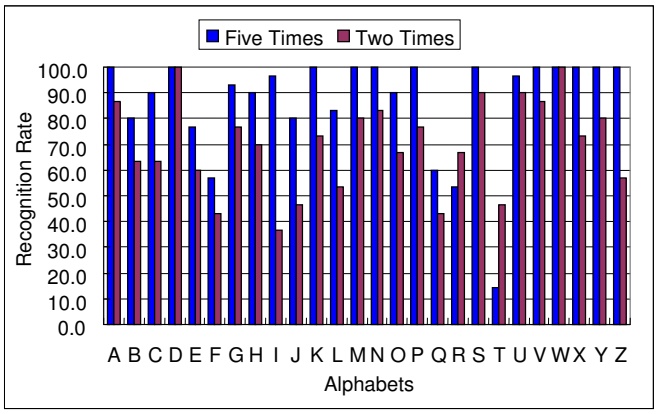

Fig. 9. Results of All Alphabets

TABLE I
RECOGNITION RATES OF ROTATION AND DECLINING

Alpha.	Recognition Rate (%)							
	Rotation				Declining			
	0°	90°	180°	270°	45°	135°	225°	315°
A	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
B	80.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
C	90.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
D	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
E	76.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0
F	56.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0
G	93.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0
H	90.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
I	96.7	60.0	96.7	73.3	20.0	0.0	80.0	3.3
J	80.0	0.0	0.0	0.0	6.7	0.0	0.0	0.0
K	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
L	83.3	0.0	20.0	50.0	6.7	10.0	10.0	10.0
M	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
N	100.0	0.0	93.3	0.0	0.0	0.0	0.0	0.0
O	90.0	0.0	80.0	0.0	0.0	0.0	0.0	0.0
P	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Q	60.0	0.0	20.0	0.0	0.0	0.0	0.0	0.0
R	53.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0
S	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
T	100.0	3.3	0.0	0.0	3.3	0.0	13.3	3.3
U	96.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0
V	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
W	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
X	100.0	13.3	100.0	0.0	0.0	0.0	0.0	0.0
Y	100.0	0.0	0.0	16.7	0.0	30.0	0.0	0.0
Z	100.0	0.0	93.3	16.7	16.7	0.0	80.0	76.7

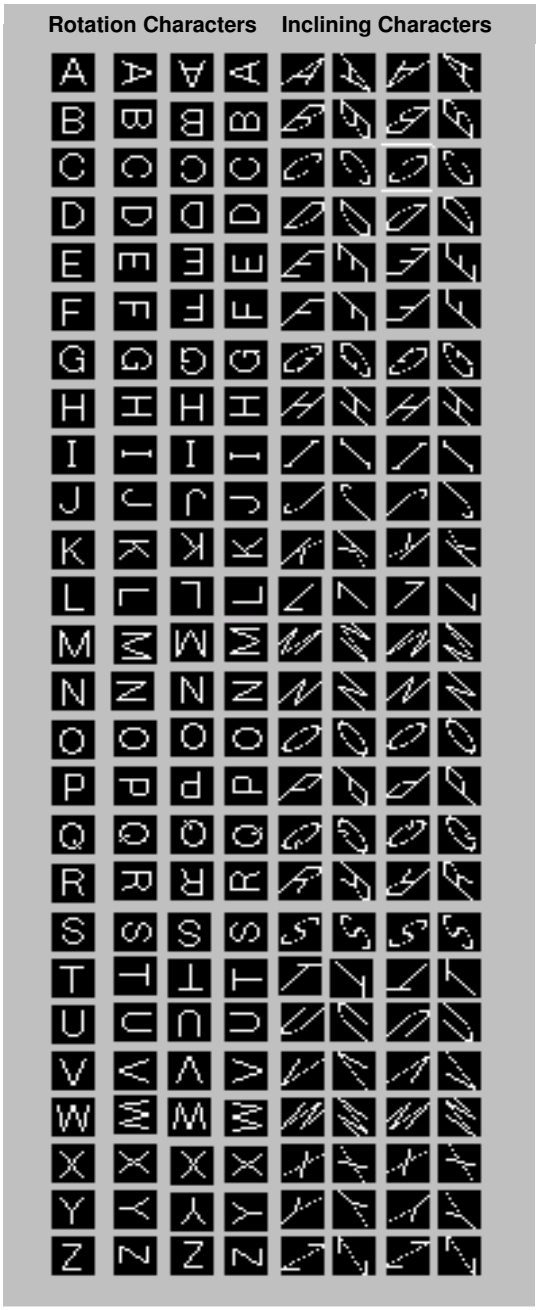

Fig. 10. Checking Images of Rotation and Declining Characters

We discuss last the performance of TAM network with double receptive fields layer architecture using two kinds of alphabets, 'A' and 'B'. The training images are shown in Figure 11. The checking images are shown in Figure 12. The resultant recognition rates are shown in Table II. The recognition rate is calculated with the average of 30 trials. The orientation selectivity of four types is shows in Figure 13- Figure 16. In the second layer of Figure 13, we should notice that whole orientation selectivity of original visual images is captured as white point values, because the whole orientation selectivity is shrinking to the center point by the Gabor filtering with the max size of receptive

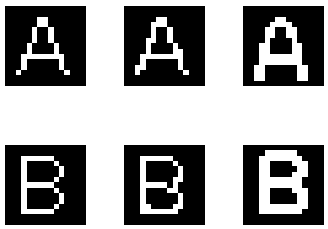

Fig. 11. Training Images of 'A' and 'B'

TABLE II

RECOGNITION RATE OF FOUR TYPES

	Char.'A' (%)	Char.'B' (%)	Average (%)
TAM with Single Layer	63.99	59.95	62.12
CVA	75.88	62.78	69.52
DL	49.44	59.63	54.23
OV	72.99	66.21	69.72
DOV	76.92	67.42	72.12
Average	68.80	64.01	66.40

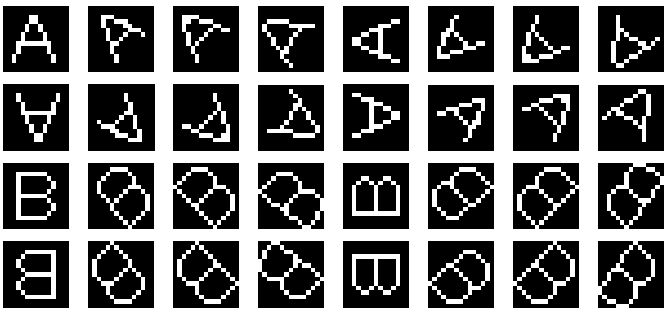

Fig. 12. Checking Images

field, and pointed as strong as white value's luminance. In the seconf layer of Figure 14, the similar orientation selectivity to the first layer is captured. However the images of orientation selectivity are more ambiguous than the first layer's orientation selectivity. In the second layer of Figure 15, the orientation selectivities of the first layer's images are multiplied by the vectors of the original visual images. In the second layer of Figure 16, the orientation selectivities of the first layer's images are multiplied by the vectors of the first layer's images. The orientation vectors of the second layer in Figure 16 are interspersed more than Figure 15, and thus the orientation selectivity of the first layer's image is emphasized in the second layer. In the Table II, we should notice that three proposed methods without DL are better than the original TAM network, and especially

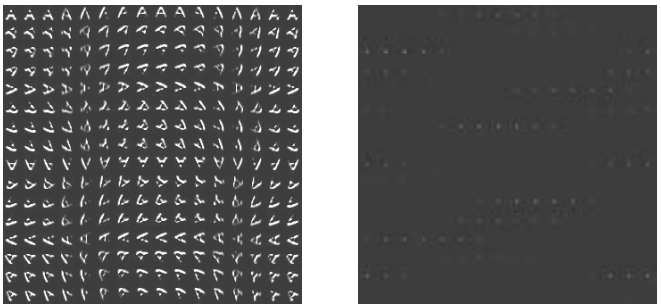

Fig. 13. Orientation Images of CVA

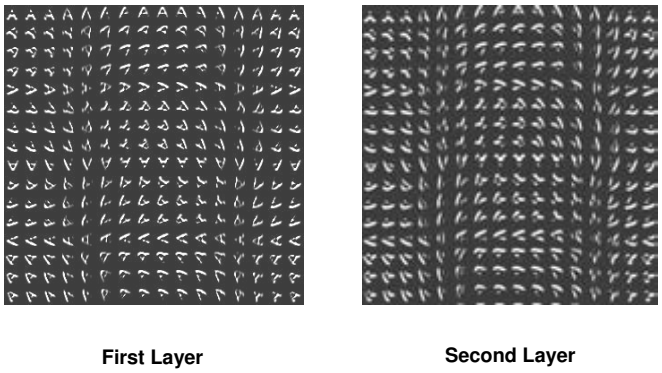

Fig. 14. Orientation Images of DL

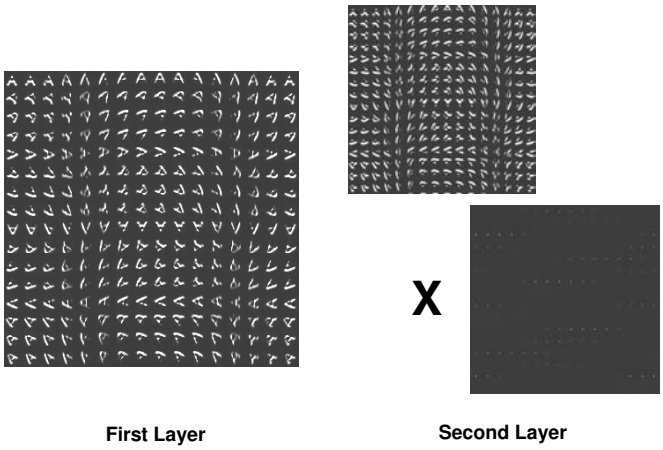

Fig. 15. Orientation Images of OV

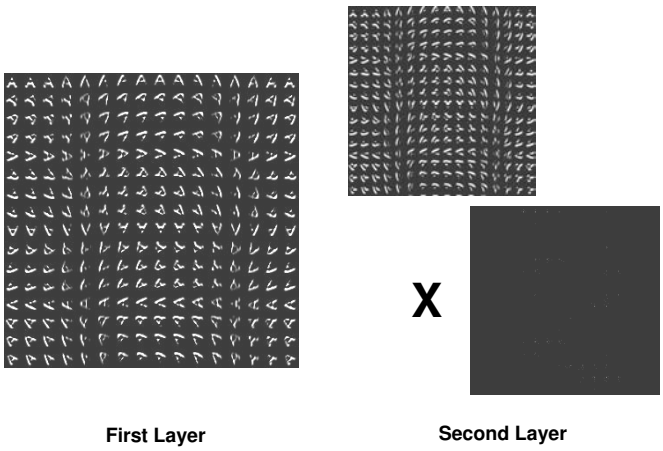

Fig. 16. Orientation Images of DOV

the recognition rate of DOV is over 70%. The two kinds of methods using orientation vectors, OV and DOV, are relatively better than others.

However, we should discuss more mechanism how to incorporate orientation vectors in the second layer. We should also discuss how to modify structure of TAM network and how to adjust precisely parameters of Gabor function in order to improve the recognition rate.

V. CONCLUSIONS

We formulated here a new TAM network with double layers architecture of receptive fields, and discuss the usefulness of TAM network through some examples of character recognition.

This is a product of research which was financially supported in part by the Kansai University Research Grants: Grant-in-Aid for Joint Research, 2006. "Building and Application of Construction Product Model in Road Maintenance Field". In addition, this research is partially supported by the Ministry of Education, Culture, Sports, Science, and Technology of Japan under Grant-in-Aid for Scientific Research number 18500181.

REFERENCES

[1] S.Yantis, "Visual Perception: Essential Readings," East Sussex, UK: Psychology Press, 2000.

[2] S.Marčelja, "Mathematical description of the responses of simple cortical cells," Optical Society of America, Vol.70, No.11, pp.1297-1300, 1980.

[3] J.Daugman, "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters," Optical Society of America, Vol.2, No.7, pp.1160-1169, 1985.

[4] A.D.Pollen and S.F.Ronner, "Visual cortical neurons as localized spatial frequency filters," IEEE Transactions of System, Man and Cybernetics, Vol.SMC13, pp.907-916, 1983.

[5] D.C.Lee, "Adaptive processing for feature extraction: Application of two-dimensional Gabor function," Remote Sensing, Vol.17, No.4, pp.319-334, 2001.

[6] K.Okajima, "Two-dimensional Gabor-type RF as derived by mutual information maximization," Neural Networks, Vol.11, pp.441-447, 1998.

[7] S.Grossberg, "How does the cerebral cortex work? Learning, attention, and grouping by the laminar circuits of visual cortex," Spatial Vision, Vol.12, No.2, pp.163-185, 1999.

[8] H.Neumann and W.Sepp, "Recurrent V1-V2 interaction in early visual boundary processing," Biological Cybernetics, Vol.81, pp.425-444, 1999.

[9] K.Fukushima, "Neocognitron: A hierarchical neural network capable of visual pattern recognition," Neural Networks, Vol.1, No.2, pp.119-130, 1988.

[10] K.Fukushima, "Recognition of partly occluded patterns: a neural network model," Biological Cybernetics, Vol.84, No.4, pp.251-259, 2001.

[11] J.R.Williamson, "Self-organization of topographic mixture networks using attentional feedback," Neural Computation, Vol.13, pp.563-593, 2001.

[12] K.Fukushima, "Use of non-uniform spatial blur for image comparison: Symmetry axis extraction," Neural Networks, Vol.18, No.1, pp.23-32, 2005.

[13] I.Hayashi and J.R.Williamson, "Acquisition of fuzzy knowledge from topographic mixture networks with attentional feedback," Proceedings of the International Joint Conference on Neural Networks (IJCNN'01), pp.1386-1391, 2001.

[14] I.Hayashi and H.Maeda, "A formulation of fuzzy TAM network with Gabor type receptive fields," Proceedings of the 4th International Symposium on Advanced Intelligent Systems (ISIS2003), pp.620-623, 2003.

[15] I.Hayashi, H.Maeda and J.R.Williamson, "A formulation of receptive filed type input layer for TAM network using Gabor function," Proceedings of 2004 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE2004), No.1335, 2004.

[16] I.Hayashi and J.R.Williamson, "Orientation selectivity by TAM network using Gabor function type receptive filed," Proceedings of the 5th International Conference on Recent Advances in Soft Computing (RASC2004), pp.122-127, 2004.

[17] I.Hayashi and J.R.Williamson, "A Study of Orientation Selectivity of TAM Network Incorporated Receptive Field Structure," Proceedings of the 3rd International Symposium on Autonomous Minirobots for Research and Edutainment (AMiRE2005), pp.187-192, 2005.