International Journal of Innovative Computing, Information and Control
Volume **3**, Number **4**, August **2007**

ICIC International ©2007 ISSN 1349-4198
pp. 789–798

HANDWRITTEN CHARACTER RECOGNITION BASED ON RELATIVE POSITION OF LOCAL FEATURES EXTRACTED BY SELF-ORGANIZING MAPS

KEIICHI HORIO AND TAKESHI YAMAKAWA

Graduate School of Life Science and Systems Engineering
Kyushu Institute of Technology
2-4 Hibikino, Wakamatsu, Kitakyushu 808-0196, Japan
{ horio; yamakawa }@brain.kyutech.ac.jp

Received October 2006; revised February 2007

ABSTRACT. *This paper describes a new pattern recognition method which is based on relative position of local features. We use a self-organizing map to detect a position of features, thus the relative position of them are automatically defined based on arrangement of the competing units. The local features are detected using filter constructed by adaptive subspace self-organizing maps.*

Keywords: Pattern recognition, Self-organizing map, Local feature, Relative position

1. **Introduction.** Many researchers have reported various pattern recognition methods so far, and recently machine learning techniques such as neural networks, support vector methods, convolutional nets and so on have become important roles in the field of handwritten character recognition [1]-[4]. Elastic matching (EM), which is also called deformable template, flexible matching or nonlinear template matching, has been gathering attention [5]. In EM, one character image 'A' is treated like a "rubber sheet" and fitted to another character image 'B', which is usually template, as close as possible. Thus EM is defined as an optimization problem with respect to a linear or nonlinear pixel-to-pixel mapping from 'A' to 'B'. Recognition by EM is robust to deformations of handwritten characters. However, EM requires high computational cost, because optimization problem, which is to find an ideal transformation of character image 'A', should be solved.

On the other hand, a visual system of a human has advantages concerning robustness and adaptability. In the human visual system, simple features in local areas, i.e. line orientation, are extracted in V1, and extracted features are integrated in higher brain [6]-[8]. A neocognitron which is one of the most famous neural network models was proposed in accordance with this biological knowledge [9]. But it is known that the neocognitron includes complex network structure and difficulty in tuning parameters.

In this paper, we propose a new handwritten character recognition method in which a concept of EM and the biological knowledge of the human visual system are adopted. Here, the concept of EM means that the character image 'A' is transformed to match it with the template image 'B' with keeping its topology. To realize such transformation, a self-organizing map (SOM) [10] is employed. Some gazing points in the image and their relative position can be extracted using SOM. Then an essential feature around each gazing point, which is a universal feature such as a line orientation, is extracted using

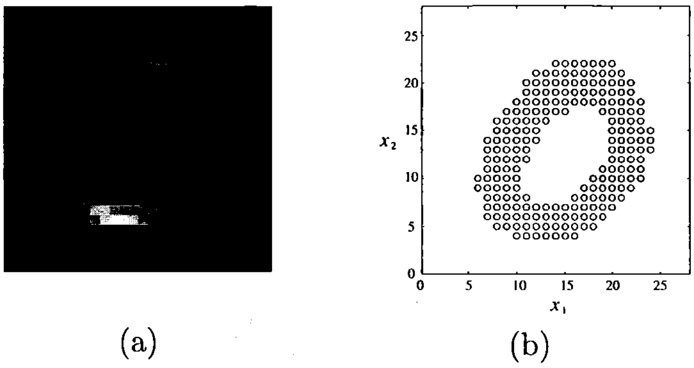

FIGURE 1. An example of feature extraction from an image. (a) Image to be recognized. (b) Set of coordinate vector extracted from the given image.

adaptive subspace SOM (ASSOM) [11]. A set of the essential features is defined as an overall feature of the given image. Finally, template matching is achieved using the overall features of the given image and the template image.

Although EM includes optimization problem to realize an ideal linear or nonlinear mapping from 'A' to 'B', only learning of SOM is needed in the proposed method. It facilitates a reduction of computational cost. Furthermore, EM usually uses pixel-to-pixel matching in calculating a distance between two images. On the other hand, feature-to-feature matching is employed in the proposed method. Thus it is expected that the recognition performance can be improved.

In the Section 2, a key idea of the character recognition based on relative position of local features is introduced, and its application to handwritten character recognition is described in detail in Section 3. In the Section 4, handwritten character recognition simulations using MNIST database [12] are explained, and performances of the proposed method are discussed. Finally we give some conclusions in Section 5.

2. **Pattern Recognition Based on Relative Position of Local Features.** It is known that orientation selectivity cells which extract lines with specific gradient exist in a primary visual cortex. The lines, *i.e.* extracted local features, are integrated in higher visual cortex, and the patterns are recognized. In the recognition phase, absolute positions of the local features in the image are not significant, and relative position of the local features is very important. By considering this biological knowledge, we propose a new pattern recognition method. In the proposed method, some gazing points which include specific features are extracted firstly. Secondly the local features around the gazing points are extracted. Thus the local features and their relative position is extracted by these processes. Only the relative position of the features is used in the proposed method, thus it is expected that recognition performance is not affected by pattern transformation such as a scaling or shifting.

2.1. **Extraction of gazing points.** In this paper, handwritten digits are used for patterns to be recognized. Figure 1(a) shows an example of handwritten digit image. In many pattern recognition methods, the image is represented by one vector called feature vector. In the simplest way, each element of the feature vector is represented as intensity

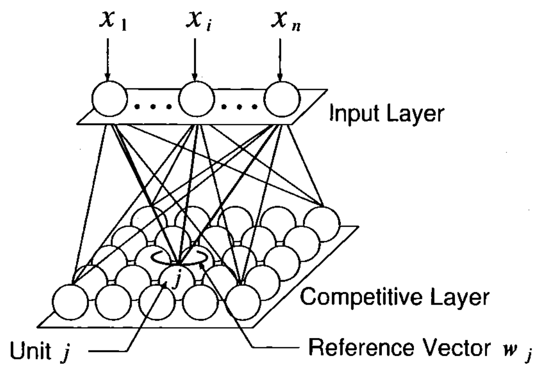

FIGURE 2. Structure of SOM.

of the corresponding pixel. That is, the feature vectors are represented as P^2-dimensional vectors, if the size of image is $P \times P$ pixels. In this case, transformations such as shifting and scaling affect a recognition performance critically, because it is very difficult to identify a change of vector in high dimensional ($P \times P$) vector space. To cope with these problems, many preprocessing methods including feature extraction have been proposed [13], however, conclusive methods do not appear so far.

In this study, we define a set of feature vectors as an overall feature of a given image. Individual feature vectors should be extracted from local images around gazing points in the given image. It might be common sense that gazing points, *i.e.* important points for recognizing a character, are on the area of which intensity is not zero shown in Figure 1(b). In Figure 1(b) each vector is 2-dimensional vector and its elements represent coordinate of the pixel in the image of which intensity is not '0'. In the image shown in Figure 1(a), black pixel means that its intensity is zero. These vectors should not be candidates for gazing points.

The gazing points are selected from the candidates shown in Figure 1(b). Furthermore the relative position of the gazing points is very important as mentioned above. Thus SOM is employed to realize gazing point extraction.

2.1.1. *Self-organizing map.* SOM is one of the most major neural network models and it has been successfully applied to pattern classification, data analysis and so on. A structure of SOM is shown in Figure 2. SOM consists of two layers, the input and the competitive layers including n and N units, respectively. The competing units are usually arranged on the 2-dimensional competitive layer. The unit j on the competitive layer is connected to the units on the input layer through reference vector $\boldsymbol{w}_j \in \mathrm{R}^n$.

Let us define a set of input vectors $X = \{\boldsymbol{x}_l \in \mathrm{R}^n | l = 1, \cdots, L\}$. An input vector is selected from the set of input vectors and applied to the input layer. Then the best matching unit (BMU) which has the closest reference vector to the input vector is assigned. The reference vectors of not only the BMU but also its neighboring units are updated to be closer to the input vector. By repeating this updating process corresponding to all input vectors, statistical characteristics and topology of the input vectors are detected. In other words, the quantized vectors of input vectors, *i.e.* typical vectors, are represented as reference vectors with keeping their topology defined by an arrangement on the competitive layer.

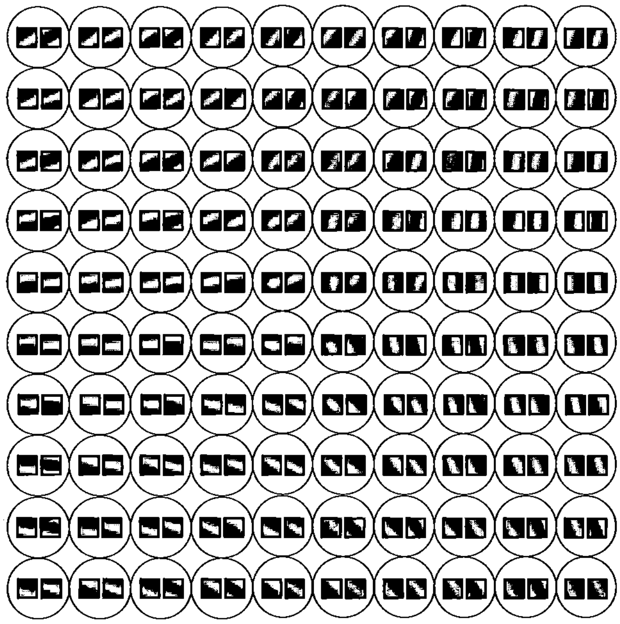

FIGURE 3. Learning result of ASSOM. Each unit responds to images of lines with specific orientation. Units are arranged based on the similarity of orientation.

2.1.2. *Gazing points extracted by SOM.* In this study, SOM is used to extract the gazing points which are important points for recognizing characters. As mentioned above, the handwritten characters are represented by the set of coordinate vectors in the campus, and the vectors should be candidates of gazing points. By using SOM in which the coordinate vectors are used as the set of input vectors, the positions of the typical features on the images are detected by sets of reference vectors. Furthermore the relative position of them is preserved, because arrangement of the competing units is fixed on the competitive layer.

2.2. **Extraction of local features.** Local features should be defined as patterns of local images around gazing points extracted using SOM. However, the important information of the local features is angle of lines, and the minute offset of the lines should be neglected. To achieve such feature extraction ASSOM is used.

ASSOM is the extension of SOM. In SOM, each competing unit includes a reference vector. On the contrary each unit of ASSOM includes linear subspace spanned by some basis vectors. The periodic patterns which are different regarding a phase are represented by the one linear subspace. It means that the patterns are classified on the competitive layer regarding only orientation of the patterns.

In order to use ASSOM as an extractor of local features around gazing points, a learning of ASSOM is achieved using 7×7 images with various orientation lines. Here, the number of units on the competitive layer is 10×10, and the number of basis vectors included in the unit for spanning the linear subspace is two. Figure 3 shows the learning result of ASSOM. Each circle represents the unit on the competitive layer, and two images in each circle are its two basis vectors. For example, the upper left unit represents the image with rising diagonal stroke from bottom left to top right. In other words, this upper left unit strongly responds to the image with rising diagonal stroke from bottom left to top right

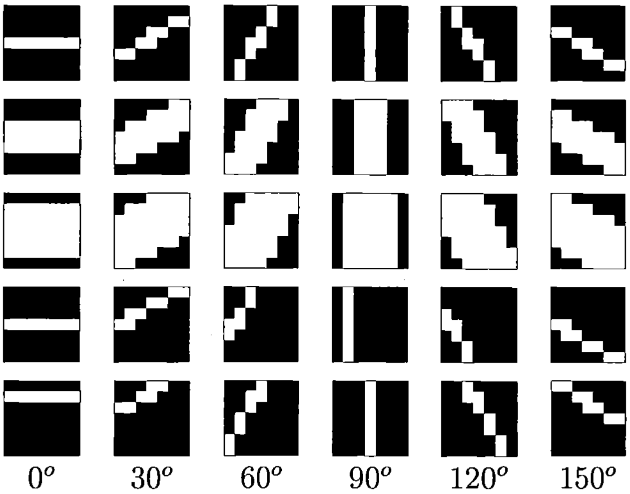

FIGURE 4. Testing images (7×7 pixels) including lines with various widths, shifting and orientation.

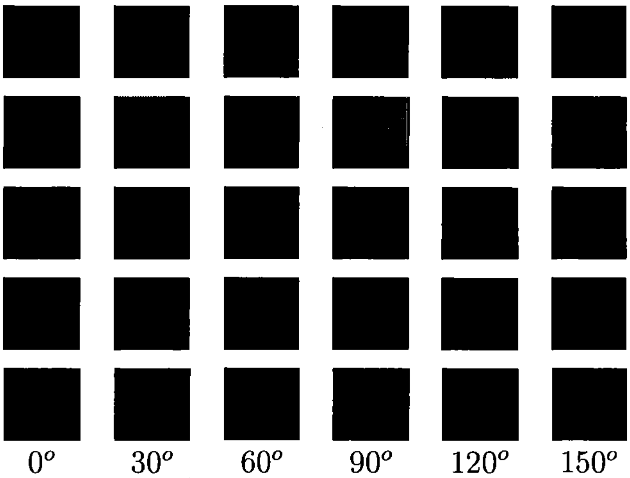

FIGURE 5. Responses of ASSOM (10×10 units shown in Figure 3) for testing images shown in Figure 4.

It is shown that images with similar orientation are represented by the units arranged nearby each other on the competitive layer.

As testing images, lines including various widths, shifting and orientation are prepared as shown in Figure 4. The top images of each column show basic images with typical six orientations. Images in the second and third rows represent thick lines of the basic images. And images of the forth and fifth rows are left-shifted and upward-shifted of the basic images. Figure 5 shows the responses of ASSOM for the testing images shown in Figure 4. Each image in Figure 5 represents a response of the competing units of ASSOM shown in Figure 3, thus each image is 10×10 pixels. Intensity means the strength of the responses. The position of each image in Figure 5 corresponds to that of each image in Figure 4. For example, upper left image in Figure 5 means a response of ASSOM when the upper left image in Figure 4 is applied. Left part of the competitive layer is activated. From Figure

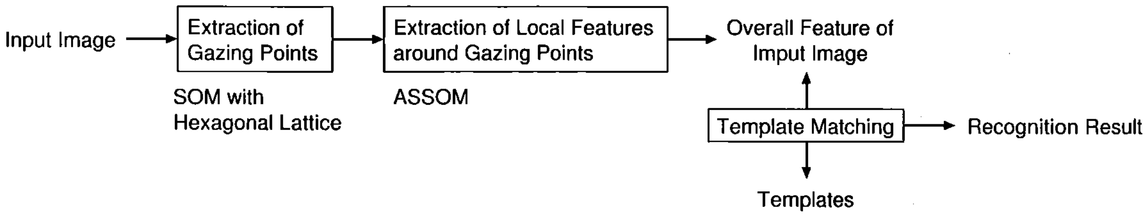

FIGURE 6. Proposed handwritten digit recognition system.

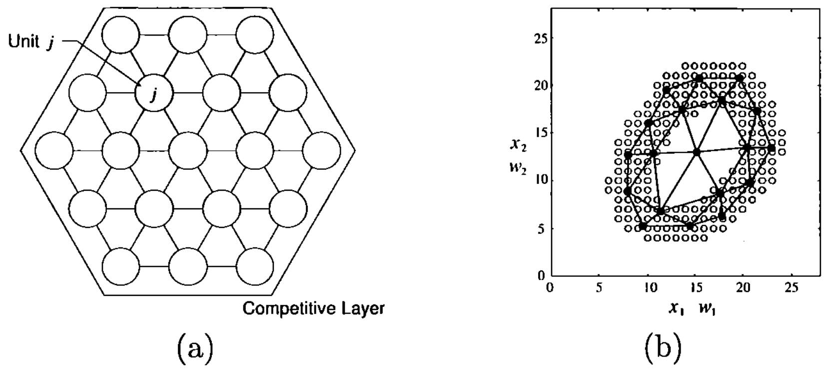

FIGURE 7. (a) Arrangement of the units on the competitive layer and (b) an example of extraction of gazing points.

5, it is shown that ASSOM extract features which are specialized in line orientation. In other words, images including shifting or difference widths indicate similar responses of ASSOM, although the strength of the responses is different. The responses of the units, where the local image around the gazing point is applied, are used as the local feature.

3. **Application to Handwritten Digit Recognition.** In this paper, the proposed pattern recognition algorithm is applied to a handwritten digit recognition problem. Figure 6 shows the proposed handwritten digit recognition system. This system is based on template matching. However, it is peculiar to the way of extracting features of an inputted image. As mentioned in the previous section, some gazing points and their relative position are extracted by SOM at first, and the local features around the gazing points are extracted by ASSOM. The responses of ASSOM for local images around the gazing points are employed as a overall feature of the inputted image, and a template matching is finally achieved for recognizing the inputted character image.

In SOM for extraction of the gazing points, 19 units are arranged on the 2-dimensional competitive layer in hexagonal lattice structure as shown in Figure 7(a). By arranging the units in this form, it is expected that robustness for a rotation of the character is realized. In Figure 7(a), circles are units and lines between units represent neighboring units. As an example, learning of SOM, in which the set of vectors shown in Figure 1(b) is used as the set of input vectors, is achieved, and Figure 7(b) shows the learning results. In Figure 7(b), white and black circles are the input vectors and the reference vectors, respectively. Lines between reference vectors representing neighbors correspond to that

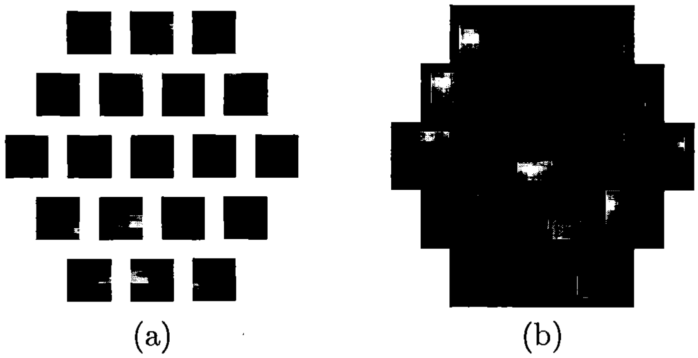

FIGURE 8. The local images and features. (a) 19 local images around 19 gazing points of the image shown in Figure 1(a) extracted by SOM. (b) The responses of ASSOM for corresponding local images.

on the competitive layer shown in Figure 7(a). The positions of the reference vectors indicate the gazing points, and lines between them represent the irrelative position.

The local images around the gazing points (7×7 pixels) extracted by SOM are shown in Figure 8(a). The arrangement of the local images corresponds to that of the units on the competitive layer. To extract essential features, *i.e.* line orientation, each local image is applied to ASSOM after learning shown in Figure 3. The responses of ASSOM for the corresponding local images shown in Figure 8(a) are shown in Figure 8(b), and the set of responses are defined as a overall feature of the inputted character image. In the proposed method, the relative position of the local features (responses of ASSOM for local images) is employed as the overall feature. Thus it is easily understood that a similar feature can be extracted even if characters are transformed such as shifting or scaling.

A template matching is employed to recognize the inputted image. Images for making the templates are processed mentioned above, and their overall features are memorized as templates. The testing image is also processed and the similarities overall features between testing images and those of templates are calculated. As a result, recognition of the inputted image is achieved.

4. **Simulation Results.** We have chosen the MNIST database of handwritten digits [12] in order to benchmark the proposed method and compare it to other typical methods.

4.1. **MNIST database.** The MNIST database consists of 60,000 training samples from approximately 250 writers and 10,000 test samples from a disjoint set of 250 other writers. We used the original 784-dimensional dataset which resembles 28×28 pixel grey level images of the handwritten digits.

4.2. **Performances.** To make the templates 5 images for each class were selected from training samples using k-means method. Figure 9 shows the 50 images prepared for the templates. The processes described in the previous section were done to make the templates, *i.e.* overall features of images. 10,000 testing images were tested, and the recognition rate was 90.0%. When the ordinary template matching in which the image data was directly used is tested, recognition rate was 84.3%.

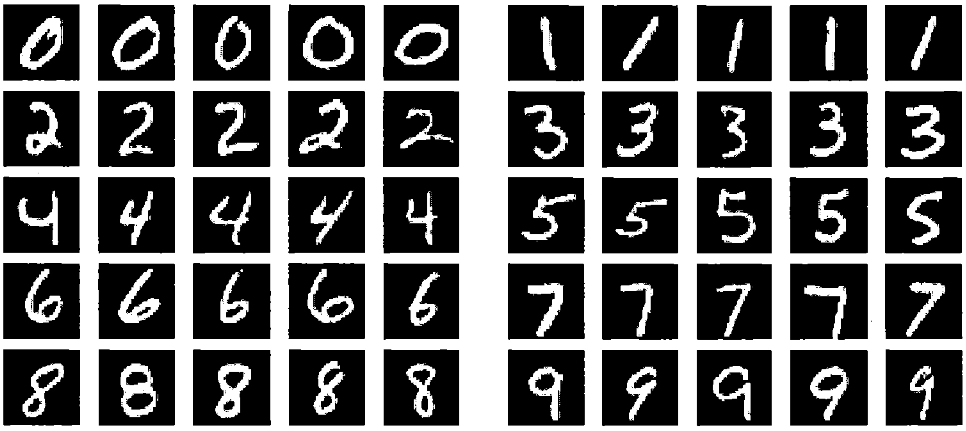

FIGURE 9. Images for making templates. 5 images are selected for each class by k-means method.

TABLE 1. Recognition rate(%) for 10,000 testing images corresponding to number of templates k for each class.

k	1	5	10	20
TM	66.8	84.3	87.9	90.2
GP+TM	56.7	66.9	72.5	77.7
GP+ASSOM+TM	76.5	90.0	91.3	93.2

Recognition rates corresponding to the number of template are shown in Table 1. In the table, "TM" is the ordinary template matching using the image data directly. In "GP+TM", the gazing points is extracted by SOM and template matching using local images around them are achieved. "GP+ASSOM+TM" is the proposed method. From the results it is shown that the proposed method performs highly accurate recognition even if the number of template is small. The recognition rate of the proposed method with only 5 templates is almost the same as that of the ordinary template matching with 20 templates. Recognition rates of GP+TM is bad. It should be a reason that the shift of the patterns in the local image is a major factor for recognition. Some examples of testing images which are not correctly recognized by the proposed method are shown in Figure 10. It is shown that these images are difficult even for a human to recognize correctly.

As mentioned above, in the proposed method the relative position of the local features is used for recognition, thus it is expected that recognition performance is not affected by the scaling or shift of the patterns. To verify this, some testing images which are scaled down, scaled up or rotated is prepared. Table 2 shows the recognition rates for the transformed images. Here the number of templates for each class is 20. It is shown that the recognition rates of the ordinary template matching method drastically become bad. Recognition rate can be improved when the images with various transformations are employed as templates. However, a large number of templates are required to realize this. On the other hand, the performance of the proposed method does not deteriorate even if the patterns are transformed. It is very important for handwritten character recognition to be robust for such transforms.

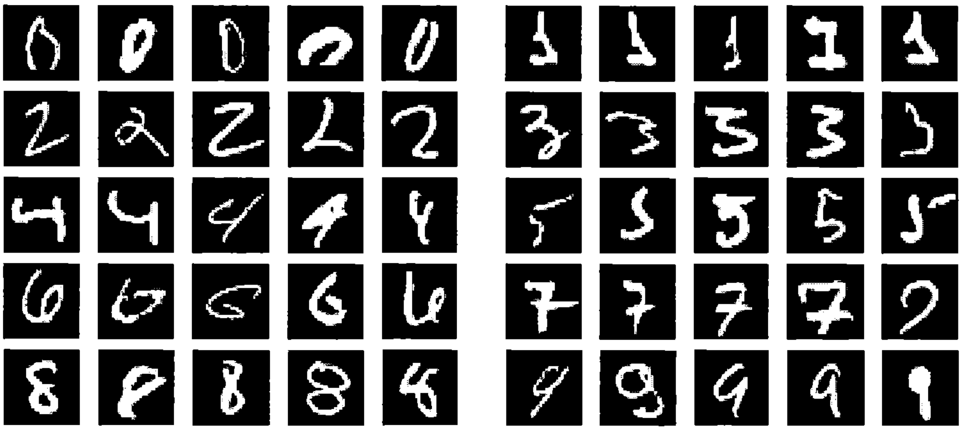

FIGURE 10. Some examples of the images which were not successfully recognized.

TABLE 2. Recognition rate for 10,000 testing images with transformation (%).

	Standard	Scaling Down	Scaling Up	Rotation(30deg.)
TM	90.2	57.3	27.0	53.5
Proposed	93.2	84.3	91.7	68.8

5. **Conclusions.** In this paper, we proposed the new pattern recognition method which was inspired by biological knowledge of the visual system of humans. In the proposed method, the local features and their relative position were used to recognize the pattern. To extract the gazing points and the local feature around them, SOM and ASSOM were employed, respectively. It was verified by simulation that the proposed method has robustness for scaling and shifting of the patterns. For the rotated images, the recognition rate of the proposed method deteriorates. The reason for this is that the arrangement of the units in the competitive layer of SOM (Figure 7) is not suitable for 30 degree rotation. To improve the performances, the neighboring relation of the units on the competitive layer should be reconsidered, and the hybrid system of SOM and ASSOM should be realized to cope with the problem of unstable learning.

Acknowledgment. The authors would like to thank Mr. Akira Aikawa for helping experiments presented in this study. This work was supported in part by the 21st Century Center of Excellence Program (center #J19), "World of Brain Computing Interwoven out of Animals and Robots (PI: T. Yamakawa)" granted in 2003 to Department of Brain Science and Engineering, Kyushu Institute of Technology by Japan Ministry of Education, Culture, Sports, Science and Technology.

REFERENCES

[1] Vapnik, V. N., *Statistical Learning Theory*, John Wiley & Sons, 1998.
[2] Bishop, C. M., *Neural Networks for Pattern Recognition*, Oxford University Press, 1995.
[3] LeCun, Y., L. Bottou, Y. Bengio and P. Haffner, Gradient-based learning applied to document recognition, *Proceedings of the IEEE*, vol.86, no.11, pp.2278-2324, 1998.
[4] Plamondon, R. and S. N. Srihari, Online and off-line handwriting recognition: A comprehensive survey, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.22, no.1, pp.63-84, 2000.

[5] Uchida, S. and H. Sakoe, A survey of elastic matching techniques for handwritten character recognition, *IEICE Trans. on Information and Systems*, vol.E88-D, no.8, pp.1781-1790, 2005.
[6] Hubel, D. H. and T. N. Wiesel, Receptive fields, binocular interaction and functional architecture in the cat's visual cortex, *Journal of Physiology*, vol.160, pp.106-154, 1962.
[7] Hubel, D. H. and T. N. Wiesel, Receptive fields and functional architecture of monkey striate cortex, *Journal of Physiology*, vol.195, pp.215-243, 1968.
[8] De Valois, R. L. and K. K. De Valois, *Spatial Vision*, Oxford University Press, 1990.
[9] Fukushima, K., S. Miyake and T. Ito, Neocognitron: A neural network model for a mechanism of visual pattern recognition, *IEEE Trans. on Systems, Man and Cybernetics*, vol.13, pp.826-834, 1983.
[10] Kohonen, T., *Self-Organizing Map*, Springer-Verlag, 1995.
[11] Kohonen, T., Emergence of invariant-feature detectors in the adaptive-subspace self-organizing map, *Biological Cybernetics*, vol.75, pp.281-291, 1996.
[12] The MNIST database of handwritten digits, http://yann.lecun.com/exdb/mnist/
[13] Trier, O. D., A. K. Jain and T. Taxt, Feature extraction methods for character recognition - A survey, *Pattern Recognition*, vol.29, no.4, pp.641-662, 1996.