

Hand Shape Recognition based on Hyper Column Model

Shimada, Atsushi
Department of Intelligent Systems, Kyushu University

Tsuruta, Naoyuki
Department of Electronics Engineering and Computer Science, Fukuoka University

Taniguchi, Rin-ichiro
Department of Intelligent Systems, Kyushu University

<https://hdl.handle.net/2324/4032>

出版情報 : Proceedings of the Joint Symposium of ISEE2005 and the Core University Program on NGI, pp.114-117, 2005-09

バージョン :

権利関係 :

Hand Shape Recognition based on Hyper Column Model

Atsushi Shimada

Department of Intelligent Systems
Kyushu University

Email: atsushi@limu.is.kyushu-u.ac.jp

Naoyuki Tsuruta

Department of Electronics Engineering
and Computer Science Fukuoka University

Email: tsuruta@tl.fukuoka-u.ac.jp

Rin-ichiro Taniguchi

Department of Intelligent Systems
Kyushu University

Email: rin@limu.is.kyushu-u.ac.jp

Abstract—In this paper, we propose a recognition method of hand shapes using Hyper-Column Model (HCM). HCM is a model to recognize images, and consists of Hierarchical Self-Organizing Maps (HSOM) and Neocognitron (NC). HCM complements disadvantages of HSOM and NC, and inherits advantages from them. There is a problem, however, that HCM does not suit general image recognition since its learning method is an unsupervised one with competitive learning which is used by Self-Organizing Map (SOM). Therefore, we extended HCM to a supervised learnable model with an associative memory of SOM. We have found that an ability of HCM with supervised learning is superior to the one with unsupervised learning.

I. INTRODUCTION

Artificial neural networks (ANNs) were traditionally applied to recognition of handwritten characters and to pixel wise clustering in remote sensing, in which the dimensionality of the input data was small. For these applications, it was reported that ANNs are superior to the principal component analysis of the linear method and to the k-nearest neighbor approach of the nonparametric method. Recently, applications of ANNs have expanded into general image recognition such as face recognition and visual surveillance. In these applications, there are three problems.

- 1) In spite of the high dimension of the input images, the regions of object parts occupy small areas of the whole space. Therefore, it is necessary to reduce the dimension to eliminate the redundancy and to allow the network to learn according to the region actually expanding the object.
- 2) The absolute dimensionality of the image region is very large even if the redundancy is eliminated, since images have large variations as a function of object locale, illuminant, and so on. Therefore, techniques to reduce the system size to a realizable scale are necessary.
- 3) In such cases, ANNs need a very large number of neurons. Therefore, learning methods which depend on the initial states of the connection weights cannot perform well, even if a network model such as the multilayered perceptron is theoretically very powerful. Thus, learning methods whose performance does not depend on the initial states of connection weights are needed.

Tsuruta proposed Hyper-Column Model (HCM)[1] which is a new image recognition model combining Hierarchical Self-Organizing Map (HSOM)[2] and Neocognitron (NC)[3]. The

learning method is quite simple with unsupervised learning, but it is powerful enough since it does not depend on the initial states of the connections. In addition, HCM can reduce the dimensionality of general images and can perform better than HSOM and NC, since HCM overcomes the disadvantages of HSOM and NC and inherits their advantages directly. HCM cannot, however, show high ability when the boundary between categories is very complicated, such a case that objects of different categories have similar features. In general image recognition problems, it is known that the recognition ability with supervised learning method is superior to the one with unsupervised learning method.

In this paper, we propose a supervised learning method in HCM by using an associative memory of SOM. Section II presents a description of HCM. In section III, we propose a supervised learning method in HCM. Experimental results of image recognition are presented in section IV. In addition, it is shown that HCM with supervised learning is superior to the one with unsupervised learning.

II. HYPER-COLUMN MODEL

A. Self-Organizing Map (SOM)

Self-Organizing Map (SOM) is one of the most widely used artificial neural network algorithms that uses unsupervised learning. This algorithm was proposed by Kohonen[4]. As shown in Fig. 1, SOM is usually represented in a two dimensional neural network. Each input vector \mathbf{I} has connections with all of neurons ($1 \cdots u \cdots N$). Each neuron has a weight vector \mathbf{W}_u . In training of SOM, a neuron c that satisfies equation (1) is chosen as the winner neuron and then weight vectors are updated by using equation (2).

$$\|\mathbf{I} - \mathbf{W}_c\| = \min_u (\|\mathbf{I} - \mathbf{W}_u\|) \quad (1)$$

$$\mathbf{W}_u(t+1) = \mathbf{W}_u(t) + h_{cu}(\mathbf{I}(t) - \mathbf{W}_u(t)) \quad (2)$$

h_{cu} is a neighborhood kernel defined by equation (3).

$$h_{cu} = \alpha(t) \cdot \exp\left(-\frac{\|r_c - r_u\|^2}{2\sigma^2(t)}\right) \quad (3)$$

With increasing $\|r_c - r_u\|$ and t , h_{cu} converges to zero. $\alpha(t)$ is a monotonically decreasing function of t ($0 < \alpha(t) < 1$), and $\sigma^2(t)$ defines the width of the kernel.

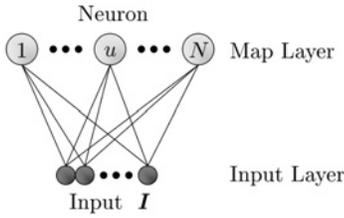


Fig. 1. Structure of SOM

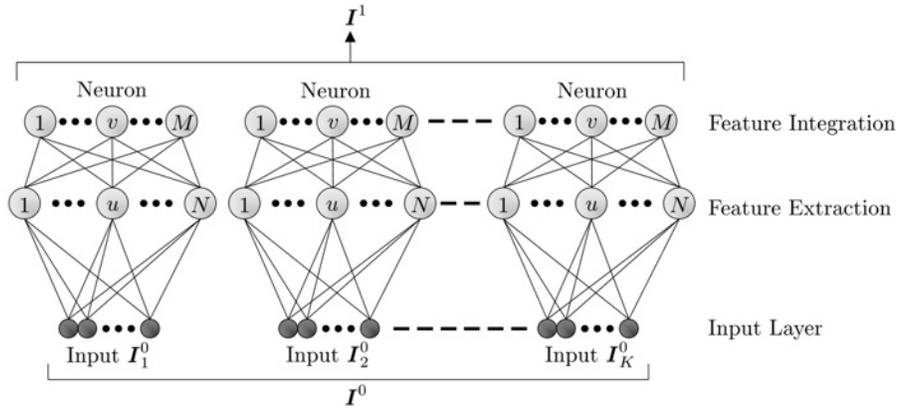


Fig. 2. Structure of HCM

B. Hierarchical Self-Organizing Map (HSOM)

Hierarchical Self-Organizing Map (HSOM) proposed by Lampinen[2] is a two layer of SOM network. For each input vector I , the winner neuron is chosen from the first layer map and its index is given to the second layer.

The advantages of HSOM are summarized into the following points.

- 1) HSOM solves the complex region problem, when a sufficiently dense data set is given for continuous variation of the input data.
- 2) HSOM can be a good preprocess to resolve the dimensional reduction problem.

However, HSOM has the following three disadvantages.

- 1) Image recognition methods based on HSOM are regarded as a memory-based method. Therefore, in case where the dimension and the size of data distribution are large, the network size increases.
- 2) Maps are organized according to the continuity of data distribution in the space defined by the distance between images using template matching. Therefore, it is hard to organize maps in cases where the distance between images does not smoothly vary with locale and scale of target object. In such cases, a heuristic, such as image blurry, must be adapted or the number of training samples must be increased so that the variations of distance are smooth enough.
- 3) The recognition method is also based on the distance of images and, therefore, is regarded as one of nearest neighbor method. Therefore, a high accurate segmentation of target region and a normalization of locale and size should be needed.

C. Neocognitron (NC)

Neocognitron (NC) is proposed by Fukushima[3] as a hierarchical network consisting of several layers of neuron-like cells. The lowest stage of the network is the input layer. Each succeeding stage has a layer consisting of cells called S-cells followed by another layer of cells called C-cells. S-cells are the feature-extracting cells. The C-cells are inserted on the network to allow for positional errors in the features.

The “structural” advantages of NC are summarized in the following two points, which can alleviate the disadvantages of HSOM.

- 1) Every feature map is rather small since NC has a hierarchical structure of divide-and-conquer type. This characteristic resolves the disadvantage 1 of HSOM.
- 2) Toward the disadvantage 2 and 3 of HSOM, NC does not need any preprocesses of blurring nor normalization due to its shift invariant recognition mechanism. NC, also, needs small number of training sample data because of its learning mechanism.

The original NC has the following two disadvantages when it is directly applied to general image recognition. The advantages of HSOM match with those disadvantages.

- 1) The learning process of NC is strongly depends on the initial state of weight vectors, and the feature extraction does not often perform well due to its simple competitive learning without neighborhood learning.
- 2) The shift integration layer integrates only shifted patterns, and does not requantize feature spaces. Therefore, its dimension reduction is not enough for variation of target shape, which requires large number of network neurons.

D. Hyper-Column Model (HCM)

Hyper-Column Model (HCM) is proposed for visual recognition of objects with variations in its position, size, and orientation[1]. HCM is a kind of competitive neural networks with unsupervised learning. The network is composed of hierarchical layers derived from NC by replacing the unit cell plains (each C-cells and the lower directly connecting S-cells) in NC with two-layers HSOM. These HSOM cell plains allow features extraction through the first SOM layer followed by features integration in the second SOM layer, as shown in Fig. 2. This feature integration process allows the cell plains to integrate more features in lower number of neurons. NC does not perform such feature integration process.

HCM uses unsupervised learning algorithm to construct its feature map. The training is performed layer by layer from the bottom layer to the upper layer. The feature extraction layer in

the L th HCM layer receives the data from the $L - 1$ th HCM layer as the input. For each input vector, the winner neuron is chosen by using equation (1) and the weight vectors are updated by using equation (2). In the feature integration layer, the winner position index of the previous layer is considered as the input to this layer, and the winner neuron is chosen by using equation (1). The weight vectors are updated by using equation (2).

In the recognition phase, the competition also applies layer by layer starting from the bottom. The network can use the same competition algorithm used in the learning phase. The winner position index of all HSOMs in the first layer (feature extraction layer) are treated as the input of the second layer (feature integration layer), and so on until the last layer.

III. SUPERVISED LEARNING IN HCM

The structure of HCM is similar to the one of NC, but HCM uses unsupervised learning algorithm that is used in SOM. In general image recognition problems, it is known that the recognition ability with supervised learning method is superior to the one with unsupervised learning method. Therefore, introducing a supervised learning method into HCM is needed.

We introduce a method of supervised learning in SOM proposed by Ichiki[5] into HCM. In this learning method, we can regard the input vector as the one which is composed of different two components; an input part \mathbf{I} and an output part \mathbf{T} . As a result, the networks can be considered as a supervised learning machines.

$$\mathbf{X} = a \begin{pmatrix} \mathbf{I} \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ \mathbf{T} \end{pmatrix} \quad (4)$$

where $a > 1$ in order that the input part can affect the information of the map more than the output part does. The weight vector \mathbf{W}_u is represented as follows,

$$\mathbf{W}_u = \begin{pmatrix} \mathbf{W}_u^I \\ \mathbf{W}_u^T \end{pmatrix} \quad (5)$$

In recognition process, an input vector is \mathbf{I} . When input \mathbf{I} is input into the network, the neuron c is the winner neuron. An output \mathbf{O} is represented as follows, where weight vectors $\|\mathbf{W}_u\|$ are normalized to n .

$$\mathbf{O} = \mathbf{W}_c^T \times v \quad (6)$$

$$v = \frac{\sqrt{a^2 \|\mathbf{I}\|^2 + \|\mathbf{T}\|^2}}{n} \quad (7)$$

IV. EXPERIMENT

A. Condition

The input data were images of human hands which consist of 10 categories shown in Fig. 3. Each category has 10 images for training data, and 500 images for test data. The size of images was 100×100 , and each pixel had an 8-bit gray value. The overall view of HCM used in this experiment is shown in Fig. 4 and the construction of each layer in HCM is shown in table I. The neurons in each feature extraction layer were arranged in a ring shape. In experiment, HCM learns training

data with unsupervised learning and with supervised learning. We have applied the supervised learning with only the top layer of HCM, since the teacher signals, in the other layers, impede the integration of similar features which belong to different categories.

B. Learning result

The maps generated by each HCM are shown in Fig. 5. The horizontal axis shows the sample number. From number 0, each 10 samples belong to the same category. The vertical axis shows the neuron number. HCM with unsupervised learning cannot generate one neuron cluster for the whole of one category 1, 2, 6, 7 and 8. On the other hand, HCM with supervised learning can do it in most cases.

C. Recognition accuracy

Table II shows the recognition result for 5000 test images. The average rate of correct recognition with unsupervised learning was 72.5%. The images of category 7 tended to be misunderstood as category 2. There are some overlaps between the neurons which belong to category 2 and category 7 as can be expected from Fig. 5(a). The images which belong to category 9, also, tended to be misunderstood as category 6. In most images which were not recognized correctly, the position of thumb was nearby the palm. We think that HCM with unsupervised learning could not create an appropriate boundary between category 6 and category 9 since the images of category 9 are similar to the ones of category 6.

On the other hand, the average rate of correct recognition with supervised learning was 93.3%. The supervised learning method gave good results without decreasing the recognition rates.

V. CONCLUSIONS

We proposed a supervised learning method in Hyper-column Model (HCM) and applied to recognition of hand shapes. In the experiments described in this paper, the recognition ability of HCM with supervised learning was superior to the one with unsupervised learning. However, HCM has a problem that it runs out of neuron when the number of category increases. We are now researching for reconstruction of the network and recognition in complex backgrounds.

REFERENCES

- [1] Naoyuki Tsuruta, Rin-ichiro Taniguchi and Makoto Amamiya, "Hyper-column Model: A Combination Model of Hierarchical Self-Organizing Maps and Neocognitron for Image Recognition", *Systems and Computers in Japan*, Vol.31, No.2, pp.49–61, 2000.
- [2] J. Lampinen and E. Oja, "Clustering properties of hierarchical self-organizing maps", *Journal of Mathematical Imaging and Vision*, Vol.2, No.3, pp.261–272, 1992.
- [3] K. Fukushima, "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position", *Biological Cybernetics*, Vol.36, No.4, pp.193–202, 1980.
- [4] T. Kohonen, "Self-Organizing Maps", *Springer Series in Information Science*, 1995.
- [5] Ichiki H., Hagiwara M and Nakagawa M, "Kohonen feature maps as a supervised learning machine", *Proc. of International Conference on Neural Networks*, pp.1944–1948, 1993.

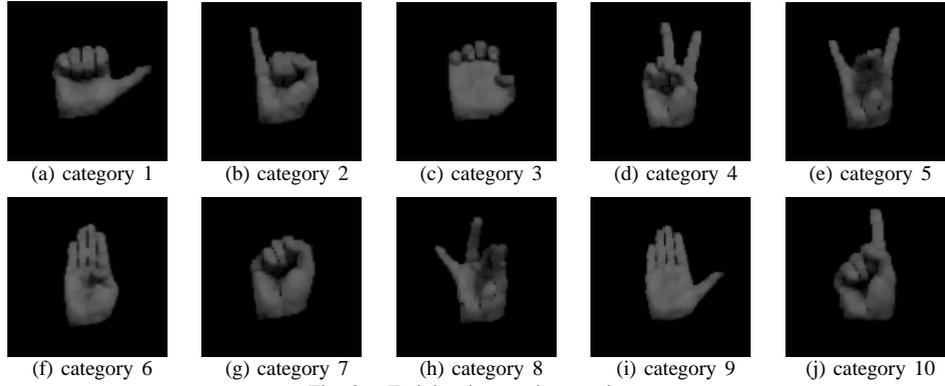


Fig. 3. Training images in experiment

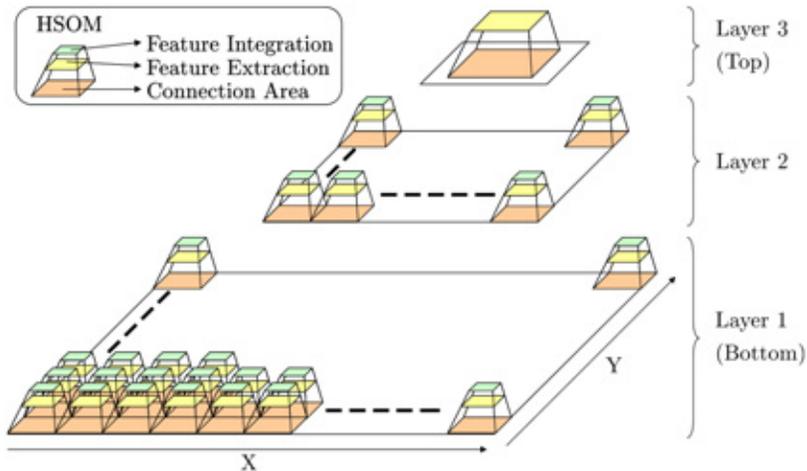
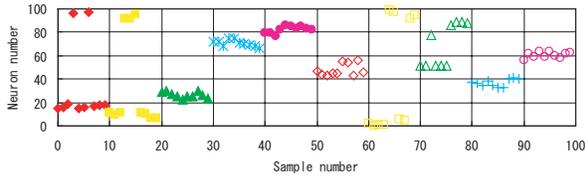
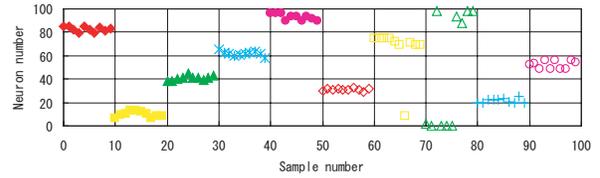


Fig. 4. Overall view of HCM used in experiment



(a) Unsupervised learning



(b) Supervised learning

Fig. 5. Correspondence between maps of HCM and sample data in experiment

TABLE I
CONSTRUCTION OF HCM USED IN EXPERIMENT

		HSOM		C.Area	Neurons
Layer 3 (Top)	Y	1	FIL	-	-
	X	1	FEL	8×8	100
Layer 2	Y	8	FIL	-	10
	X	8	FEL	5×5	100
Layer 1 (Bottom)	Y	20	FIL	-	5
	X	20	FEL	5×5	100

FEL ... Feature Extraction Layer
 FIL ... Feature Integration Layer
 C.Area ... Size of Connection Area

TABLE II
RATES OF CORRECT RECOGNITION FOR 10 CATEGORY TEST DATA (%)

Category	1	2	3	4	5	6	7	8	9	10	Average
Unsupervised learning	75.4	63.4	91.6	72.4	80.2	89.0	57.2	51.4	62.4	82.2	72.5
Supervised learning	98.4	91.4	98.4	98.2	89.2	90.4	98.2	97.2	80.8	91.0	93.3