'92 KACC 1992. 10. 19~21

# An Approach to Visual Pattern Recognition by Neural Network System

Yasuhiro HATAKEYAMA* and Yukinori KAKAZU**

*Division of Information Engineering
Graduate School of Engineering
Hokkaido University
N13-W8, Sapporo 060, JAPAN

**Department of Precision Engineering
Faculty of Engineering
Hokkaido University

*, **TEL. +08-11-736-3818
FAX. +08-11-758-1619


## ABSTRACT

In this paper, a visual pattern recognition system is proposed, which can recognize both a pattern and its location. This system, referred to as the expanded neocognitron, has the following capabilities: (1) A higher performance in extraction of features, and (2) A new capability for recognizing the locations of patterns. This system adopts the learning and recognizing mechanism of the neocognitron. First, the ability to classify pattern is enhanced by improving the mechanisms of feature extraction and learning algorithm. Second, the function of detecting the location of each pattern is realized by developing an architecture which does not reduce structure, i.e., the unit density is constant all the way from the input stage to the output stage.


## INTRODUCTION

Visual information processing is one of the important key techniques for intelligent systems like an intelligent robot which moves in an unknown environment. In such a case, the system should have high recognition ability for positional deviation and deformed patterns. Studies that intended to enable systems to obtain both position and direction of a target pattern from visual image input have been attempted from various approaches.

In the field of signal processing and character recognition, neural network systems realize two advantages, i.e., self-organization and generalization [CARP89]. In particular, the neocognitron has the high ability to recognize the shifted and distorted patterns [FUKU89]. Our approach is applying the neocognitron system to the visual pattern recognition by improving and expanding it.

The neocognitron has a hierarchical structure, and its pattern processing is carried out feed forward as shown in Figure 1. In this paper, the architecture for developing the visual pattern recognition is based on this system self-organized by unsupervised learning. This is because an adaptive and autonomous mechanism of recognition and real time method are required in intelligent vision systems.

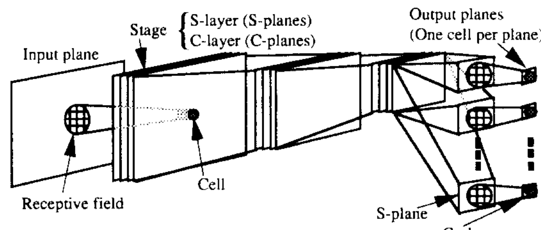


Figure 1. Overview of the hierarchical structure of the neocognitron.

There are two kinds of units, an S-cell having a feature extraction function and a C-cell with an blurring function. These cells independently form layers, a pair of which named the S-layer and C-layer, respectively. In particular, the S-cell with plasticity in learning is essential as a key in pattern recognition. In the proposed system, the conventional S-cell model is improved to be effective for learning and extracting a local feature.

## A VISUAL PATTERN

Considering visual pattern recognition, we assume that an input image contains patterns at the arbitrary positions in it. To classify the pattern from the visual image input, it is necessary that the system can recognize it without being affected by the its location. Therefore, the input image and the pattern in it need to be described by an expression with respect to position of unit.

Each layer of the neocognitron consists of multiple planes. Suppose the cells in each plane are allocated on a quadric lattice, then the $l$th layer $U^l$ is represented as

$$U^l = (U_1^l, \cdots, U_k^l, \cdots, U_{K_l}^l),\ \ U_k^l \subset \boldsymbol{R}^2, \quad k = 1, \cdots, K_l,\ \ l = 1, \cdots, L, \qquad \cdots (1)$$

where $U_k^l$ is the $k$th plane, $L$ and $K_l$ are the number of layers and planes, respectively. The cell which belongs to the $k$th plane of the $l$th layer is denoted as follows:

$$u^l = (\boldsymbol{x}, k)^l,\ \ \boldsymbol{x} \in U_k^l,\ \ k = 1, \cdots, K_l, \qquad \cdots (2)$$

where $\boldsymbol{x}$ represents the coordinate of the cell $u^l$ which has the receptive field for all the planes of the $l$-1th layer. Let $\boldsymbol{p}_k$ be a pattern in the receptive field on the $k$th plane, then the pattern $\boldsymbol{P}(u^l)$ that $u^l$ receives is denoted by,

$$\boldsymbol{P}(u^l) = (\boldsymbol{p}_1, \cdots, \boldsymbol{p}_{K_{l-1}}). \qquad \cdots (3)$$

Thus, the cell deals with the local patterns of the previous layer.

In the neocognitron system, S-layer is composed of a set of S-planes which are groups of cells that respond to the specified local feature.

## THE FEATURE EXTRACTION

After idealized training, The S-cell can extract the local feature precisely. However, effective learning is not always ensured in self-organization by unsupervised learning, because the characteristic of an insufficiently learned S-cell is different from one which is completely learned. Since it is usually assumed that the characteristics of any local features at any level of successive processes are unknown, we can not set the parameters properly. To solve this problem, we propose an improved S-cell which is highly sensitive and stable characteristic even in the training process [HATA92].

The connection coefficients of S-cell are variable and can be tuned by training. Consequently, the S-cell performs as a template of a feature. Let $\boldsymbol{a}_k$ and $b_k$ be the coefficients of the excitatory and inhibitory connections, respectively. The output of the proposed S-cell is expressed as

$$f_k(u) = \phi\left(r\left(\frac{\tau + (\boldsymbol{P}, \boldsymbol{a}_k)}{\tau + r' b_k(\boldsymbol{g}, (\boldsymbol{P}\cdot\boldsymbol{P}))^{1/2}} - 1\right)\right). \qquad \cdots (4)$$

$$\phi(\xi) = \begin{cases} \xi & ;\text{if } \xi \geq 0, \\ 0 & ;\text{otherwise}, \end{cases} \qquad \cdots (5)$$

where $r$ is the parameter for threshold and $r' = r/(r+1)$. $\boldsymbol{P}$ is a local pattern which the S-cell $u$ receives, and the vector $\boldsymbol{g}$ is a filter defined by the quadric gaussian distribution over the receptive field. The operator symbol (,) means calculation of the inner product between vectors, and $(\cdot)$ returns the vector whose elements are the product of the corresponding elements of two vectors.

In this proposed model, the bias $\tau$, can be tuned to be highly sensitive. Tuning the bias to a sufficiently small value, the characteristic does not change as appears in Figures 2. If $\tau$ = 1.0, then that model is equal to the conventional S-cell. In this case, its response depends on threshold and norm of local feature. On the other hand, if $\tau$ is sufficiently small, then the proposed S-cell becomes highly sensitive.

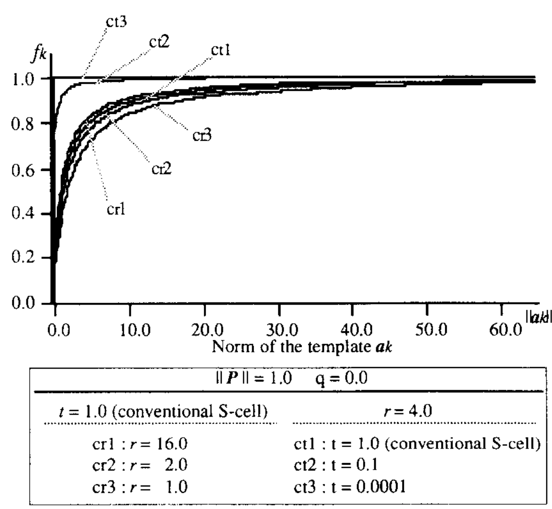


Figure 2. Stability of the proposed S-cell.

## THE COMPETITIVE LEARNING

With a competitive learning of the neocognitron there are two problems: First, the number of planes is decided before training and can not be changed; the number of planes equals the number of detectable features. However, it is unknown also that what kind of features there will be. Therefore, the conventional structure is limited as to the preset number of features that can be detected. Second, the conventional learning method can not distinguish between similar features.

To solve these problems, we realize a flexibility in the plane structure. Namely, we introduce a dynamics to the number of planes. Let $\Pi_l(t)$ be the number of planes in the $l$th layer at training time $t$. The dynamics of the plane structure is defined as,

$$\Pi_l(t+1) = \Pi_l(t) + 1\left[\min_k (G_k) - \gamma\right], \qquad \cdots (6)$$

where $\gamma$ is the threshold for the increment of the number of the planes, and 1[] means the Heaviside function. $G_k$, means degree of growth of the $k$th plane, which is defined as follows:

$$G_k = \frac{b_k}{b_k + 1}. \qquad \cdots (7)$$

According to this dynamics, the number of planes increases until no features are detected.

The competitive learning algorithm locally evaluates the activities of cells and detects the cell to be learned. In such detection of the maximum point, there is the difficulty that the feature which is the neighbor of the maximum point can not be detected. Therefore, it is necessary for the activities of the S-cells to be revaluated in order to carry out precise learning.

The algorithm to revaluate the local activities is performed by mutual inhibition between the planes. Let the revaluated cell be $u_\kappa$, then the revaluation is described by

$$f_k(u_\kappa) = f_k(u_\kappa) - \sum_{k'=1}^{K_l} d_{\kappa k'}(u_\kappa), \qquad \cdots (8)$$

$$d_{\kappa k'}(u_\kappa) = \begin{cases} \alpha_1(\boldsymbol{g}, \boldsymbol{P}_{k'}(u_\kappa)) & ; \text{if } k' \neq \kappa, \\ \alpha_2(G_\kappa - f_k(u_\kappa)) & ; \text{otherwise}, \end{cases} \qquad \cdots (9)$$

where $\alpha_1$ and $\alpha_2$ are constant. In the expression (9), if $k' \neq \kappa$, then $d_{\kappa k'}$ means the inhibition from the $k'$ th plane to the potential energy of the $k$th planes, and if $k' = \kappa$, then the activity itself is revaluated according to the degree of growth. This revaluation enhances the high activity cell sufficiently learned, and gives a chance to its neighboring cell. Figure 3 shows that the proposed algorithm can more precisely learn a lot of local features than the conventional one.

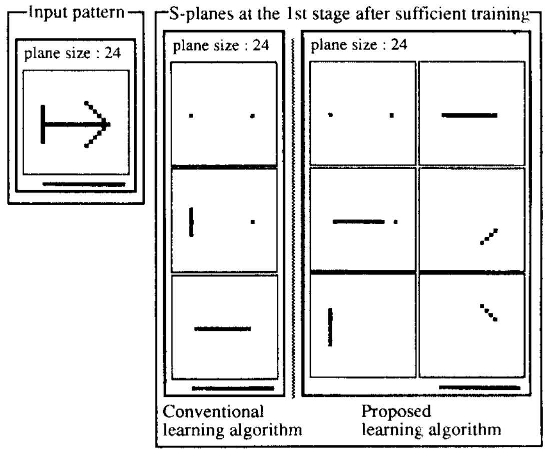


Figure 3. Comparison of learning results.

## DETECTING THE LOCATION OF PATTERN

In the visual pattern recognition, shift invariance is one of the most desirable characteristics. The neocognitron has the ability to perform this shift-invariant pattern recognition. However, it is pointed out that the performance of the neocognitron is not intrinsically shift invariant [BARN90]. The reducing plane structure causes this defect.

The conventional neocognitron is designed to integrate features by decreasing the density of the cells. On the other hand, our approach runs contrary to this idea. The size of the receptive field of the S-cell is increased according to the level of the layer.

By adopting this structure, the expanded neocognitron can recognize not only the pattern but its location. Moreover, it is able to classify the patterns on an input screen in one feed forward processing.

## EXPERIMENT

To observe the ability of the expanded neocognitron, An experiment of pattern recognition have been carried out. In this experiment, training patters are five geometrical shapes; a rectangle and four triangles having the individual directions. Table 1 shows the structure after training.

Four example images and the results of recognition are shown in Figure 4. (a) two shapes with noise, (b) four triangles which are different from each other in the direction, (c) two same triangles touched at a point, and (d) a rectangle and a triangle which coincide with each other in an edge. The output layer consists of five planes that correspond to each pattern. A dot in a plane presents the existence of the pattern at that point in the input screen. These recognition results indicate that the expanded neocognitron successfully performs as the visual pattern recognition system.

## CONCLUSIONS

To develop the visual pattern recognition system, we proposed the expanded neocognitron which was constructed by expanding the feature extraction mechanisms of the conventional neocognitron on the following points: (1) the highly sensitive S-cell, (2) new competitive learning algorithm, and (3) non-reducing plane structure. The proposed system has the following capabilities: (1) Higher in performance of feature detection, and (2) Capable of recognizing locations of patterns. The recognition abilities on above were demonstrated by the experiments of self-organization and pattern recognition.

## REFERENCES


[BARN90] Barnard, E., Casasent, D., "Shift Invariance and the Neocognitron," Neural Networks, Vol.3, No.4, pp.403-410 (1990).

[CARP89] Carpenter, G. A.: "Neural Network Models for Pattern Recognition and Associative Memory," Neural Networks, Vol.2, No., pp.243-257 (1989).

[FUKU89] Fukushima, K., "Analysis of the Process of Visual Pattern Recognition by the Neocognitron," Neural Networks, Vol.2, No.6, pp.413-420 (1989).

[HATA92] Hatakeyama, Y., Yukinori, K., "Analysis and Improvement of the Feature Detection Ability of Neocognitron for Adaptive Image Processing," Journal of Robotics and Mechatronics, Vol.4, No.2, pp.128-134 (1992).


| Plane size : 32×32 | Number of layers : 5 | | | | |
|---|---|---|---|---|---|
| | S | C | S | C | S |
| Size of receptive field | 5×5 | 3×3 | 7×7 | 3×3 | 11×11 |
| Number of planes | 18 | 18 | 10 | 10 | 5 |

Table 1. Structure of the expanded neocognitron for recognizing five patterns.

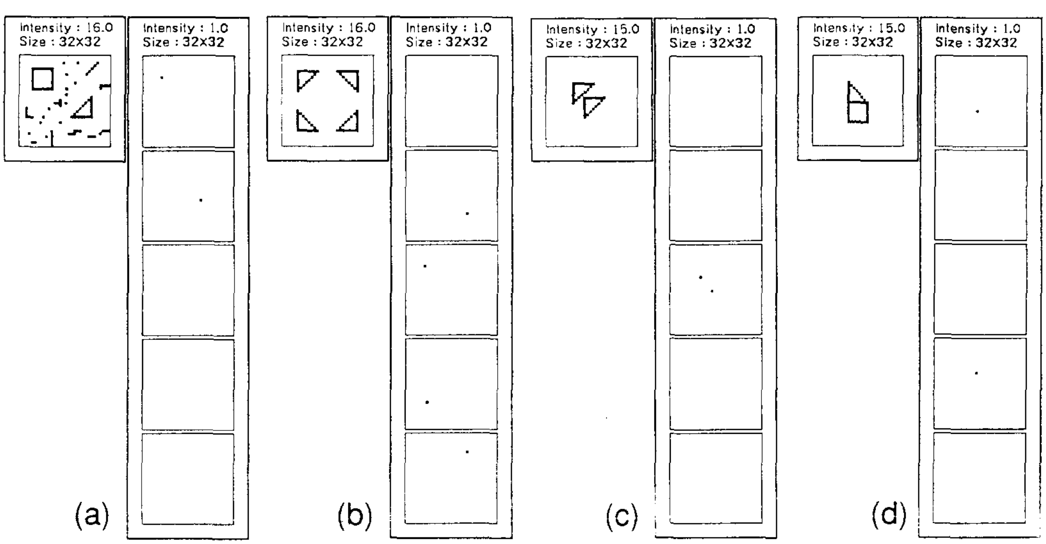


Figure 4. Four example images and the results of recognition.