HANDWRITTEN ALPHABET AND DIGIT CHARACTER RECOGNITION USING FEATURE EXTRACTING NEURAL NETWORK AND MODIFIED SELF-ORGANIZING MAP

Kenji NAKAYAMA Yasuhide CHIGAWA Osamu HASEGAWA
Dept. of Electrical and Computer Eng., Faculty of Tech., Kanazawa Univ.
2-40-20, Kodatsuno, Kanazawa 920 Japan

ABSTRACT

A new pattern recognition method is proposed for handwritten alphabet and digit characters. Features, which include an end of a line, a corner, a junction point, a cross point, a line segment between the feature points, and an isolated circle, are employed to express relative topological structure of handwritten characters These features are extracted by mutually connected neural networks combined in a layer form. The feature point distribution of a standard pattern is mapped onto that of a distorted pattern, through a modified self-organizing map (SOM). The distorted pattern is recognized based on similarity between both feature point distributions. The modified SOM has the following advantages. First, the number of feature points is small, which are classified into several groups. Second, the mapping is carried out in the variable ring shape region, in order to find out a suitable pairing of the feature points. Third, the feature points are selected from both the standard and the distorted patterns in order to avoid any vibration. Finally, neighborhood are selected along lines of the patterns. These improvements can provide stable and fast feature point mapping. Computer simulation demonstrates the proposed method can adapt a variety of pattern distortions.

I INTRODUCTION

Handwritten character recognition is very important application fields of neural networks. Because, the conventional computers are not so good for this purpose. Neural networks are expected to do well due to their analogy to human brain information processing. It is, however, not so easy to simulate the human visual information process, due to a lack of enough knowledge. However, many approaches based on mainly artificial neural networks have been proposed.

Neocognitron [1] is the first neural model, which can be applied to distorted pattern recognition. Recently, many approaches, based on multilayer neural networks, modified back-propagation and self-organizing methods have been proposed [2]-[6]. Some of them extract invariant features in the handwritten characters through different methods from neural systems. The features are applied to pattern matching using neural networks. In another methods, the input patterns are directly applied to the neural networks, in which feature extraction and pattern matching are simultaneously carried out.

In this paper, we propose a new pattern recognition scheme [7] for handwritten alphabet and digit characters. Binary patterns are taken into account in this paper. A system consists of a feature extracting neural network and a modified self-organizing map (SOM). In order to express relative topological structure of handwritten characters, useful features are employed. Several improvements are introduced in order to effectively apply Kohonen's SOM [8] to handwritten character recognition. Computer simulation is demonstrated using alphabet and digit characters.

II INVARIANT FEATURES IN HANDWRITTEN CHARACTERS

It is very important to extract features, which can express relative topological structure of handwritten characters. For this purpose, we employ the following features.

- End of a line
- Corner (Angle$\leq$90 degrees)
- Junction point (T type)
- Cross point (+ type)
- Arbitrary point on an isolated circle
- Middle point between another feature points

The middle point equivalently express a line segment between two feature points. Figure 1 shows examples of the feature points. A symbol ○ is an inactive unit, and symbols ●, ★ and ☆ are active units. Furthermore, ★ and ☆ indicate the feature point and the middle point, respectively.

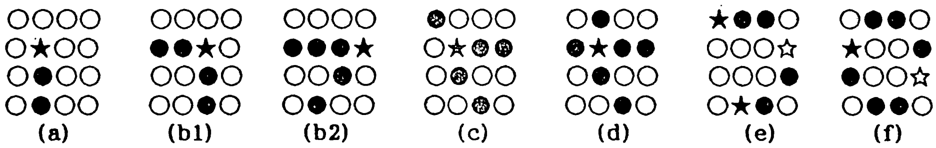

Fig.1 Examples of feature points. (a) End of a line. (b1), (b2) Corners. (c) Junction point. (d) Cross point. (e) Middle point. (f) Arbitrary point on an isolated circle, and middle point.

III FEATURE EXTRACTING NEURAL NETWORK

3.1 Blockdiagram

Figure 2 shows a blockdiagram of a feature extracting neural network. A character is applied to the I-layer. In the F-layer, the end of a line is extracted in the F_1 network. Candidates for the corner, the junction point and the cross point are detected in the F_2 network. The F_3 network extracts the corner of 90 degrees consisting of diagonal lines.

The TR-layer has one neural network, in which lines of the pattern are traced starting from the feature points. By counting the number of traces starting from the same point, its feature is determined. At the same time, the middle point is determined as a meeting point of two traces.

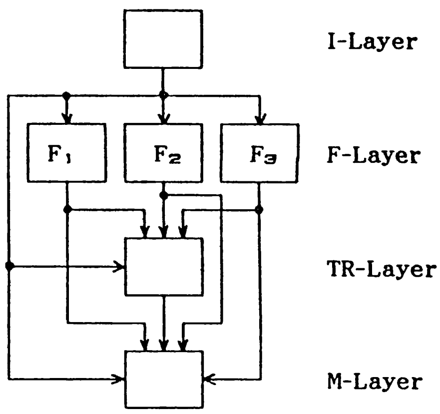

Fig.2 Blockdiagram for feature extracting neural network.

3.2 Features Extracted in F-Layer

In this paper, a pattern is assumed to be a fine pattern. In actuality, of course, some pre-processing is necessary to obtain a fine pattern.

The pattern is first applied to the I-layer, whose outputs are 1 or 0. Connection weights w_{IFij} from the ith unit in the I-layer to the jth unit in the F-layer are determined by

$$w_{IFij} = \begin{cases} 8, & i=j \\ 1, & \text{adjoining units} \\ 0, & \text{otherwise.} \end{cases} \qquad (1)$$

A unit has 8 adjoining units, as shown in Fig.3.

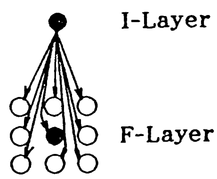

Fig.3 Connections from I-layer to F-layer. Unit ● has 8 adjoining units ○.

(a) F_1 Neural Network

The end of a line is detected in the F_1 network, by controlling the unit to be active, when its input is equal to 9.

(b) F_2 Neural network

Candidates for Feature Points:

By using the connection weights given by Eq.(1), the units, which locate on the right position or the adjoining positions of the feature points, have the maximum input. These units have at least three adjoining active units. Therefore, the unit is controlled to be active, when its input is equal to or greater than 11. Examples are shown in Fig.4.

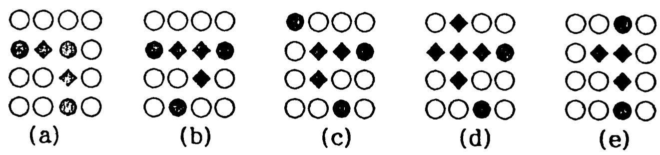

Fig.4 Examples of the units ◆, whose input is equal to or greater than 11.

Competitive Learning:

Since several units can be active, in the initial state, an appropriate unit is selected through the following competitive learning. Each unit is connected with the adjoining units, and has a self-loop. Assuming the ith unit and the jth unit locate in the adjoining, the input from the former to the latter is determined by

$$u_{F2ij}(n) = \begin{cases} \varepsilon, & v_{F2j}(n-1) > v_{F2i}(n-1), \quad i \neq j, \quad n \geq 1 \\ \delta, & v_{F2j}(n-1) = v_{F2i}(n-1) \\ -1, & v_{F2j}(n-1) < v_{F2i}(n-1) \end{cases} \qquad (2a)$$

$$u_{F2ij}(n) = v_{F2ij}(n-1), \quad i=j, \quad \text{self-loop} \tag{2b}$$

where, $0 < \varepsilon \ll \delta \ll 1$, and $v_{F2i}(n)$ and $v_{F2j}(n)$ are the outputs of both units. The initial output $v_{F2j}(0)$ is determined by

$$v_{F2j}(0) = \phi(\Sigma w_{IFij} v_{Ii} - 10) \tag{3}$$

$$\phi(x) = \begin{cases} x, & x>0 \\ 0, & x \leq 0 \end{cases} \tag{4}$$

where, v_{Ii} is the output of the I-layer unit. The jth unit output for $n \geq 1$ is determined by

$$v_{F2j}(n) = \phi(\Sigma u_{F2ij}(n) - \eta), \quad n \geq 1 \tag{5}$$

where η is a random number($0 < \eta < \varepsilon$), which makes small differences among the maximum inputs.

The unit, whose output is greater than that of the adjoining unit, can receive positive input. On the contrary, the unit, whose output is less than that of the adjoining unit, receives negative input. By repeating this state transition, the unit, which has the maximum input at n=0, or locate on the central point of the cluster of the active units, can be the winner.

(c) F_3 Neural Network

A corner of 90 degrees, consisting of diagonal lines as shown in Fig.5, cannot be detected in the F_2 network. Because it has only two adjoining active units. The F_3 network is used for detecting this kind of corners.

Figure 6 illustrates the units, which are connected with each other. The units ★ and ◎ locate on the same coordinates in each layer. The intermediate units $\diamond_1 \sim \diamond_4$ are prepared for each unit ◎. They receive the outputs from a pair of two units (①, ②), (①, ③), (②, ④), and (③, ④), respectively, through weights $w_{IAik}=1$. These units are controlled to be active, when their input is equal to 2. Their output is 1.

The unit ◎ receives the outputs from the units ★ and $\diamond_k$ through excitatory connections $w_{IF3jj}= 8$ and $w_{AF3kj}=1$, respectively, and from the units ○ in the I-layer through inhibitory connections $w_{IF3ij}=-8$. Thus, the input of the unit ◎ becomes 8+1=9, only when the unit ★ is the corner, as shown in Fig.5. Therefore, the unit ◎ is controlled to be active, only when its input is equal to 9.

○○★○○ ○●○○
○●○●○ ○○●○
●○○○● ○○○★
○○●○
○●○○

Fig.5 Examples of corners ★, which have only two adjoining activated units.

①○②
○★○ I-layer
③○④

$\diamond_1$ $\diamond_2$ $\diamond_3$ $\diamond_4$ Intermediate units

· · ◎ · · F_3 network

Fig.6 Units in I-layer and F-layer, and intermediate units, which are connected in layer form.

3.3 Trace along Pattern in TR-Layer

The F_2 network selects only candidates for feature points, as mentioned previously. Their feature are determined in the TR-layer by tracing along the pattern starting from the candidates.

A network in the TR-layer is a mutually connected neural network. A unit is connected to its neighborhood through connections w_{Tij}. Furthermore, it receives the outputs from the I-layer and the F-layer through connections w_{ITij} and w_{FTij}, respectively. They are determined by

$$w_{ITij} = \begin{cases} 8, & i=j \\ 0, & i \neq j \end{cases} \tag{6}$$

$$w_{FTij} = \begin{cases} 1, & i=j \\ 0, & i \neq j \end{cases} \tag{7}$$

$$w_{Tij} = \begin{cases} 1, & \text{adjoining units} \\ 0, & \text{otherwise} \end{cases} \tag{8}$$

· · ○○$◎_I$○○ · · I-Layer

· ○ $◎_F$ ○ · F-Layer

○ ○ ○
○ $◎_T$○ TR-Layer
○ ○ ○

Fig.7 Units in I-layer, F-layer and TR-layer, which are connected in layer form.

Figure 7 shows these relations. The units $◎_I$, $◎_F$ and $◎_T$ locate on the same coordinates in each layer. The units ○ in the TR-layer are the adjoining units of the unit $◎_T$. In the initial state, the unit, which locates on the pattern, receives 8 from the I-layer. Furthermore, if this unit corresponds to the feature points, then it receives 1 from the F-layer. Therefore, its input becomes 8+1=9. By setting threshold to be 9, only the feature units can be activated. They become starting units.

In the next step, the activated units output 1 to the adjoining units through the connections

w_{TIJ}=1. Therefore, input of the adjoining units, which locate on the pattern, becomes 9. Thus, unit activation is propagated along lines of the pattern. In order to trace the lines with two active units, each unit stays in an active state during two time slots.

3.4 Determining Features and Middle Points in M-Layer

The end of a line, the corner, the junction point and the cross point have one, two, three and four traces, respectively. The number of the traces is detected in the M-layer. This layer receives the outputs from the I-layer, the F-layer and the TR-layer, as shown in Fig.8. Units $\Diamond_I$, $\Diamond_F$, $\Diamond_T$ and $\Diamond_M$ locate on the same coordinates in each layer. Units $\bigcirc_{D1}$, $\bigcirc_{D2}$ detect the traces, which include a pair of two active units. These units are prepared for possible directions. They receive the outputs from the active units ① and ②, through connections w_{TD1IJ}=0.5 and w_{TD2IJ}=1.5, respectively. By setting threshold to be 2, they become active, only when they detect the trace.

The unit $\bigcirc_F$ is used for counting the number of traces. This unit receives the outputs from the corresponding unit $\Diamond_F$ in the F-layer and $\bigcirc_{DK}$ through connections w_{FFJJ}=10 and w_{DFKJ}=1, respectively. It is activated when its input is equal to or greater than 11. Its input represents the number of traces plus 10, and is directly transferred to the unit $\Diamond_M$ in the M-layer.

The unit $\bigcirc_P$ detects a middle point, at which two traces meet. This unit receives the outputs from the F-layer $\Diamond_F$ through large inhibitory connections w_{FPJJ}=-10, in order to avoid to select the feature point as a middle point. This unit also receives the outputs from the I-layer $\Diamond_I$, the TR-layer $\Diamond_T$ and $\bigcirc_{DK}$ through w_{IPJJ}=8, w_{TPJJ}=1 and w_{DPKJ}=1, respectively. Furthermore, it receives the output from the units ○ in the TR-layer through inhibitory connections. This unit is activated when its input is equal to 10, and outputs 10 to the unit $\Diamond_M$ in the M-layer.

· · ○○$\Diamond_I$○○ · · I-layer

· · ○○$\Diamond_F$○○ · · F-layer

○○②①$\Diamond_T$①②○○○ TR-layer

$\bigcirc_{D1}$ $\bigcirc_{D2}$ Trace units

Middle point $\bigcirc_P$ $\bigcirc_F$ Feature point

· · ○○$\Diamond_M$○○ · · M-layer

Fig.8 Units in I-layer, F-layer and TR-layer, and intermediate units, which are connected in layer form.

The unit $\Diamond_M$ has two inputs from the units $\bigcirc_F$ and $\bigcirc_P$. Thus, if its input is equal to 10, then the unit $\Diamond_M$ represents the middle point. On the other hand, the input is equal to or greater than 11, then the unit $\Diamond_M$ stands for the feature point. A value, given by (the input - 10), is the number of traces, which determines the feature.

3.5 Detection of Isolated Circle

An isolated circle, which does not have the previous features, is not detected in the F-layer. After the tracing in the TR-layer, one unit of the circle is selected at random. The circle is traced starting from this unit, and the middle point is determined.

IV FEATURE POINT DISTRIBUTION MAPPING

4.1 Seléction of Standard Patterns

Let P(m) and Q be the standard and the distorted patterns, respectively. Their feature points are expressed by (α, β, · ·) and (a,b , · ·), respectively. α and a stand for the end of a line, for instance. Candidates of the standard patterns are selected based on the following similarity.

$$S = \frac{\mathbf{p}\cdot\mathbf{q}}{|\mathbf{p}|\cdot|\mathbf{q}|} \tag{9}$$

$$\mathbf{p}=[N(\alpha),N(\beta),N(\gamma),\cdots] \tag{10a}$$

$$\mathbf{q}=[N(a),\ N(b),\ N(c),\ \cdots] \tag{10b}$$

N(α) and N(a) mean the number of feature points of α and a, respectively. The numerator is an inner product of **p** and **q**, and a symbol | | in the denominator stands for a norm of a vector.

4.2 Modified Self-Organizing Feature Mapping

Feature points of both P(m) and Q are distributed on the same plane. In order to simplify the following description, an NxN grid plane is used. Each feature point locates on the grid, which has the coordinates (i,j). The feature point distribution of P(m) is mapped onto that of Q. Because the standard patterns hold reasonable topological structure. This constraint is used in the mapping.

Step1: A feature point is selected in the region, whose coordinates are (1,j), (N,j), (i,1), (i,N), 1≤ i,j ≤ N. This region is called the 1st-region. An example is shown in Fig.9. The feature point is selected from both patterns, in order to avoid vibration in the mapping process.

Step2: When a feature point α_m, which is an element of α, is selected, the corresponding feature point, which locates close to α_m, is selected from a. Let this point be a_n. It can be found the outside of the 1st-region.

On the other hand, when a_n is first selected in the 1st-region, then the corresponding α_m will be chosen from α. The following processes are the same for both cases.

Step3: The feature points, which directly connected to α_m along the standard pattern, are selected as the neighborhood for α_m. We call these points 1st-order neighborhood. The next feature points are called 2nd-order neighborhood. Let rth-order neighborhood be α_{mrk}.

Step4: α_m is shifted toward a_n by a distance $d(\alpha_m)$. At the same time, the rth-order neighborhood are shifted toward the same direction by $d(\alpha_{mrk})$. Shifting distance is gradually decreased as r increases as follows:

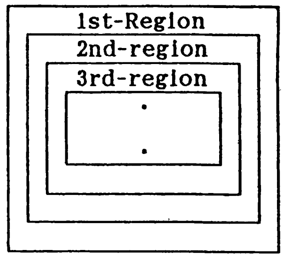

Fig.9 Variable ring shape region, in which the feature point mapping is carried out.

$$d(\alpha_m) > d(\alpha_{m1k}) > d(\alpha_{m2k}) > \cdot\ \cdot \tag{11}$$

By using the neighborhood, obtained in Step2, and the above weighted shifting distance, relative topological structure of P(m) can be held through the mapping.

Step5: Select another feature point in the same region, and repeat Step2 through Step4. If all feature points are selected in the 1st-region, then the mapping moves on the next region, that is the 2nd-region: (2,j), (N-1,j), (i,2), (i,N-1), $2 \leq i,j \leq N-1$. In this region, Step 2 through Step4 are repeated. Thus, the mapping region is gradually reduced toward the central point, as shown in Fig.9. This variable region can make it possible to find out a suitable pairing of the feature points.

After the region reaches the central point, the mapping process returns to Step1. The above steps are repeated until the mapping process converges.

4.3 Pattern Recognition

After the feature point mapping, the right character is selected using the following similarity between the feature point distributions of the mapped version of P(m) and Q.

$$C = \frac{[\Sigma(\alpha_m, a_n)/r_{mn}]_N + [\beta]_N + [\gamma]_N \cdot\ \cdot}{[\Sigma(\alpha_m, a_n) + \Sigma\alpha_m + \Sigma a_n]_D + [\beta]_D + [\gamma]_D \cdot\ \cdot} \tag{12}$$

In the above equation, each term is determined as follows: If α_m is exactly mapped onto a_n, then $(\alpha_m, a_n)=1$, $\alpha_m=0$, $a_n=0$ and $r_{mn}=1$. If α_m locates close to a_n, then $(\alpha_m, a_n)=1$, $\alpha_m=0$, $a_n=0$ and $r_{mn}>1$. r_{mn} is proportional to the distance between α_m and a_n. If there is no corresponding feature point close to α_m and a_n, then $(\alpha_m, a_n)=0$, $\alpha_m=1$ and $a_n=1$. $[\beta]_{N,D}$ and $[\gamma]_{N,D}$ indicate the same operation as in the previous $[\ ]_{N,D}$ for α and a. When the feature point distribution of P(m) is completely mapped onto that of Q, the similarity C becomes 1. Otherwise, it is less than 1.

V SIMULATION

5.1 Feature Extraction

Alphabet and digit characters in a printed form with 16x16 dots are used as the standard patterns. Figure 10 shows examples of feature points extracted by the proposed neural network. A small point · indicates a dot of the pattern. The middle points are not detected in a short line segments. This can be improved by increasing the number of dots. Although some feature points are somewhat shifted from their right positions, they can be compensated for in the modified SOM, as shown in the following.

5.2 Feature Point Mapping

An example of the distorted pattern and its feature points are shown in Fig.11. The candidates for this pattern, selected based on Eq.(9), include the patterns 「4」, 「5」, 「N」, 「Z」 and 「z」. In order to shift the feature points continuously, a 32x32 grid plane is employed in this process.

Examples of the feature point mapping are shown in Fig.12 using 「N」, 「Z」 and 「5」 as the standard patterns. A solid line and a dotted line connect the feature points in the distorted and the mapped standard patterns, respectively. These lines do not exactly follow the patterns. The pattern 「N」 is exactly mapped onto its distorted version. The similarity C becomes 1. On the other hand, the other patterns cannot approach to the distorted 「N」, due to their relative topological structure.

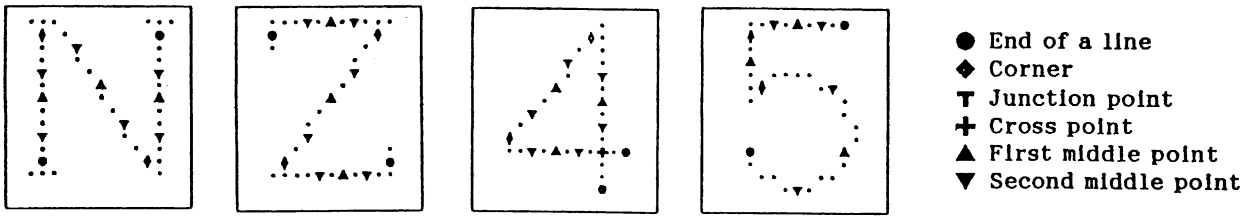

Fig.10 Examples of extracted feature points.

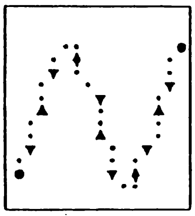

Fig.11 Distorted pattern of 「N」.

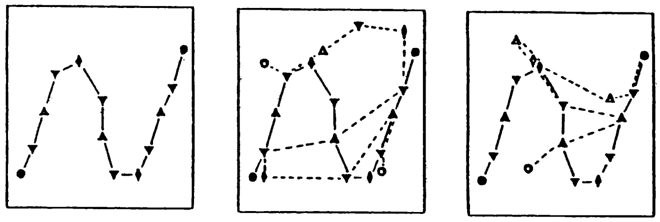

Fig.12 Examples of the feature point mapping.

5.3 Distorted Pattern Recognition

We have tried so many distorted patterns. Figure 13 shows some examples of patterns which can be recognized. From these results, the proposed method can adapt rotation, shifting, expansion and contraction and distortion, which can occur in the handwritten characters.

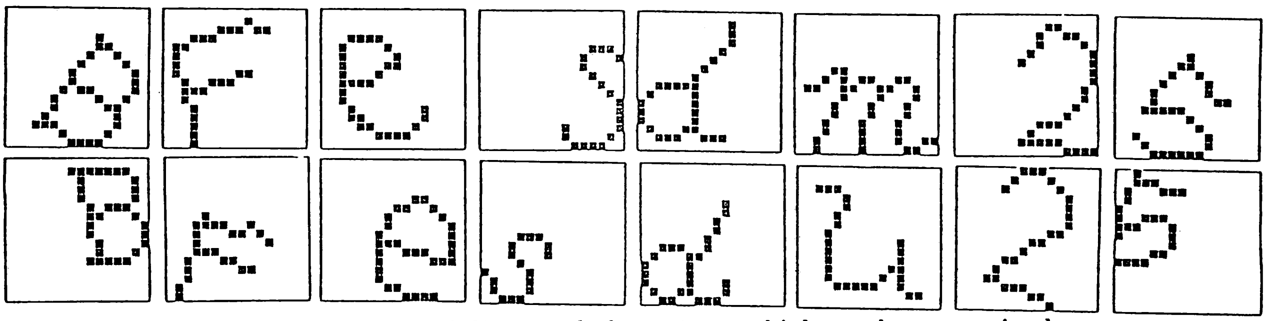

Fig.13 Examples of distorted characters, which can be recognized.

VI CONCLUSIONS

A new patten recognition scheme, based on neural networks, has been proposed for handwritten character recognition. Computer simulations have demonstrated the proposed method can adapt a variety of distortions. By adding typical handwritten characters in the standard patterns, more complicated and distorted patterns can be recognized.

REFERENCES

[1]K.Fukushima, S.Miyake and T.Ito, "Neocognitron: A neural network model for a mechanism of visual pattern recognition", IEEE Trans. vol.SMC-13, pp.826-834, 1983.

[2]Y.Chen et al., "Backpropagation applied to handwritten zip code recognition", Neural Network, vol.1, no.4, pp.541-551, 1989.

[3]Y.Hirai and Y.Tsukui, "Position independent neuro pattern matching and its application to handwritten numeral character recognition", Proc. IJCNN'91 San Diego, pp.Ⅲ-695-701, 1991.

[4]T.P.Vogl et al., "Classification of hand-written digits and Japanese Kanji", Proc. IJCNN'91 Seattle, pp. Ⅰ-97-102, 1991.

[5]S.Lee and J.Pan, "Handwritten numeral recognition based on hierarchically self-organizing learning networks", Proc. IJCNN'91 Singapore, pp.1313-1322, 1991.

[6]Y.Idan and R.C.Chevallien, "Handwritten digits recognition by a supervised Kohonen-like learning algorithm", Proc. IJCNN'91 Singapore, pp.2576-2581, 1991.

[7]Y.Chigawa, O.Hasagawa and K.Nakayama, "Handwritten digit recognition by using feature extracting neural network and self-organizing feature map", IEICE Japan, Report of Technical Meeting on Neuro Computing, vol.NC91-50, pp.15-22, Oct. 1991.

[8]T.Kohonen, "The self-organizing map", Proc. IEEE, vol.78, no.9, pp.1464-1480, Sept. 1990.