Recognition and Segmentation of Connected Characters With Selective Attention

KUNIHIKO FUKUSHIMA AND TARO IMAGAWA*
Osaka University
(Received 16 January 1992; revised and accepted 8 May 1992)

Abstract—We have modified the original model of selective attention, which was previously proposed by Fukushima, and extended its ability to recognize and segment connected characters in cursive handwriting. Although the original model of selective attention already had the ability to recognize and segment patterns, it did not always work well when too many patterns were presented simultaneously. In order to restrict the number of patterns to be processed simultaneously, a search controller has been added to the original model. The new model mainly processes the patterns contained in a small “search area,” which is moved by the search controller. A preliminary experiment with computer simulation has shown that this approach is promising. The recognition and segmentation of characters can be successful even though each character in a handwritten word changes its shape by the effect of the characters before and behind.

Keywords—Neural network, Selective attention, Visual pattern recognition, Character recognition, Segmentation, Recognition of connected characters, Cursive handwriting.

1. INTRODUCTION

Machine recognition of connected characters in cursive handwriting of English words is a difficult problem. It cannot be successfully performed by a simple pattern matching method because each character changes its shape by the effect of the characters before and behind. In other words, the same character can be written differently when it appears in different words in order to be connected smoothly with the characters in front and in the rear.

Fukushima (1986, 1987, 1988a) previously proposed a “selective attention model,” which has the ability to segment patterns, as well as the function of recognizing them. When a composite stimulus consisting of two patterns or more is presented, the model focuses its attention selectively to one of them, segments it from the rest, and recognizes it. After that, the model switches its attention to recognize another pattern. The model also has the function of associative memory and can restore imperfect patterns. These functions can be successfully performed even for deformed versions of training patterns, which have not been presented during the learning process.

However, the model does not always work well when too many patterns are presented simultaneously. The model has been modified and extended to be able to recognize connected characters in cursive handwriting (Imagawa & Fukushima, 1990, 1991). A search controller has been added to the original model in order to restrict the number of patterns to be processed simultaneously. The new model processes the patterns contained in a small “search area,” which is moved by the search controller. The positional control of the search area does not need to be accurate, as the original model, by itself, has the ability to segment and recognize patterns, provided the number of patterns present is small.

In the recognition of cursive handwriting, the information of the height or vertical position of characters sometimes becomes important. For instance, character “?” in script style can be interpreted as a deformed version of character “z.” They differ only in their heights. Since our selective attention model has the ability of deformation-resistant pattern recognition,
both of them might be recognized as the same character. In order to discriminate between them, we have introduced a mechanism to measure the height of the character which is recognized and segmented.

2. STRUCTURE AND BEHAVIOR OF THE MODEL

The model is a hierarchical multilayered network and consists of a cascade of many layers of neuron-like cells. The cells are of the analog type: their inputs and outputs take non-negative analog values. Figure 1 illustrates the multilayered structure of the hierarchical network. Each rectangle in the figure represents a group of cells arranged in a two-dimensional array.

The network has backward as well as forward connections between cells. Figure 2 shows how the different kinds of cells, such as $u_s$, $u_c$, $w_s$, and $w_c$, are interconnected in the network. Each circle in the figure represents a cell. Letters $u$ and $w$ indicate the cells in the forward paths and backward paths, respectively. Although the figure shows only one of each kind of cell in each stage, numerous cells actually exist arranged in a two-dimensional array, as shown in Figure 1. We will use notation $u_{c1}$, for example, to denote a $u_c$-cell in the $l$-th stage, and $U_{c1}$ to denote the layer of $u_{c1}$-cells. The highest stage of the network is the $L$-th stage ($L = 4$).

A more detailed diagram illustrating spatial interconnections between neighboring cells appears in Figure 3.

2.1. Forward Paths

The signals through forward paths manage the function of pattern recognition. If we consider the forward paths only, the model has almost the same structure and function as the "neocognitron" model (Fukushima, 1980, 1988b), which can recognize input patterns robustly, with little effect from deformation, changes in size, or shifts in position.

Cells $u_s$ are feature-extracting cells. They correspond to S-cells in the neocognitron. With the aid of subsidiary inhibitory cell $u_{s1}$, they extract features from the stimulus pattern. The $u_s$-cells of the first stage have fixed input connections and extract line components of various orientations. In all other stages higher than the first, $u_s$-cells have variable input connections, which are reinforced by unsupervised learning.

The $u_c$-cells, which correspond to C-cells of the neocognitron, are inserted in the network to allow for positional errors in the features of the stimulus. Each $u_c$-cell has fixed excitatory connections from a group of $u_s$-cells which extract the same feature, but from slightly different positions. Thus, the $u_c$-cell's response is less sensitive to shifts in position of the stimulus patterns.

![FIGURE 1. Multilayered structure of the hierarchical network.](image-url)
Recognition of Connected Characters

The processes of feature-extraction by $u_{5}$-cells and toleration of positional shift by $u_{c}$-cells are repeated in the hierarchical network. During this process, local features extracted in a lower stage are gradually integrated into more global features. This structure is effective for endowing the network with robustness against deformation in pattern recognition.

The layer of $u_{c}$-cells at the highest stage, that is, layer $U_{CL}$, works as the recognition layer. The response of the cells of this layer shows the final result of pattern recognition. Even when two patterns or more are simultaneously presented to the input layer $U_{co}$, usually only one cell, corresponding to the category of one of the stimulus patterns, is activated in the recognition layer $U_{CL}$. This is partly because of the competition between $u_{5}$-cells by lateral inhibition, and also because of the attention focusing by gain control signals from the backward paths, which will be discussed below.

Mathematically, the output of the cells in the forward paths are calculated as follows in the computer simulation. In the mathematical descriptions below, the output of a $u_{c}$-cell, for example, is denoted by $u_{C}(n, k)$, where $n$ is a two-dimensional set of coordinates indicating the position of the cell’s receptive-field center in the input layer $U_{co}$ and $k (=1, 2, \ldots, K_{t})$ is a serial number indicating the type of feature which the cell responds. In other words, $k$ is a serial number of the cell-plane defined in connection with the neocognitron. Variable $t$ represents the time elapsed after the presentation of stimulus pattern and takes a discrete integer value. Sometimes in such expressions, $k$ is abbreviated for stage $U_{co}$ in which we have $K_{o} = 1$, and $n$ is omitted for the highest stage which has only one $u_{c}$-cell for each value of $k$.

Among $u_{5}$-cells, there is a mechanism of backward lateral inhibition. Since the calculation of backward lateral inhibition is time-consuming in computer simulation, the computation of the output of a $u_{5}$-cell is divided into two steps. More specifically, before calculating the final output of a feature-extracting cell $u_{S}$, a temporary output $u_{S}$, in which the effect of lateral inhibition is ignored, is calculated first:

$$u_{S}(n, k) = \frac{a_{t} + \sum_{v=1}^{K_{t}} \sum_{n_{v}=0}^{A_{t}} a_{t}(n_{v}, n, k) \cdot u_{C}(n + v, k)}{1 + r_{t}(n, k) \cdot b_{t}(k) \cdot u_{S}^{2}(n)} - 1,$$

where $\varphi[x] = \max(x, 0)$. The output of subsidiary cell $u_{S}$, which sends inhibitory signal to this $u_{5}$-cell, is given by

$$u_{S}^{2}(n) = \sqrt{\sum_{v=1}^{K_{t}} \sum_{n_{v}=0}^{A_{t}} c_{t}(n_{v}) \cdot \{u_{C}(n + v, k)\}^{2}}.$$

Incidentally, this is equal to the root-mean-square of the responses of the $u_{c}$-cells. Parameter $a_{t}$ is a positive constant determining the level at which saturation starts in the input-to-output characteristic of the $u_{5}$-cell. $a_{t}(n_{v}, n, k)$ is the strength of the excitatory input connection coming from cell $u_{C}(n + v, k)$ in the preceding stage $U_{i-1}$, and $A_{t}$ denotes the summation range of $v$, that
is, the size of the spatial spread of the input connections to one $u_{S}'$-cell. $b_l(k)$ ($\geq 0$) is the strength of the inhibitory input connection coming from subsidiary cell $u_{S}'(n)$. $c_l(v)$ represents the strength of the fixed excitatory connections, and is a monotonically decreasing function of $|v|$. The positive variable $r_l(n, k)$, which will be given by eqn (9), determines the efficiency of the inhibitory input to the $u_{S}'$-cell.

From the above temporary output $\hat{u}_{S}'(n, k)$, in which the effect of lateral inhibition is ignored, the final output of the $u_{S}'$-cell is calculated. The calculation is made approximately, however, for the sake of economy of the computation time: The final output of the $u_{S}'$-cell is calculated by applying the following recursive equation twice, beginning with $u_{S}'(n, k) = \hat{u}_{S}'(n, k)$:

$$u_{S}'(n, k) := \varphi \left[ u_{S}'(n, k) - \sum_{v \in E_{l+1}} c_l(v) \cdot u_{S}'(n + v, k) \right] - \sum_{v \in E_l} \hat{c}_l(v) \cdot u_{S}'(n + v, k).$$

where $c_l(v)$ and $\hat{c}_l(v)$ are the strength of the connections for lateral inhibition, and $E_l$ denotes the size of the spatial spread of these connections. The notation := is used in the sense of recursive call in computer languages (for example, ALGOL). This means that lateral inhibition works quickly compared with other time delays in the network.

The input connections $a_l(v, \kappa, k)$ and $b_l(k)$ are fixed for the first stage ($l = 1$). They are adjusted in such a way that the $u_{S}'$ cell can extract line components of a particular orientation. In the computer simulation discussed later, each $u_{S}'$ cell has $3 \times 3$ excitatory input connections, which have spatial distribution as illustrated in Figure 4.

In all other stages higher than the first, the input connections of $u_{S}'$-cells are variable and reinforced by means of an algorithm similar to that used for the unsupervised learning in the neocognitron (Fukushima, 1980, 1988b) when all backward signal flow is stopped. Thus, each $u_{S}'$-cell comes to respond selectively to a particular feature of the stimuli presented during the learning phase.

The output of a $u_{S}'$-cell is given by

$$u_{S}'(n, k) = g_l(n, k) \cdot \varnothing \left[ \sum_{v \in E_l} d_l(v) \cdot u_{S}'(n + v, k) \right].$$

where $\varnothing[x] = \varphi[x] / (1 + \varphi[x])$. Parameter $d_l(v)$ denotes the strength of the fixed excitatory connections and is a monotonically decreasing function of $|v|$. The size of the spatial spread of these connections is $D_l$. The variable $g_l(n, k)$ denotes the gain of the $u_{S}'$-cell, and its value is controlled by the signal from the $w_{S}'$-cell in the backward path and also from the search controller as discussed in Sections 2.4 and 2.5.

The input layer $u_{S}'$ receives not only the input pattern $p$ but also positive feedback signals from the recall layer $w_{S}'$, as in Figure 2. Hence $u_{S}'$-cells of the input layer are different in nature from those of other stages. Expressed mathematically,

$$u_{S}'(n) = g_l(n) \cdot \max[p(n), w_{S}'(n)].$$

The gain $g_{S}'(n)$ is given by eqn (13) in the same manner as for the intermediate stages. The output of a $w_{S}'$-cell will be given by eqn (6), and its value at $t < 0$ is zero.

2.2. Backward Paths

The signals through backward paths manage the function of selective attention and associative recall. The cells in the backward paths are arranged in the network in a mirror image of the cells in the forward paths. The forward and the backward connections also make a mirror image to each other but the directions of signal flow through the connections are opposite.

The output signal of the recognition layer $U_{CL}$ is sent to lower stages through the backward paths and reaches the recall layer $W_{C0}$ at the lowest stage of the backward paths. The backward signals are transmitted retracing the same route as the forward signals. The route control of the backward signals is made by the gate signals from the cells of the forward paths. More specifically, from among many possible backward paths diverging from a $w_{S}'$-cell, only the ones to the $w_{S}'$-cells which are receiving gate signals from the corresponding $u_{S}'$-cells are chosen (Fukushima, 1986, 1987, 1988a) (Figure 3). Guided by the gate signals from the forward paths, the backward signals reach exactly the same positions at which the input pattern is presented.

As mentioned before, usually only one cell is activated in the recognition layer $U_{CL}$, even when two or more patterns are presented to the input layer $U_{C0}$. Since the backward signals are sent only from the activated recognition cell, only the signal components corresponding to the recognized pattern reach the recall layer, $W_{C0}$. Therefore, the output of the recall layer can also be interpreted as the result of segmentation, where only components relevant to a single pattern are selected from the stimulus. Even if the stimulus pattern which is now recognized is a deformed version of a training pattern, the deformed pattern is segmented and emerges with its deformed shape.

The following is a more detailed description of the response of the cells. Mathematically, the output of a $w_{S}'$-cell and the subsidiary cell $w_{S}'$ in the backward paths is given by
backward paths.

the same equation (8) can be applied if we put

where \( \alpha_i \) is a positive constant.

Uc-cells are sent directly back to wc-cells through

\( \text{w}_{\text{L}}(n, k) = u_{\text{L}}(n, k) \). In other words, the output of

Quantitatively, the output of a Ws-cell is given by

when it receives a signal from both
corresponding

signals from we-cells but also a gate signal from the

that each Ws-cell receives not only excitatory backward

activated. To elicit a similar response from the Ws-cells

at least one preceding Us-cell is enough, and usually
activate a Uc-cell in the forward path, the activation of

cells which receive excitatory backward signals from

in the backward paths, the network is synthesized so

connections diverge from a we-cell towards Ws-cells

layer \( 1 whirlto \) layer \( 2 whirlto \), however, some control signals
control signals from the forward paths are required to

guide the backward signal flow between these layers.

To control the route of the backward signal flow from
layer \( \text{W}_{\text{L}} \) to layer \( \text{W}_{\text{S}} \), however, some control signals
from the forward paths are necessary. Corresponding
to the fixed forward connections which converge to a
us-cell from a number of us-cells, many backward
connections diverge from a wc-cell towards ws-cells
(Figure 3). It is not desirable, however, for all the ws-cells
which receive excitatory backward signals from the
wc-cell to be activated. The reason is as follows: To
activate a uc-cell in the forward path, the activation of
at least one preceding us-cell is enough, and usually
only a small number of preceding us-cells are actually
activated. To elicit a similar response from the ws-cells
in the backward paths, the network is synthesized so
that each ws-cell receives not only excitatory backward
signals from wc-cells but also a gate signal from the
corresponding us-cell, and the ws-cell is activated only
when it receives a signal from both us- and wc-cells.
Quantitatively, the output of a ws-cell is given by

\[
\text{w}_{\text{S}}(n, k) = \min \left[ \text{w}_{\text{S}}(n, k), \sum_{r \in \text{Ox}} \text{d}(r) \cdot \text{w}_{\text{S}}(n, k) \right], \quad (8)
\]

where \( \alpha_i \) is a positive constant.

In the highest stage, where no wc-cell exists,
the same equation (8) can be applied if we put
\( \text{w}_{\text{L}}(n, k) = \text{w}_{\text{L}}(n, k) \). In other words, the output of
us-cells are sent directly back to wc-cells through
backward paths.

2.3. Threshold Control

Take, for example, a case in which the stimulus contains
a number of incomplete patterns which are contami-
nated with noise and have several parts missing. Even
when the pattern recognition in the forward path is
successful and only one cell is activated in the recogni-
tion layer \( \text{U}_{\text{L}} \), it does not necessarily mean that the
segmentation of the pattern is also completed in the recall layer \( \text{W}_{\text{C}} \).

When some part of the input pattern is missing and
the feature which is supposed to exist there fails to be
extracted in the forward paths, the backward signal flow
is interrupted at that point and cannot proceed any
further because no gate signals are received from the
forward cells. In such a case, the threshold for extracting
features is automatically lowered around that area and
the model tries to extract even vague traces of the un-
detected feature. More specifically, the fact that a feature
has failed to be extracted is detected by \( \text{w}_{\text{S}} \)-cells from
the condition that a wc-cell in the backward paths is
active but that feature-extracting us-cells around it are
all silent (Figures 2 and 3). The signal from \( \text{w}_{\text{S}} \)-cells
weakens the efficiency of inhibition by \( \text{u}_{\text{S}} \)-cells, and
virtually lowers the threshold for feature extraction by
the us-cells. Thus, us-cells are made to respond even
in incomplete features, to which, in the normal state,
no \( \text{u}_\text{s} \)-cell would respond.

Thus, once a feature is extracted in the forward
paths, the backward signal can then be further trans-
mitted to lower stages through the path unlocked by
the gate signal from the newly activated forward cell.
Hence, a complete pattern, in which defective parts are
interpolated, emerges in the recall layer \( \text{W}_{\text{C}} \). Even if
the stimulus pattern which is now recognized is a de-
formed version of a training pattern, interpolation is
performed, not for the training pattern, but for the de-
formed stimulus pattern. From this restored pattern,
noise and blemishes have been eliminated because no
backward signals are returned for components of noise
or blemishes in the stimulus. Thus, the segmentation
of patterns can be successful, even if the input patterns
are incomplete and contaminated with noise. Com-
ponents of other patterns which are not recognized at
this time are also treated as noise.

A threshold-control signal is also sent from the no-
response detector shown at far right in Figure 2. When
all of the recognition cells are silent, the no-response
detector sends the threshold-control signal to the us-
cells in all stages through path \( \text{x} \) shown in Figure 2,
and lowers their threshold for feature extraction until
at least one recognition cell becomes activated.

Mathematically, the efficiency of inhibition to a us-cell
is determined by \( r/[n, k) \) in eqn (1), and its value
is controlled by two kinds of threshold-control signals,
\( x_S \) and \( x_x \), as follows:

\[
r/[n, k) = \frac{r^p}{1 + x_S(n, k) + x_x(n, k)}, \quad (9)
\]
The size of the area $Xs$ increases by an amount proportional to the output $38 K$. Fukushima and T. Imagawa

This difference in size is effective in preventing a spurious input signals in the forward paths (c.f. eqn (4)).

A $w_{cx}$-cell, which monitors the failure of extracting a feature in the forward paths, receives an excitatory connection from a $w_{ct}$-cell and inhibitory connections $d_j(v)$ from $u_{sr}$-cells around it in the forward paths. Its output is given by

$$w_{cx}(n, k) = \phi \left[ w_{ct}(n, k) - \sum_{v \in D_j} d_j(v) \cdot u_{sr}(n + v, k) \right]. \quad (11)$$

The size of the area $D_j$, from which the inhibitory signals from the preceding $u_{sr}$-cells are gathered, is a little wider than the spread of $D_j$, from which a $u_{ct}$-cell receives input signals in the forward paths (c.f. eqn (4)). This difference in size is effective in preventing a spurious output from $w_{cx}$-cells even when a stimulus pattern is a slightly deformed version of a learned pattern.

The other threshold-control signal $x_{ct}$ is generated by the no-response detector. The no-response detector monitors the response of the $u_{ct}$-cells and increases the level of $x_{ct}$ if all the $u_{ct}$-cells are silent. The level of $x_{ct}$ supplied to the $l$-th stage is

$$x_{ctl} = \begin{cases} x_{ctl}' + \beta_{ct} & \text{if } w_{ctl}(k) = 0 \text{ for all } k \\ \beta_{ct} \cdot x_{ctl}' & \text{else} \end{cases} \quad (12)$$

In other words, $x_{ct}$ is increased by a constant amount $\beta_{ct}$ if all the $u_{ct}$-cells in the recognition layer are silent. The increase of the level of $x_{ct}$ is continued until at least one $u_{sr}$-cell, and consequently one $u_{ct}$-cell, is activated. Once at least one $u_{ct}$-cell is activated, the increase in $x_{ct}$ stops and begins to decay with an attenuation ratio $\beta_{ct}'$.

### 2.4. Gain Control

The gains of $u_{ct}$-cells in the forward paths are variable and controlled by two kinds of gain-control signals: one from the corresponding backward cells $w_{ct}$, and the other from the search controller (Figure 2). Mathematically, gain $g_c(n, k)$ in eqns (4) and (5) is given by

$$g_c(n, k) = g_{wt}(n, k) \cdot g_{st}(n). \quad (13)$$

where $g_{wt}(n, k) (>0)$ is controlled by the signal from the backward cell $w_{ct}(n, k)$, and $g_{st}(n) (>0)$ is controlled by the signal from the search controller. This section discusses the former and the latter will be discussed in Section 2.5.

When a $w_{ct}$-cell is activated, it sends a gain control signal to the corresponding $u_{ct}$-cell and increases the gain between the inputs and the output of the $u_{ct}$-cell. Thus, only the forward signal flow in the paths in which backward signals are flowing is facilitated. (This method of gain control is somewhat different from that in the original model (Fukushima, 1987, 1988a), in which the gain of a $u_{ct}$-cell is decreased when the corresponding $w_{ct}$-cell is not activated.)

Since the backward signals are usually sent from only one activated recognition cell, only the forward paths relevant to the pattern which is now recognized are facilitated. This means that attention is selectively focused on only one of the patterns in the stimulus. A $u_{ct}$-cell is fatigued if it receives a strong gain control signal. It can maintain high gain only when it is receiving a large gain-control signal. Once the gain control signal disappears, the gain of the $u_{ct}$-cell drops rather rapidly and cannot recover for a long time. This fatigue is effectively used for switching attention to another pattern. It prevents the model from recognizing the same character twice.

Mathematically, $g_{wt}(n)$ in eqn (13) consists of two components: $g_{wt1}(n, k) (>0)$ and $g_{wt2}(n, k) (>0)$. They represent the effect of facilitation and fatigue, respectively.

$$g_{wt}(n, k) = 1 + \alpha_{wt1} \cdot g_{wt1}(n, k) - g_{wt2}(n, k). \quad (14)$$

where $\alpha_{wt1} (>1)$ is a constant determining the degree of facilitation.

The values of $g_{wt1}$ and $g_{wt2}$ vary as follows:

If $w_{ct}(n, k) > 0$:

$$g_{wt1}(n, k) = \gamma_{t1} \cdot g_{wt1}(n, k) + (1 - \gamma_{t1}) \cdot w_{ct}(n, k). \quad (15)$$

$$g_{wt2}(n, k) = \gamma_{t2} \cdot g_{wt2}(n, k) + (1 - \gamma_{t2}) \cdot w_{ct}(n, k). \quad (16)$$

If $w_{ct}(n, k) = 0$:

$$g_{wt1}(n, k) = \gamma_{t1} \cdot g_{wt1}(n, k), \quad (17)$$

$$g_{wt2}(n, k) = \gamma_{t2} \cdot g_{wt2}(n, k). \quad (18)$$

where $\gamma_{t1}$, $\gamma_{t2}$ and $\gamma_{t1}$ are positive constants (<1) determining the speed of build-up and decay of the gain. The values of $\gamma_t$ and $\gamma_{t2}$ are determined to be small. The value of $\gamma_{t2}$ is much larger: It is nearly equal to 1 for small $l$, but somewhat smaller for larger $l$. The initial values of $g_{wt1}$ and $g_{wt2}$ are zero, that is, $g_{wt1}(n, k) = 0$ and $g_{wt2}(n, k) = 0$.

Therefore, the values of $g_{wt1}$ and $g_{wt2}$ increase very rapidly with the same time constant, when the corresponding backward cell $w_{ct}$ is active. Since we have $\alpha_{wt1} > 1$ in eqn (14), the value of $g_{wt}$, and consequently the gain of the $u_{ct}$-cell, are increased very fast.

When the $w_{ct}$-cell becomes silent, however, the values
of $g_{B1}$ and $g_{B2}$ decrease with different time constants. The value of $g_{B1}$, which controls the degree of facilitation, decreases very rapidly, while the value of $g_{B2}$, which causes the effect of fatigue, does not decay for a long time, unless the corresponding $w_c$-cell is activated again. Hence, the effect of fatigue remains in the $U_c$-cell and does not recover for a long time. Quantitatively, this tendency is stronger for a lower stage, but somewhat weaker for a higher stage. This difference in time constant is effective in making the network easily process an input string which contains two or more characters of the same category. It is a matter of course that cells $u_c$ are not fatigued at all, if they have not been facilitated before.

2.5. Search Area

Although the original model of selective attention already has the ability to recognize and segment patterns, it does not always work well when too many patterns are presented simultaneously. In order to restrict the number of patterns to be processed simultaneously, a search controller is introduced into the new model. The search controller sends a gain-control signal to $U_c$ in all stages except $U_{cL}$. The signal decreases the gain of the $u_c$-cells situated outside of the search area. The effect of this signal is represented by $g'_{S}(n)$ in eqn (13).

It should be noted that the process of controlling the search area in our model is not identical to a simple process of limiting visual field or gating input signals by a "searchlight" mechanism (Crick, 1984). The search controller controls the gain of the $u_c$-cells, not only in the input layer $U_{c0}$, but also in all other $U_c$-layers except $U_{cL}$. The size of the search area is controlled to be larger at a higher stage than in a lower stage.

The position of the search area is shifted to the place in which a larger number of line-extracting cells are activated. To be more specific, the output of layer $U_{c1}$ is filtered by a spatial filter with Gaussian distribution, and the place of maximal activity is detected. The center of the search area is moved to this place.

The boundary of the search area is not sharply restricted: The gain of the $u_c$-cells are controlled to decrease gradually around the boundary. Since the present model has been designed to recognize a character string written in a single line only, the spatial distribution of the gain is controlled to be Gaussian in the horizontal direction, but is uniform in the vertical direction.

Mathematically, $g'_{S}(n)$ in eqn (13) is given by

$$g'_{S}(n) = \exp \left( -\frac{(n_x - \mu)^2}{2 \sigma^2} \right).$$  \hspace{1cm} (19)

where $n_x$ represents the $x$ coordinate of $n$, that is, $n = (n_x, n_y)$. The $x$ coordinate of the center of the search area is represented by $\mu$. The positive parameter $\sigma$ is set to be larger for larger $l$, and we have $g'_{S} = 1$.

2.6. Switching Attention

Once a character has been recognized and segmented, the attention is switched to recognize another pattern. To be more exact, there is a detector in the network which determines the timing of attention switching. The detector monitors the following two conditions: whether the number of activated recognition cells $u_{cL}$ is only one, and whether the total activity of layer $U_{cL-1}$ has nearly reached a steady state. When both of these conditions are simultaneously satisfied, the detector sends a command to switch attention.

The fatigue of the cells is effectively used in the model for switching attention to another pattern. Once a command to switch attention is given to the network, the backward signal flow is cut off for a short period. Since the gain control signals from $w_c$-cells disappear, the gain of $u_c$-cells falls to the level determined by the degree of fatigue of the cells. The stronger the fatigue has been before switching attention, the smaller the gain of the cell becomes after switching attention. The effect of the threshold control signal is also reset at this moment.

Because of this method of controlling the gain of the $u_c$-cells, signals corresponding to the previous pattern have difficulty flowing through the forward paths. Usually another recognition cell $u_{cL}$, hitherto silent, will be activated. If no $u_{cL}$-cell is activated, the non-response detector works until at least one $u_{cL}$-cell is activated.

In order to find a new position to which the search area is to be moved, the output of layer $U_{c1}$ is filtered by a spatial filter with Gaussian distribution again, and the place of maximal activity is sought. The gain control signal from the search controller is extinguished during this process. Once the place of maximum activity is detected, the search area is moved to the place, and the process of recognition and segmentation is restarted.

If the level of the maximum activity is less than a certain threshold, however, the model stops working, assuming that all characters in the input string have
already been processed, and that no more characters are left unrecognized.

3. COMPUTER SIMULATION

A preliminary experiment was performed with computer simulation to check the ability of the model. The input layer of the model has a rectangular shape, and consists of 57 × 19 cells.

In this experiment, the model was taught only a small number of characters, instead of the whole set of 26 alphabetical characters. The network was trained with unsupervised learning in a similar way as for the original model (Fukushima, 1987, 1988a). The five training patterns shown in Figure 5 were repeatedly presented to the network during the training phase. The size of each training pattern was a 19 × 19 pixel array. These training patterns were presented only in this shape, and anything like a deformed version of them was not presented during the training.

The connecting strokes between characters change their shapes considerably, depending on the combination of characters. Sometimes, when a character is placed in front of or at the end of a character string, a connecting stroke might disappear there. In order to decrease the effect of such deformation, the tail ends of the connecting strokes of each training pattern are made to fade away as shown in Figure 5, rather than chopped off abruptly.

In this experiment, the same pattern “e,” shown in Figure 5, is used not only as the training pattern for “e” but also the training pattern for “l.e.” It should be noted that both “e” and “l.e” have almost the same shape when written in script style, and the only difference between them resides in their heights. After finishing the training, the same recognition cell in layer $U_{cc}$ comes to be activated by both “e” and “l.e,” because our selective attention model can recognize the shape of patterns robustly, with little effect from deformation. The two characters can easily be discriminated, however, by comparing the heights of the segmented patterns, which appear in layer $W_{co}$. Hence we can say that, in this experiment, our model has been taught to recognize, not five, but six characters.

Figure 6 shows how the response of layer $W_{co}$, in which the result of segmentation appears, changed with time when a handwritten character string “lake,” shown at the top of the figure, was presented to the input layer $U_{co}$. Time $t$ after the first presentation of the character string is indicated in the figure. The mark $\vee$ indicates the position where the center of the search area is moved. It can be seen from this figure that character “l.e” was recognized first and segmented, then followed by “l.e,” “a,” and “e.” Attention was switched just after $t = 3, 8, \text{and } 12$. The model stopped working just after $t = 21$, when all the characters in the input string had...
Recognition of Connected Characters

FIGURE 7. Some examples of input character strings which have been successfully recognized and segmented (Imagawa & Fukushima, 1990).

been completely recognized and segmented. Although the characters in the input string are different in shape from the training characters shown in Figure 5, recognition and segmentation of the characters have been successfully performed.

Figure 7 shows some examples of input character strings which have been successfully recognized and segmented. It can be seen from the figure that input strings ("act") are processed correctly, even if the spacing between the characters changes. Recognition and segmentation can be successful even if input strings ("tea") are contaminated with noise or have some missing parts. A string ("toot"), which contains two of the same character with somewhat different shapes, can also be processed successfully.

4. DISCUSSION

We have modified the original model of selective attention and extended its ability to be able to recognize connected characters in cursive handwriting.

A preliminary experiment with computer simulation, in which only a small number of characters have been taught to the model, has shown that this approach is promising. The recognition and segmentation of characters can be successful even though each character in a handwritten word changes its shape by the effect of the characters before and behind.

However, some problems still remain to be solved. For example, when characters in a word are deformed too much, a connecting part of two adjacent characters sometimes has a shape similar to a local feature of a different character. In such a case, there is a possibility of failure in recognition and segmentation. This tendency might probably increase when the characters to be recognized are increased in number. It is a future goal to test the performance of the model with a larger number of training patterns, and to fix the problems which might arise. However, we expect that these problems can be settled with some modification of the model. We believe that the use of selective attention is a correct approach for connected character recognition of cursive handwriting.

REFERENCES


