A Conceptual Development Framework for Intuitive Human Pattern Recognition

Yung-Sheng Chen*, Jiunn-Liang Lin† and Wen-Hsing Hsu‡

* Department of Electrical Engineering, Yuan Ze University
135 Yuan-Tung Road, Chung-Li 320, Taiwan, R.O.C.
e-mail: eeyschen@ee.yzu.edu.tw

† Computer & Communication Research Laboratories
Industrial Technology Research Institute
Bldg. 51, 195-11 Sec. 4, Chung Hsing Rd., Chutung
Hsinchu 310, Taiwan, R.O.C.

‡ Department of Electrical Engineering
National Tsing Hua University
101 sec. 2, Kuang Fu Road, Hsinchu 300, Taiwan, R.O.C.

Name and present address of the corresponding author:

Prof. Yung-Sheng Chen
Department of Electrical Engineering, Yuan Ze University
135 Yuan-Tung Road, Chung-Li 320, Taiwan, Republic of China
TEL: +886-3-4638800 (ext. 409)  FAX: +886-3-4639355
e-mail: eeyschen@ee.yzu.edu.tw
List of Figures

1  Illustrations of tolerance (or perceptual constancy) of pattern scalings. (a) Maximum receptive field ($RF_{\text{max}}$) and minimum receptive field ($RF_{\text{min}}$), (b) a standard learned pattern, (c) an enlarged case, (d) a lessened case, (e) an over-ranged case, (f) an under-ranged case, and (g) the case of the thick variation of the primitives of the learned pattern. .................................. 23

2  A higher complexity pattern positioned in $RF_{\text{min}}$ is not easily recognized (compare it to Figure 1.) .......................................................... 24

3  A reverberating circuit. Such recurrent circuits allow neural activity to be maintained for quite some time and may, according to some, form a basis for short-term memory (Lorete de Nó, 1938). .............................. 24

4  A stimulation-response model including the generalization process. .... 24

5  Conceptual modeling of the intuitive neural-like network (I-net). .... 25

6  A simple proportional relationship. Here $s_i$ and $p_i$ denote the size and position of a pattern on the $i$th layer in I-net, respectively. ................. 26

7  The result is the same if the information is directly propagated to layer 3 from layer 1 or is propagated to layer 3 from layer 1 through layer 2. . . 26

8  Created results in different layers when (a) the size of the seed pattern is $64 \times 64$, and (b) the size of the seed pattern is $48 \times 48$. ..................... 27

9  Examples of eight original patterns and the resultant patterns after the generalization process. ......................................................... 28

10 (a) The interconnections in the monitor for a pattern to be learned. (b) The interconnections in the monitor for all patterns. ................................. 29

11 Results when patterns of the finished generalization process are transmitted to the other layers of I-net. The size of seed patterns is $64 \times 64$. . . 30

12 Some test patterns. Where “*” denotes that the pattern is treated as noise-heavy or as an unknown pattern by the I-net. ............................. 31
List of Tables

1. Results of a psychological experiment using the method of limits. Where “NO.” means the iteration number of generalization by the method described in the text, “Y” means that the subject says that we can read this pattern, and “N” means that the subject says that we cannot read this pattern. The illustrated pattern size is 64 x 64. .......................... 32

2. Relationships between the stimulation (pattern, background) and the stored object (pattern, background). .............................. 32
ABSTRACT

In human visual perception, there is an intuitive tolerance (perceptual constancy) for per-
ceiving raw stimulus patterns which may be rotated, scaled, deformed, or noisy to some
extent for a learned pattern. Such an intuitive property of perception is a major feature
of human pattern recognition, where a recognition mechanism (consider the character
recognition for example) may not necessarily follow in a stroke-based (feature-oriented)
scheme; rather, it may follow in a whole-image-based (spatial-oriented) scheme. In this
paper, we present a conceptual development framework for intuitive human pattern
recognition. The implementation model is called I-net and basically consists of three
parts: an attention mechanism, a generalization mechanism, and a recognition mecha-
nism. The paper begins with describing the intuitive properties involved in recognizing
2-D binary patterns, following an introduction to intuitive human pattern recognition.
Then the details of the I-net are described and experimental results are presented. The
paper concludes that the proposed framework provides another direction for approaching
human pattern recognition.

Keywords:

Human Visual Perception, Perceptual Constancy, Pattern Recognition, Selective Atten-
tion, Generalization Process.
1 INTRODUCTION

In our daily life, the world is filled with disorderly stimuli. Within the organization of human perception, some stimuli are organized into patterns which are meaningful to people. Although the stimulus may be not exactly like a complete pattern, it still can be perceived as such by people under the organization principle (Morgan and King, 1966). The main property of this principle is known as perceptual closure in psychology. In our perceptual processes, it tends to organize the world by filling in gaps in stimulation so that we perceive a whole object and not disjointed parts. Based on the organization principle, we have successfully proposed some approaches to interpretive models of line continuation (Chen and Hsu, 1989; Chen et al., 1993), the closure process for line patterns (Chen and Lin, 1995; Chen and Lin, 1998), computer vision of a dotted image (Chen and Hsu, 1995a) a color blindness plate (Chen and Hsu, 1994; Chen and Hsu, 1995b). The main goal of these fundamental researches is to guide us in further study of so-called intuitive human pattern recognition. To better understand what intuitive pattern recognition means, the tree-recognition example presented by Abu-Mostafa and Psalties may be adapted for an introduction as follows.

Are all these objects trees? Even a young child can answer correctly; a conventional computer, however, has enormous difficulty in doing so. Although there is a fair amount of regularity among the trees shown (each has a trunk and branches, for example), there is also a major component of arboreal irregularity among them. A generalized definition of a tree based on the underlying regularity could lead to erroneous identifications (such as mistaking a telephone pole, which has a “trunk” and “branches,” for a tree). Hence, any effective program designed to recognize trees would essentially have to be a list of all types of trees, which cannot be done in a few lines of computer code (Abu-Mostafa and Psalties, 1987). Such a recognition problem, Abu-Mostafa and Psalties defined as a random problem; and they suggested the solution to random problems lies in memorizing all possible solutions rather than bottom-up and/or logic processing. This suggested solution may be confirmed by so-called intuitive thought in psychology (Piaget, 1952): During the stage of intuitive thought, the child begins to group objects and events into classes, but the grouping is based upon some dominant and outstanding perceptual characteristic of the situation. The child, at this stage, is not able to make
general statements; his thought is tied to the immediate perceptual characteristics of the situation. Experiments on conservation illustrate the type of thought characteristic of this stage of development (Piaget, 1952).

In reality, such random problems may include optical character recognition (OCR). In the past, various systems have been proposed and widely used (Ogawa and Taniguchi, 1988; Chen et al., 1988; Govindan and Shinaprasad, 1990; Impedovo et al., 1991; Lee and Chen, 1992). In general, an OCR system may be divided into preprocessing, feature extraction, and matching. The major preprocessing techniques available are smoothing, normalizing, thinning, and line-segment approximation (Impedovo et al., 1991), which will facilitate the extraction of features, such as feature points and strokes. We posit the feature extraction performed is based upon a feature-extractable property. Usually, there is a limitation in the conventional method, that is, to process pattern recognition, the concerned pattern should possess a feature-extractable property. It means that the concerned pattern should either be well-defined or have a good statistical characteristic in its noisy form. Feature extraction from such patterns can be performed well by a currently known method (Impedovo et al., 1991). However, consider the complex dotted image, including a meaningful dotted figure. Pattern recognition, using the conventional method, will become difficult or impossible, since the feature-extractable property does not exist in such an image (Chen and Hsu, 1994; Chen and Hsu, 1995a; Chen and Hsu, 1995b). Such a random problem may also be regarded as one type of those mentioned in reference (Abu-Mostafa and Psalties, 1987).

In this paper, we consider another random problem, recognition from a degraded stimulus pattern, which is simply described as follows. In human visual perception, there is an intuitive property; the inherent tolerance (or perceptual constancy) of perceived degraded stimulus patterns which may be rotated, scaled, deformed, or noisy to some extent for a known pattern. Such an intuitive property of perception is a major feature of human pattern recognition, where recognition (consider character recognition for example) may not necessarily follow a stroke-based scheme\(^1\) (satisfying the feature-extractable

\(^1\)As mentioned previously, the stroke-based scheme has been widely and often used in the currently known OCR systems, where computational problems that lend themselves to algorithmic solutions share a characteristic property: they are structured and based on a set of features. Most of the problems currently being solved with computers belong to this class of structured problems, and it is now a
property); rather, it may follow a whole-image-based scheme (solving the random problem emphasized by Abu-Mostafa and Psalties (1987)). The recognition mechanism is operated by top-down processing (or active processing) rather than bottom-up processing (or passive processing). And the capability of a whole-image-based scheme is similar to the concept of the solution to random problems lying in memorizing all possible solutions.

After the introduction, the other parts of this paper are organized as follows. The intuitive properties for recognizing 2-D binary patterns are comprehensively described in the next section. Then the design of an I-net for intuitive pattern recognition system is outlined. The mechanisms of the I-net, including an attention mechanism, a generalization mechanism, and a recognition mechanism, are presented consecutively with some illustrations and experiments. Concluding remarks are given in the final section.

2 INTUITIVE PROPERTIES OF RECOGNIZING 2-D BINARY PATTERNS

From the viewpoint of psychology (Morgan and King, 1966), the term perception generally refers to the awareness of objects, qualities, or events stimulating the sense organs it; also refers to a person’s experience of the word. The term perceptual constancy generally refers to the tendency of objects to be perceived in the same way despite wide variations in the energies impinging upon the receptors. Hence, we can say that the stimulation-response performed fast intuitively by the human brain is due to the fact that all the possible solutions for events in the real-world have been memorized or generalized in our brains. The term memorization is usually called long-term memory components which are distributed in the cerebral cortex from the viewpoint of anatomy (Reynolds and Flagg, 1983). And the term generalization is regarded as an internal mental process. We group these two terms under the name generalization process which will be dealt with universal practice for computer programmers to look for an algorithm whenever they have to solve a problem. However, problems such as pattern recognition in natural environments lack the structure that would allow simple algorithmic solutions. It is this departure from the properties of structured problems and the methods for solving them that characterize a random problem (Abu-Mostafa and Psalties, 1987).
more clearly, later.

Since only 2-D binary patterns are focused on in this paper and according to the explanations of human perception mentioned above, we present several fundamental intuitive properties of binary pattern recognition which are easily performed by the human brain. These properties are concerned with the tolerance of perceiving raw stimulus patterns either rotated or scaled to some extent for a learned pattern. According to the tolerance to rotations and scalings, the intuitive properties hold automatically for tolerance of other deformations, distortions or noisy interferences. These properties are used for the construction of the I-net presented later, and are illustrated as follows.

2.1 Tolerance of pattern scalings

If a pattern is learned, human beings can easily recognize its enlarged or reduced forms. The bases for forming enlarged or reduced cases are twofold: (1) the linear variation of the size of the pattern, and (2) the linear variation of the thickness of the primitives of the pattern. These two cases can also happen simultaneously. This property is subject to two factors: (1) the range of the receptive field of the sense organ, i.e. the effective stimulated region on the retina of the eyeball, and (2) the complexity of the stimulus pattern.

For instance, we assume a maximum receptive field and a minimum receptive field as shown in Figure 1a, and let a standard learned pattern be shown in Figure 1b. If an enlarged or a lessened stimulus pattern (shown in Figure 1c and 1d respectively) stimulates our visual system, we can intuitively recognize them. However, if the enlarged stimulus pattern is over-ranged (shown in Figure 1e) or the lessened stimulus pattern is under-ranged (shown in Figure 1f), then the visual system will have difficulties in recognizing what the stimulus pattern is. The system should self-adjust its position or distance corresponding to the given pattern and try again. In addition, if the thickness of the primitives of the learned pattern is changed finitely, such a new pattern (shown in Figure 1g) may also be perceived easily.

Note that, for the same size of receptive field, the degree of tolerance of pattern scalings to the higher-complexity pattern is smaller than that to the lower-complexity pattern. For example, in Figure 2, the reduced version of another higher-complexity
learned pattern cannot be easily recognized by the minimum receptive field of Figure 1a. The complexity measure of a pattern can be obtained by many methods (Rosenfeld and Kak, 1982).

2.2 Tolerance of pattern rotations

The property introduced here is fundamentally based on the following three facts: (1) the world of perception is, within limits, quite plastic and modifiable (Morgan and King, 1966); (2) the human body is naturally formative (e.g. the head is on the top of human body) and is naturally soft (e.g. the waist and the neck of the body are often bent slightly and softly); and (3) the neural network in our brain is highly adaptable (Kohonen, 1988). Hence, we can say that the adaptation of the neural network for human visual perception follows the stimulus pattern received by the retina of the eyeball. Since the received (learned) pattern is usually not fixed due to (1) and (2), the internal adaptation of the neural network is thus performed gradually and will reach a stable point eventually. Furthermore, in our brain, the network may have a special internal process to generate a large number of internal representations of a learned pattern (e.g. various rotated versions) within the long-term memory (Reynolds and Flagg, 1983). Therefore, although people need not learn all the possible forms for a pattern, the internal process may facilitate the human body’s easily adapting to its environment.

Based on the tolerances (pattern scalings and rotations) mentioned above for human intuitive pattern recognition, three other properties are also easily observed. They are tolerance of pattern slants, tolerance of pattern deformations, distortions or noisy interference, and tolerance of pattern variations with various strips in the primitives of the pattern, respectively. Accordingly, to induce these properties, we not only should consider the effective range of the receptive field and the range of scalings and rotations, but also must apply the closure property of perceptual grouping (Morgan and King, 1966) to the intuitive pattern recognition.

2.3 Generalization process

In this subsection, we first introduce the organization theory developed for the investigation of molecular changes in learning and memory. The theory states that memory is
carried in the nervous system by both short-term and long-term changes (Hebb, 1949). More than this, the long-term changes depend on the prior presence of short-term activity. According to this theory, when an organism is first exposed to a learning situation, activity begins in chains, or collections, of neurons in the brain; this short-term activity will not persist unless the learner repeats the situation. In this short-term stage, the activity is usually considered to be nerve impulses travelling around reverberatory, or self-excit ing circuits. The circuit is shown in Figure 3 (Lorete de Nó, 1938). This activity, if it lasts long enough, changes the synaptic relations among cells so that they are permanently reorganized. Later, after reorganization, if once a cell of the circuit is fired, the others organized with it also fire. This relatively permanent reorganization is, according to the theory, the way memories are stored in the brain.

Figure 4 shows a stimulation-response model for explanations of the previous arguments. When a stimulus pattern is presented on this model, the “sensory mechanism” will record the corresponding stimulus signals and cause “short-term memory” activity. At that time, “response” may occur instantaneously. If “attention” focuses on the event, the “learning mechanism” will be fired in which the synaptic relations among cells are changed. And the permanent reorganizations of this event will be stored in the “long-term memory” if the activity in the “short-term memory” lasts long enough. Thus, the stimulation-response relationship for this stimulus pattern is built completely.

According to the internal mental process mentioned above and the intuitive properties of pattern recognition presented previously, we have a “generalization process” connecting with the “long-term memory”. When the “generalization process” performing internally for a permanent reorganization is accomplished, many other relative organizations will also be formed and be stored permanently in the “long-term memory”. Thus the tolerance for recognizing the learned pattern is enhanced for the so-called intuitive pattern recognition. Essentially, the model shown in Figure 4 can be used to explain the perceptual constancy in psychology (Morgan and King, 1966).

For recognition of a 2-D binary pattern, the generalization process can be described by the following procedure:

Step 1. Learn a noiseless pattern and assign a specified pattern code \(P\) by learning-with-a-teacher scheme.
Step 2. Find the MA (Media Axis) (Chen and Hsu, 1993) for this pattern. The MA provides the fundamental structure for the learned stimulus pattern.

Step 3. Dilate each point of the MA until the generated pattern is blurred and cannot be perceived as the original pattern $P$. Then obtain a **Maximum Tolerance Range in Scaling** ($MTRS$) for this pattern ($P_{MTRS}$) on the specified receptive field.

Step 4. Rotate the pattern ($P_{MTRS}$) through a range of degrees until recognition is affected or delayed. Then obtain a **Maximum Tolerance Range in Rotation** ($\theta_{MTRR}$) for the learned pattern $P$ on the specified receptive field. That is, we have generated a set of rotated ($P_{MTRS}$)s whose rotating degree, say $\theta(g_p)$, falls in the range ($\theta_{MTRR}$) for the pattern $P$ recognition with respect to the specified receptive field. Here we define $\phi(g_p) = P_{MTRS}$ so that the unrotated version of $g_p$ is the pattern $P_{MTRS}$.

Step 5. Project the MA found in Step 2 on each receptive field ($RF$), and for each $RF$ do Step 3 and Step 4.

Hence, assume that the model (or visual system) is fixed by the receptive range $[RF_{min}, RF_{max}]$, according to the above procedure, for a learned pattern $P$ we will generate all the possible solutions (Abu-Mostafa and Psalties, 1987) for recognizing the various deformed but tolerable forms of this pattern. The set of generalized solutions of this pattern $P$ is thus expressed by

$$G(P) = \bigcup_{j=RF_{min}}^{RF_{max}} \{ g_p : \phi(g_p) = P_{MTRS_j} \text{ and } \theta(g_p) \in \theta_{MTRR_j} \}.$$  (1)

In the recognition stage, the input stimulus pattern will be matched with each element $g_p$ in $G(P)$ for all learned patterns. If any one of the elements in a set of $G(P)$ has the highest score above a mental threshold, then the pattern is recognized and given a pattern code $P$.  

11
3 I-NET

3.1 Introduction to I-net

The concept of I-net, based upon the previous generalization process, is modeled as a hierarchical structure consisting of many layers which are located as in Figure 5. Each layer can be treated as both a receiving layer which is used to receive the stimulation, and a matching layer which is used to produce the matching result. Each layer consists of a set of receptive-fields which consist of a set of neuron-modules. The lower the layer is, the larger the corresponding receptive-field is. All the corresponding receptive-fields over all the layers will form into a receptive-cone and are connected together. In the receptive-field, each neuron-module consists of many memory cells where the states of the corresponding positions of all learned patterns will be stored. Therefore, any learned pattern will exist in all layers of I-net, but will have different sizes among them.

During the learning stage, the stimulation will be located by the attention mechanism, and its size will be found, too. The size facilitates choosing a layer to be activated. After learning, the code of the pattern will be automatically registered into the memory cells of the activated neuron-modules of the specified layer. The generalization process occurs in the specified layer. The result is reached, after the process has the maximum tolerance of other deformations, distortions or noisy interferences. Then, the result and code are broadcast into the memory cells which belong to the same activated receptive-cone containing at least an activated receptive-field which contains the activated neuron-modules. Similar patterns of all sizes are created after the broadcast.

During the recognizing stage, the input pattern is located by the attention mechanism, too. It will be matched with every layer by the recognition mechanism. If a set of memory cells containing the same code in a layer has the highest matching score above a mental threshold (Chen and Hsu, 1995b), and agrees with the voting result of the recognition mechanism which is used to monitor the consistency of the structures of two patterns matched to each other; then the pattern is recognized as the code. Otherwise, the input pattern will be categorized as an unknown pattern.
3.2 Interconnections of layers in I-net

There are two characteristics of the interconnections in the I-net: (1) the structures of patterns formed on the other layers in the I-net after transmission is the same as that of the source pattern, (2) regions of stimuli will be formed on all layers in proportion depending on the locations of layers in the I-net. Although the characteristics exist between any two layers in the I-net, the interconnections only exist in neighboring layers. The transmission is like wave propagation. The source layer will transmit information to neighboring layers through the interconnections. Then, the information will be propagated to the neighboring layers again and treated as new source layer information.

The interconnections stated above can support the fundamental intuitive property in scaling for human visual pattern recognition. As to the fundamental intuitive property in rotation, it will be supported by the directions of geometrical rotations of information on each layer. There exists a plane which consists of many layers for each layer in the I-net in order to keep the results consistent with the rotations of pattern in all degrees. In this paper, we are only concerned with the interconnections with the scaling property.

*Figure 6* shows the simple proportional relationships \( \frac{s_2}{s_2} = \frac{p_2}{p_2}, \frac{s_3}{s_3} = \frac{p_3}{p_3} \), and \( \frac{s_1}{s_1} = \frac{p_1}{p_1} \). Obviously, the relations exist in the patterns on any three layers in the I-net under the fundamental considerations as described above. The I-net is a multilayered structure. Each layer keeps patterns with a certain size corresponding to the location of the layer in the I-net. Extract the pattern from each layer, it will form a triangle-relationship in size, when one-dimensional cases are considered. From the relationship, see *Figure 7* for reference, we find that there is the same result if the information is directly propagated to layer 3 from layer 1 or is propagated to layer 3 from layer 1 through layer 2. Therefore, we could simulate the I-net by several layers, since it has numerous layers to keep patterns of all sizes for visual perception in the human brain. And it does not influence the correct results on layers.

In our experiments, we chose twenty-eight patterns which are divided into two groups, one has an image size 64 × 64, and the other has a size 48 × 48. For simplicity, the I-net is constructed of four layers whose sizes are 64 × 64, 48 × 48, 32 × 32 and 16 × 16. When a pattern enters the I-net, it will create the other three patterns with different sizes after being transmitted through the interconnections between layers. *Figures 8a* and *8b* show
one of the results where the sizes of seed patterns are $64 \times 64$ and $48 \times 48$, respectively.

4 ATTENTION MECHANISM

To perform computer vision on the concerned image, the design of the attention mechanism is principally based upon the concept of selective attention in human visual perception (Morgan and King, 1966). From the psychovisual aspect, two external factors, called concentration and numbers of stimuli in a certain area of image, are adopted in this paper for binary image processing. They can support our model to perform the selective attention. The corresponding algorithms will be (1) finding the center of attention, and (2) getting the size of the input pattern.

4.1 Finding the center of attention

The center of attention is the most attractive point in the view field. Eyes always are moving to locate that point at the center of the scene and cover the full pattern. To locate the most attractive pattern is the main work of the attention mechanism. Under the two external factors, concentration and number of stimuli, a neural network including input plane and S-plane is designed to accomplish this work.

Each block of cells in the input plane is connected to a cell in the S-plane, whose size is the maximum size for a pattern. And the position of the cell in the S-plane is the position of the center of the block related to the input plane. The weights on each block form into a 2-D gaussian shape.

The output of the S-cell is the sum of all the excitatory inputs weighted with the interconnecting coefficients. Following the S-plane, there exists a subnet, called MAXNET (Lippmann, 1987), so that the position of the cell with maximum output in the S-plane can be located. The position is the center of attention.

With the S-plane and MAXNET, we get the position of the input pattern in our view field if there is the only one pattern. The center of the scene in the eye will be moved onto it. It means that the input pattern has been attended to. Of course, the information of the input pattern’s size can also be obtained at the same time.
4.2 Getting the size of the input pattern

Getting information about the input pattern is done in the learning path. During the learning stage, we assume that there is no noise and that all the stimuli belong to the learning pattern. Then we have the following procedure for getting the size of an input pattern.

Step 1. Set $(x_c, y_c)$ to the center of attention found by the MAXNET.

Step 2. Produce a window centered at $(x_c, y_c)$, whose size is the smallest image size in the I-net.

Step 3. While there are stimuli outside the window, enlarge the window to the next larger size of image in the I-net.

Step 4. Set input pattern’s size to the size of the window.

5 GENERALIZATION MECHANISM

Generalization has been studied for a long time by many psychologists. There are many psychological experiments to prove its existence in human perception. In this section, we introduce our ideas on generalization in human pattern learning and implement the generalization mechanism based upon some psychological experiments.

For human visual pattern recognition, the pattern becoming thicker may be recognized correctly until the pattern cannot clearly express the original. However, there exists a problem of where the boundary is which determines whether one can recognize the pattern. For the answer, we have done psychological experiments with the method of limits to find the relation between the recognized results and the extents to which the patterns are thickened.

Experiments start by presenting gradually thicker patterns until the patterns are too thick to be recognized correctly. Repeating the experiment for many different people, we made a table of the thickened numbers and recognized results of the patterns. Then the boundary was obtained. The thinner or bigger the pattern, the higher are the thickened
numbers. That is in agreement with our expectations. The thinner and the bigger patterns own more space for being thickened. This is the so-called generalized space\textsuperscript{2}.

In our experiments on handwritten Chinese characters, see Table 1, we find that any patterns can be recognized under some boundaries. The thinner or the bigger the pattern is, the higher the iteration numbers are. The results show that any pattern has its “limit” of generalization space for human pattern recognition. Figure 9 shows some resultant patterns, which will be sent properly to each layer of the I-net for further pattern recognition, based upon the proposed psychology-oriented generalization process. Such a procedure can be regarded as the so-called formation of long-term memory. A more concrete algorithm has been presented in (Chen and Hsu, 1995b), that includes dilation, thinning, and distance computation to perform the pattern generalization. In this algorithm, the spatial distance between the input pattern and the generalized pattern is computed and used as the criterion of stopping condition. The operation of pattern generalization is stopped when the spatial distance < 1.

6 RECOGNITION MECHANISM

During the recognizing stage, the input pattern will be first selected by the attention mechanism, then it will be transmitted to the I-net and recognized by the recognition mechanism of the I-net. The recognition mechanism consists of two parts: competition and monitor. The part of competition is like the traditional neural networks except for the matching function table, which will be presented later. The monitor is used to check the fundamental structure of an input pattern.

An input pattern is matched with all the generalized patterns stored in the I-net, like

\textsuperscript{2}For a binary image, the generalization is a growing process for stimuli in order that the pattern will be thick enough to tolerate all the deformations and noises. But where is the boundary of the growth? It depends on the complexity of a pattern. The simpler the pattern is, the more iterations the generalization process of the pattern needs. Generally speaking, the complexity function in image processing is related to the presence of equal parts, periodicities, and symmetries in patterns. However, in the generalization process, the complexity depends only on the generalized space of the pattern but not on the presence of equal parts, periodicities, and symmetries in patterns. The larger the generalized space of the pattern is, the lower the complexity of the pattern is, and the more iterations the generalization process of the pattern will need.
the recall operation in human pattern recognition. The one with the highest matching score may be treated as the only one for possible matching. If the highest matching score is over a mental threshold and it is in agreement with the monitor’s positive result; then the code of the voted generalized pattern, assigned by the learning mechanism, will be output to represent the input pattern. Otherwise, the input pattern will be regarded as an unknown pattern which was never learned before.

6.1 Competition: winner takes all

Competition is frequently used in neural networks, e.g., ART, Hamming net. The learned pattern in the I-net with the highest matching score will be picked out for the input pattern in the recognizing process. How to calculate the matching scores of the learned patterns? That is the critical part for every neural net. It may be divided into two parts for introduction, i.e., (1) the degree of lateral interaction, Mexican hat, and (2) matching function.

Neurons, are not isolated from other neurons. There exists interactions between the neurons (Kohonen, 1988). The lateral interaction, Mexican hat, has been extensively documented in sensory systems and is very pronounced in the retina of the eye. All the neural cells will be affected if a cell receives the excitatory signal. The lateral area has the excitatory effect. This kind of lateral interaction is a very important factor in studying the activity of neurons, and it frequently makes a different illusion in human visual perception from the real image.

Usually, the matching score is calculated by the weighted connections between the input layer and the output layer. The weighted values are fixed during calculating. However, there are two reasons leading us to give it up. One is the proportion of the object and background of a binary pattern, and the other is the closure in visual perception. The closure is a perceptual process tending to organize the world by filling in gaps in stimulation so that we perceive a whole object and not disjointed parts (Morgan and King, 1966). This “filling in” thus forms a tendency to complete in perception what is physically an incomplete pattern or object. Hence both pattern and background components in a receptive field should be concerned in the I-net so that the closure property can be possibly performed. Accordingly in our matching rule, we hypothesize that, in
the interconnections between neurons, there exists a relationship between the stimulation (pattern, background) from input and the stored object (pattern, background) in the I-net, as shown in Table 2.

Table 2 states four possible cases when a stimulation is compared to an internal stored object. (1) If the stimulus pattern matches the stored pattern, then the output has $Z_1^+$ response. This is a reasonable match case. (2) If the stimulus background matches the stored background, then the output has $Z_2^+$ response. This is also a reasonable match case. (3) If the stimulus pattern matches the stored background, then the output has $Z_3^-$ response. This is an unreasonable match case such as noisy interference. And (4) if the stimulus background matches the stored pattern, then the output reveals $Z_0$ response. This falls in the area of closure such as the dotted pattern or a special font. Hence $Z_1^+$, $Z_2^+$ are positive signals for excitation, $Z_3^-$ is a negative signal for inhibition, and $Z_0$ is silent, respectively. Each input stimulation will result in actions of some neurons around the stimulated neuron (used to store the internal generalized object), which are distributed as the “Mexican-hat function” due to the lateral interaction between neurons (Kohonen, 1988). By properly selecting the values of $Z_1^+$, $Z_2^+$, $Z_0$, and $Z_3^-$, a cooperative and competitive operation can be performed; and a feasible mental threshold for performance of the closure property is possibly decided by a series of psychological experiments or a special algorithm.

Based on this significant relationship in Table 2, we have successfully developed so-called closure weights being applied to the recognition of color blindness plate (Chen and Hsu, 1995b) and used in the current I-net. In this application, a specified threshold, $T_m = 0.6$, is used to recognize the selected pattern. As a note, for such a recognition system, the adoption of a threshold value depends strongly on the system complexity and the matching function. This agrees with the findings in human visual perception (Morgan and King, 1966).

6.2 Monitor

A pattern may be deformed but its fundamental structure should not be changed if it still can be recognized. The monitor is used to monitor the completeness of the fundamental structure of certain learned patterns which the input pattern contains. If
some parts of the fundamental structure of a learned pattern are lost corresponding to the input pattern, the code of the learned pattern will be rejected as not representing the input pattern in the recognizing process. How to find the fundamental structure for a pattern? What is the fundamental structure of a pattern? This problem can be solved by a neural network, called MATNET (Chen and Hsu, 1993), which performs medial axis transformation. The structure of MATNET is analogous to that of the retina (Morgan and King, 1966), which lines the interior of the eyeball. Although the found medial axis (MA) for a pattern using MATNET does not have the connecting property, it plays an important role in the generalization process of our I-net.

Figure 10a shows the interconnections in the monitor for a pattern to be learned. During the learning stage, the input layer presents the input pattern, and the result of MATNET of the input pattern will be presented on the hidden layer. All cells within each disk of the input layer are connected to a corresponding cell of the hidden layer. Later, during the recognizing stage, any cell on the input layer will excite the the connected cell on the hidden layer if it is excitatory. Behind the hidden layer, there is a monitor cell which will connect to the cells of the hidden layer whose positions are the result of MATNET of a learned pattern. In other words, when an input pattern is presented on the input layer, the mechanism will monitor whether the input pattern includes the completeness of the fundamental structure of the learned pattern in the I-net. If the monitor cell is excitatory, it means that the input pattern contains the fundamental structure of the learned pattern, otherwise it displays a negative result. Different patterns own the different structures. Therefore, there is a plane consisting of many layers which own the results of MATNET for all learned patterns as shown in Figure 10b.

6.3 Experimental results

So far, the whole conceptual framework for an intuitive human pattern recognition (I-net) has been presented completely. The I-net is implemented on a SUN-3/160 workstation. Figure 11 shows the generalization results which are transmitted into the corresponding layers of the I-net, where the size of seed patterns is $64 \times 64$. Some test patterns shown in Figure 12 are recognized successfully except for those patterns (they are treated as
noise-heavy or unknown patterns by the I-net) marked by “*”.

Note that experiments of rotation have not yet been involved in our system, however, this will be our main goal in the near future, to investigate a practical system. In such an advanced program, the complexity of system and the capacity of generalized patterns stored in the system will be studied and analyzed. Accordingly, the proposed conceptual framework has provided us another research direction for investigating human pattern recognition.

7 CONCLUSIONS

In human visual perception, there exists an intuitive property perceiving degraded stimulus patterns which may be rotated, scaled, deformed, or noisy to some extent for a known pattern. It may be regarded as a subset of random problems in pattern recognition as indicated by Abu-Mostafa and Psalties (1987). They suggested the solution to random problems lies in memorizing all possible solutions rather than bottom-up and/or logic processing. To further explore the mentioned intuitive property in human pattern recognition, in this paper, a conceptual framework for intuitive human pattern recognition is developed and presented. The physical modeling for the conceptual framework is called I-net, which basically consists of three parts: an attention mechanism, a generalization mechanism, and a recognition mechanism. For simplicity, a pattern is learned by a supervisor scheme, and its all generalized patterns are stored in I-net for pattern recognition. According to our experiments, even though apparent rotation tolerance is not involved in the current I-net, we find that the I-net possesses quite a large tolerance for deformations and noises in recognizing a 2-D binary pattern (Chinese characters for example) due to the generalization mechanism.

Some properties may be noted in the conceptual framework of the intuitive human pattern recognition (I-net) system. The recognition may not strongly follow a stroke-based scheme (satisfying the feature-extractable property), rather, it may follow a whole-image-based scheme (providing the solution to random problems emphasized by Abu-Mostafa and Psalties (1987)). The recognition mechanism is operated by top-down processing (or active processing) rather than bottom-up processing (or passive processing). The capability of a whole-image-based scheme is similar to the concept of the
solution to random problems lies in memorizing all possible solutions. As a summary of the proposed framework, the attention mechanism selects the proper whole-image for pattern recognition. The generalization process supports the reasonable “memorizing possible solutions” for a learned pattern. The four matching weights in pattern-background relation table support “closure perception” and thus further confirm the useful whole-image matching scheme. The structure elements of a pattern (obtained by MATNET) support the “generalization” in long-term memory and “structure monitor” in pattern recognition. Accordingly, the proposed framework provides another direction for approaching human pattern recognition.

ACKNOWLEDGMENTS

This work had been supported in part by the National Science Council of Republic of China under Grant No. NSC 79-0408-E007-12.

References


Figure 1  Illustrations of tolerance (or perceptual constancy) of pattern scalings.  (a) Maximum receptive field ($RF_{\text{max}}$) and minimum receptive field ($RF_{\text{min}}$), (b) a standard learned pattern, (c) an enlarged case, (d) a lessened case, (e) an over-ranged case, (f) an under-ranged case, and (g) the case of the thick variation of the primitives of the learned pattern.
Figure 2  A higher complexity pattern positioned in $RF_{\text{min}}$ is not easily recognized (compare it to Figure 1.)

Figure 3  A reverberating circuit. Such recurrent circuits allow neural activity to be maintained for quite some time and may, according to some, form a basis for short-term memory (Lorete de Nó, 1938).

Figure 4  A stimulation-response model including the generalization process.
Figure 5  Conceptual modeling of the intuitive neural-like network (I-net).
Figure 6  A simple proportional relationship. Here $s_i$ and $p_i$ denote the size and position of a pattern on the $i$th layer in I-net, respectively.

Figure 7  The result is the same if the information is directly propagated to layer 3 from layer 1 or is propagated to layer 3 from layer 1 through layer 2.
Figure 8  Created results in different layers when (a) the size of the seed pattern is $64 \times 64$, and (b) the size of the seed pattern is $48 \times 48$. 
Figure 9  Examples of eight original patterns and the resultant patterns after the generalization process.
Figure 10  (a) The interconnections in the monitor for a pattern to be learned.  (b) The interconnections in the monitor for all patterns.
Figure 11  Results when patterns of the finished generalization process are transmitted to the other layers of I-net. The size of seed patterns is $64 \times 64$. 
Figure 12  Some test patterns. Where “∗” denotes that the pattern is treated as noise-heavy or as an unknown pattern by the I-net.
Table 1  Results of a psychological experiment using the method of limits. Where “NO.” means the iteration number of generalization by the method described in the text, “Y” means that the subject says that we can read this pattern, and “N” means that the subject says that we cannot read this pattern. The illustrated pattern size is $64 \times 64$.

<table>
<thead>
<tr>
<th>NO.</th>
<th>雲</th>
<th>水</th>
<th>意</th>
<th>本</th>
<th>天</th>
<th>土</th>
<th>不</th>
<th>刀</th>
<th>夫</th>
<th>寶</th>
<th>平</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>1</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>2</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>3</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>4</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>5</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>6</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>7</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>8</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>9</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>10</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>11</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>12</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>13</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>14</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>15</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
</tbody>
</table>

Table 2  Relationships between the stimulation (pattern, background) and the stored object (pattern, background).

<table>
<thead>
<tr>
<th></th>
<th>stimulus pattern</th>
<th>stimulus background</th>
</tr>
</thead>
<tbody>
<tr>
<td>stored pattern</td>
<td>$Z^+_1$</td>
<td>$Z^+_0$</td>
</tr>
<tr>
<td>stored background</td>
<td>$Z^-_3$</td>
<td>$Z^-_2$</td>
</tr>
</tbody>
</table>