Sequential Recognition of Superimposed Patterns with Top-Down Selective Attention

Byung Taek Kim
Brain Science Research Center
Korea Advanced Institute of Science & Technology
Yusong-gu, Taejon 305-701 Korea
btaek@yahoo.com

Soo-Young Lee
Brain Science Research Center
and Department of BioSystems
Korea Advanced Institute of Science & Technology
Yusong-gu, Taejon 305-701 Korea
sylee@mail.kaist.ac.kr

Abstract

Top-down attention is a cognitive mechanism to filter out irrelevant information from sensory input. Unlike bottom-up attention based on the sensory signal itself the top-down attention process is originated from the higher brain, which consists of previous knowledge about the sensory signals. A simple computational model is developed for the top-down attention. In this model an attention gain coefficient is assigned to each input feature, and all the attention gain coefficients are dynamically adjusted based on previous knowledge. A mulitilayer Perceptron is used to model the knowledge in the higher brain. The developed model demonstrates excellent capability of extracting and recognizing each pattern sequentially from superimposed dual-class patterns studied in visual perception.

1. Introduction

In the real-world pattern classification, accurate pattern classification can be very complex and difficult. Because real-world patterns are almost always distorted by noise or spatially overlapped, the attempt for accurate recognition of noisy and superimposed patterns has been a big issue. It has been tried with a lot of different approaches. To improve pattern recognition rate, we present an approach that is based on the human selective attention.

Human beings naturally use top-down selective attention to improve their pattern recognition ability. In pattern classification, human selective attention makes us distinguish important features of input stimulus from others and focus on a critical feature of an input stimulus.

A lot of psychologists have studied the human selective attention mechanism for a long period of time ([3] – [5]). However, only a few of their theories can be applied to engineering pattern classification cases, and yet, the controversy among each psychological theory exists.

Fukushima [6] applied the human selective attention mechanism to develop his Neocognitron, which uses selective attention and attention switching algorithms. However, his model has many unknown parameters to be determined heuristically, and the performance of the system is sensitive to the parameter values. Also, the real-time application is almost impossible due to its big load of computation. For classification of superimposed patterns, Rao [7] introduced a selective attention model based on Kalman filters. However, since his model is based on linear systems, a nonlinear extension is not straightforward.

There have been a lot of approaches tried; yet there is not a single best approach using selective attention. We present a simple and efficient algorithm which gives us improved recognition rate.

2. Psychological Views of Selective Attention

Psychologists have studied human brain's attention mechanisms for many years. One of the most important issues about human brain's attention is where attention acts in the stream of information processing. There are two different approaches to this issue. One approach is input dependent, and it is called bottom-up attention. Bottom-up attention looks for low-level stimulus features, such as the amplitude of the sound signal and the color of the visual signal. The other is top-down approach, and it is relevant to the knowledge of the object and the expectation based on the previous knowledge. For example, top-down attention looks for a location where the critical task-relevant information is expected.

There are two major streams that consist the selective attention theory. One is early filtering (or early selection) theory, and the other is late selection theory. The earliest and the best-known modern theory of the selective attention is the early selection theory proposed by Broadbent [8]. He presented two auditory channels to subjects, one to each ear, and asked them to shadow one channel. In his experiment, he observed that although subjects could not recall most of what took place in the shadowed channel, they could often recall the last few seconds of input on that channel. In his theory, he proposed that the human brain briefly stores incoming stimuli, but the stimulus information fades and is neither admitted to the conscious mind nor encoded in a way that would permit later recollection unless the attention is directed toward it. As a modification to Broasbent's theory, Treisman [9] proposed that the filter merely attenuates rather than totally preventing it for further analysis. Shortly after, the late selection theory that is contrary to the early selection theory was proposed. In our opinion the most important is not the time of attention but the attention mechanism in human brain information processing.

3. Proposed Selective Attention Model using Multilayer Perceptron

According to the Broadbent's early selection theory, there is a selection filter acting between the sensory input and the working memory. As shown in Figure 1(a), the output of the attention filter is the expected input to an attended class, which may be given internally or externally. We place an attention gain layer in front of the Multilayer Perceptron (MLP) classifier (the dotted box), as in Figure 1(b). The attention gain level with one-to-one connectivity works as the filter between the input and the classifier, and the classifier MLP plays as the knowledge in the higher brain. Here, we depicted an MLP with single hidden layer, but our approach can be applied to general MLP architecture. [1,2]

First, we fix the attention gains to 1 during the MLP training so that the system can behave as an ordinary MLP. However, attention gains are to be adjusted during classification of test patterns. In adjusting gains, we use the knowledge of trained MLP to tell which input feature is necessary and important to classify a pattern. When attention gains are adjusted, they are used either to suppress the irrelevant and noisy features in the input or to enhance the relevant features in the input. In adjusting gains, the gradient descent rule is used, and the gains are used to make the input as good an example of the output class as possible. We repeat the gain adjusting process for each of all different output classes. The class of adjusted gains that gives the smallest modulation of the input is described as the class with the strongest response, and that class is chosen to be the recognized class (the exact quantitative rule and detailed explanation of the algorithm will follow below).

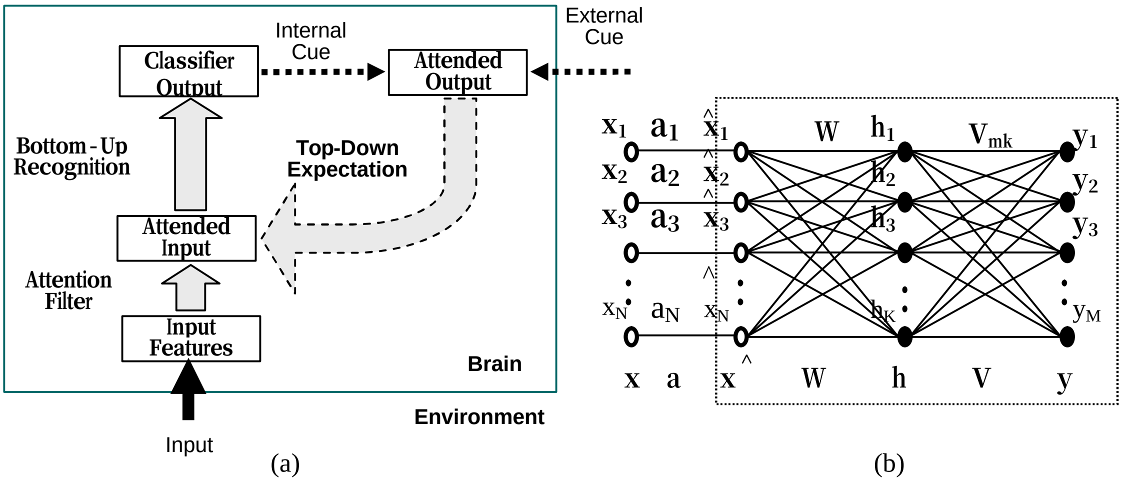
The **W** and **V** are synaptic weights of MLP, and they are adjusted via the standard training way of MLP using gradient descent algorithm with error back propagation. The process of adjusting attention gains is as follows: First, the target output vector $\mathbf{t}^{\mathbf{s}} = [t^{s}_{1}\ t^{s}_{2} \ldots\ t^{s}_{M}]^{T}$ is defined. $t_i^s = 1$ for the attention class and –1 for the rest of the classes. Then, all the attention gains are set to 1. Second, the test input is applied to the pre-trained MLP network, and the gains are adjusted to minimize the error, $E^s \equiv \frac{1}{2}\sum (t_i^s - y_i)^2$ **.** The update rule for attention gains is based on the gradient descent algorithm with error back propagation. At the (n+1)th iteration epoch, the attention gain a_k is updated as follows:

$$a_k[n+1] = a_k[n] - \eta(\partial E / \partial a_k)[n] = a_k[n] + \eta\ x_k \delta_k^{(0)}[n] \qquad (1)$$

$$\delta_k^{(0)} = \sum_j W_{jk}^{(1)} \delta_j^{(1)} \qquad (2)$$

where E is the output error, and $\delta_j^{(1)}$ is the jth attribute of the back propagated error at the hidden-layer. $W_{jk}^{(1)}$, the synaptic weight between the input $\hat{x}_k$ and the j'th neuron at the hidden layer. η is the step size, which is the learning rate.

The attention gains are to lie in [0, 1]. The summary of step-by-step selective attention algorithm is as follows:



(a) (b)

Figure 1. Proposed Network for Top-Down Selective Attention. (a) Basic concept; (b) Neural network implementation

(1) Set all the attention gains to 1, and train MLP to adjust synaptic weights (W and V).
(2) The output target vector $\mathbf{t}^s = [t^s_1\ t^s_2 \ldots\ t^s_M\ ]^T$ is defined for a class that we desire.
(3) All the attention gains are initialized to 1.
(4) Apply the test pattern, and apply the selective attention algorithm (equation (1) and equation (2)) to adjust the attention gains until the attention process converges.
(5) Compute the confidence measure M (the algorithm defined below) for each class of the given test pattern.
(6) Select the class with a minimum confidence measure M as the recognized class.

The confidence measure if defined as follows:

$$M = D_I E_O^{a} \tag{3}$$

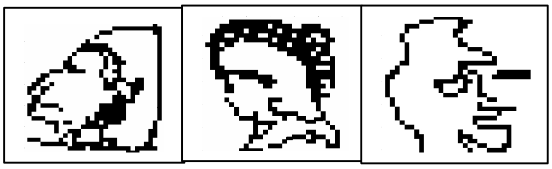
$$E \equiv \tfrac{1}{2}\sum \left\| \mathbf{y}^s - \mathbf{f}(\mathbf{x}^s, \mathbf{W}) \right\|^2 \tag{4}$$

$$E_O \equiv \sum_i [t_i - y_i(\hat{\mathbf{x}})]^2 / 2M \tag{5}$$

Where D_I is the square of Euclidean distance between two input patterns before and after the application of selective attention, and E_O is the output error after the application of selective attention. Here, we normalized D_I and E_O with the number of input pixels and output pixels, respectively. In equation (3), a is a parameter to emphasize either of D_I or E_O above. When a is 0, the confidence measure M is only determined by D_I. When a is 1, D_I and E_O are equally important to determine the confidence measure M. $\hat{\mathbf{x}}$ can be thought of as the minimal deformation of the test pattern to trigger the attended class, so the Euclidian distance between x and $\hat{\mathbf{x}}$ is a good measure for classification.

The proposed selective attention algorithm was tested on recognition of noisy and spatially superimposed image patterns. The database basically consists of 3 superimposed images, and each of which has two different output classes as shown in Figure 2 and Figure 3. 200 test patterns were generated from each of the 3 superimposed patterns using deforming methods (glass effect, wave effect, whirl and pinch effect) of image processing program's toolbox, as shown in Figure 4; The first panel shows the original image before a deforming method was applied, and the second, third, and forth show deformed images using glass effect, wave effect, whirl and pinch effect, respectively. Glass effect was performed by varying the refraction parameter (from 1.0 to 1.3 by 0.01 every time). Wave effect was performed by varying the amplitude parameter (from 0 to 9 by 0.1 every time), and whirl and pinch effect was done by varying the angle parameter (from -40° to 40° by 1° every time). Each pattern was edge-detected first, then, it was down-sampled to 32 x 32 pixels and encoded as binary pixel values. Black pixels were coded as 1, and white pixels were coded as 0. An one hidden-layer MLP was trained by back propagation using 6 training patterns, shown in Figure 3.

The numbers of input, hidden, and output neurons were 1024, 30, 6, respectively. Total of 600 test patterns were presented to the network for classification. The 600 test patterns were sorted into 5 different bins specified by a specific range of hamming distance, as shown in Table 1 below. The Hamming distance is defined as the distance between the deformed test pattern (examples shown in Figure 4) and the original pattern (shown in Figure 2)



(a) (b) (c)

Figure 2. The scaled original superimposed images: The full size images were edge-detected, down-sampled to 32 x 32 pixels, and encoded as binary pixels. (a) face and Eskimo; (b) faces of a lady and an old man; (c) trumpet player and girl face.

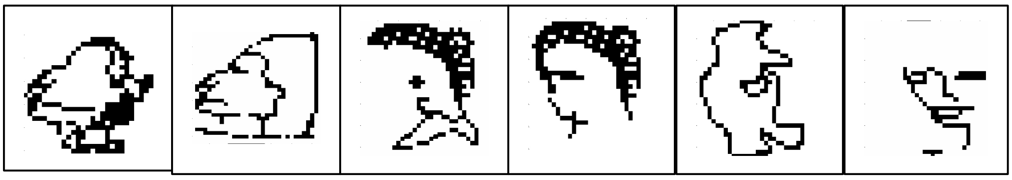


Figure 3. The Training Patterns: They were generated by separating the pattern of each class from the other in the original superimposed image (i.e., the first two images were generated from the first image in Figure 2).

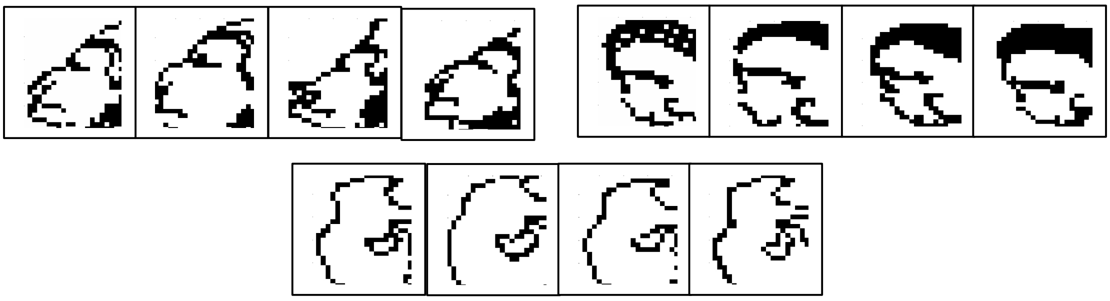


Figure 4. Examples of Test Patterns

Table 1: The Number of Test Patterns for a Specific Range of Hamming Distance

Hamming Distance	0 – 60	61 – 120	121 – 180	181 – 240	> 240
Number of Test Patterns	76	177	151	136	60

4. Attention Switching for Superimposed Pattern

Using the selective attention algorithm as described in section 3, recognition of the first class of each superimposed pattern was done successfully. However, since the test patterns are spatially superimposed, we still need to remember that there is another class to classify in each superimposed test pattern. To aid that, we present attention switching algorithm.

After the first class is recognized as depicted in section 3, attention is switched from the pixels of the recognized pattern to the remaining pixels in the test pattern. Attention switching works as follows:

The Attention gain values are set to 0 if and only if the gain values were greater than or equal to 1 after the selective attention process (if the attention gain values are greater than or equal to 1 after the selective attention process, we consider that the corresponding inputs of the gains (greater than or equal to 1) are attended during the recognition of the first pattern), and all other values are set to 1. Then, we filter the test pattern through the attention switched gains to obtain filtered test pattern, $\hat{\mathbf{x}}$, for the second round classification. Last, we apply $\hat{\mathbf{x}}$ to the standard MLP network (where all the attention gains are 1) and take the most active output class as the recognized class.

The proposed selective attention and attention switching algorithm was tested for recognition of 600 test patterns as described in section 3. Figure 5 shows examples of selective attention and attention switching algorithm in action, each consisting of the horizontal sequence. The first panel shows the superimposed test pattern. The second panel shows the attended input $\hat{\mathbf{x}}$ for the first round classification; because $\hat{\mathbf{x}}$ has analog values, we thresholded the values at 0.7 to facilitate viewing in the figure. The third panel shows the masking pattern for attention switching, generated by thresholding $\hat{\mathbf{x}}$ at 1. The fourth panel shows the residual input pattern for the second round classification. Figure 5 shows that attention switching was performed effectively, and the remaining input patterns for the second round classification are quite visible.

After performing selective attention and attention switching process, we compared the result to the one when selective attention and attention switching were not applied (We simply took the two outputs with the highest activity). The most active one was the first recognized class, and the second most active one was the second recognized one). Figure 5 and Table 2 summarize the recognition result.

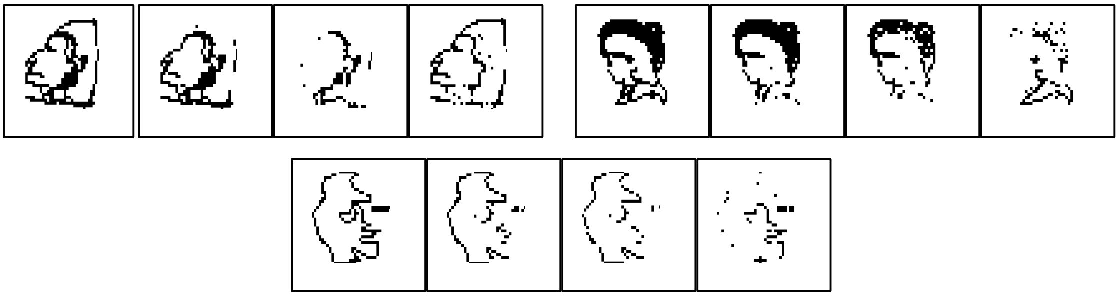


Figure 5. Examples of Selective Attention and Attention Switching. From the left to right for each test pattern, the test input, the attended pattern, the switching pattern, and the secondary input pattern, respectively.

Table 2. Recognition Result of the Test Patterns after tested with MLP without Selective Attention

			Bin 1	Bin 2	Bin 3	Bin 4	Bin 5
First Classification	Correct Recognition	Rate	100 %	83.05 %	80.79 %	57.35 %	40 %
		N.R.P/T.P.	76/76	147/177	122/151	56/136	24/60
	False Recognition	Rate	0 %	0 %	7.95 %	16.18 %	0 %
		N.R.P/T.P.	0/76	0/177	12/151	22/136	0/60
	Rejection	Rate	0 %	16.95 %	11.26 %	42.65 %	60 %
		N.R.P/T.P.	0/76	30/177	17/151	58/136	36/60
Second Classfication	Correct Recognition	Rate	100 %	97.74 %	86.09 %	79.08 %	70 %
		N.R.P/T.P.	76/76	173/177	130/151	107/136	42/60

N.R.P/T.P. means the number of recognized patterns / the total number of patterns

Bin1: Hamming Distance of 0 to 60, Bin2: Hamming Distance of 61 to 120, Bin3: Hamming Distance of 121 to 180, Bin4: Hamming Distance of 181 to 240, Bin5: Hamming Distance greater than 240

Table 3. Recognition Result using Selective Attention and Attention Switching with Varying a

a = 0

		Bin 1	Bin 2	Bin 3	Bin 4	Bin 5
First Classification	Recognition	100 %	100 %	100 %	96.32 %	73.33 %
	N.R.P/T.P.	76/76	177/177	151/151	131/136	44/60
Second Classification	Recognition	100 %	100 %	100 %	87.5 %	81.67 %
	N.R.P/T.P.	76/76	177/177	151/151	119/136	49/60

a =1/4

		Bin 1	Bin 2	Bin 3	Bin 4	Bin 5
First Classification	Recognition	100 %	100 %	100 %	100 %	81.67 %
	N.R.P/T.P.	76/76	177/177	151/151	136/136	49/60
Second Classification	Recognition	100 %	100 %	98.68 %	86.76 %	80 %
	N.R.P/T.P.	76/76	177/177	149/151	118/136	48/60

a = 1/2

		Bin 1	Bin 2	Bin 3	Bin 4	Bin 5
First Classification	Recognition	100 %	100 %	100 %	100 %	100 %
	N.R.P/T.P.	76/76	177/177	151/151	136/136	60/60
Second Classification	Recognition	100 %	100 %	98.68 %	86.76 %	80 %
	N.R.P/T.P.	76/76	177/177	149/151	118/136	48/60

a = 1

		Bin 1	Bin 2	Bin 3	Bin 4	Bin 5
First Classification	Recognition	100 %	97.74 %	98.68 %	100 %	100 %
	N.R.P/T.P.	76/76	173/177	147/151	136/136	60/60
Second Classification	Recognition	100 %	97.74 %	97.35 %	86.76 %	80 %
	N.R.P/T.P.	76/76	173/177	147/151	118/136	48/60

5. Conclusion

In this paper, we proposed and demonstrated selective attention and attention switching algorithm for recognition of noisy and superimposed patterns. The algorithms brought improved recognition rates and assured that this method was effective. This algorithm is simple and can be easily implemented to standard MLP, so the further application of this algorithm in a lot of different fields of visual and auditory pattern is expected.

References

[1] Lee, S.Y., Mozer, M.C. (1999). Robust Recognition of Noisy and Superimposed Patterns via Selective Attention, Neural Information Processing Systems 12, MIT Press.

[2] Lee, S.Y., Lee, K.W., Feng, J., Buxton, H. (2001). Perceptual Reversal over Time Course, International Conference on Neural Information Processing, vol. 1, pg26 - 32

[3] Cowan, N. (1997). *Attention and Memory: An Integrated Framework*, Oxford Univ. Press.

[4] Pashler, H.E. (1998). *The Psychology of Attention*, MIT Press.

[5] Parasuraman, R. (ed.) (1998). *The Attentive Brain*, MIT Press.

[6] Fukushima, K. (1987). Neural network model for selective attention in visual pattern recognition and associative recall, *Applied Optics*, 26:4985-4992.

[7] Rao, R.P.N. (1998). Correlates of attention in a model of dynamic visual recognition. In *Neural Information Processing Systems 10*, MIT Press.

[8] Broadbent, D.E. (1958). *Perception and Communication.* Pergamon Press.

[9] Treisman, A. (1960). Contextual cues in selective listening, *Quarterly Journal of Experimental Psychology*, 12:242-248.