

A computational model for character recognition based on multi-resolution channels and IAM

S. Y. BANG, C. S. PARK, S. K. KIM AND D. J. KIM
Dept. of Computer Science & Eng. & Brain Research Center
POSTECH, San 31 ,Hyoja-dong, Nam-gu, Pohang 790-784, KOREA

Abstract: A computational model to recognize Korean characters is presented. The model is based on IAM but improves the input processing part. We excluded the assumption that all the characters are of a standard template. Also we incorporated the multi-resolution channel theory so that we can improve the performance. We simulated the model and tried it to real data. The simulation results show that the model has a practical significance and the multi-resolution channels actually improve the performance.

Key-Words: Character recognition, IAM, Multi-resolution, Human visual processing, Neocognitron

1 Introduction

IAM (Interactive Activation Model) was proposed as a cognitive model for the visual recognition of English words[1]. Actually IAM was proposed to explain the Word Superiority Effect in English. In this model letters are recognized through the bottom-up process by the information from the input image and the top-down process by the information from the words. Although the model is simple, it can explain many aspects of letter and word recognition. It may be the only comprehensive model for visual letter and word recognition[2]. We earlier developed an IAM model for Korean character recognition based on the original IAM and succeeded in explaining the Character Superiority Effect which is found in the human perception of Korean characters[3]. This model motivated us to develop a computational model which is practical enough to be used by a real character recognition situation.

Although IAM is powerful and very persuasive as a cognitive model, it still needs many improvements and additional functions to become a more complete model for visual character recognition[2]. But here our interest is not to obtain a more complete cognitive model but to develop a computational model which is useful in a real world. When we review the IAM model from this aspect, the most problematic part is the input. In the original IAM the input letters are all given in a standardized shape i.e. template. The information from the input, therefore, always consists of a fixed number of data each of which tells whether a line segment exists or not in its prespecified location of the template. In our IAM model for Korean

characters the constraint was slightly relaxed so that each datum can be a real number instead of a binary. Further six networks are processed in parallel since there are 6 different character types of Korean characters (See the appendix for a brief description of Korean characters) and we don't know which type it is until we know the character itself. Note this is essentially different from and is an extension of the structure of the original IAM where four separate paths exist in order to process four letters of the input word in parallel since in the latter case we know that there are always 4 letters while in our model we don't know which type it is. In anyway the use of the common standardized template for all characters is unacceptable in a real situation even if the input characters are printed fonts.

On the other hand many models have been proposed which can recognize characters even if they are distorted to a certain degree. Neocognitron is one of them[4]. In this model micro features are gradually combined to larger features and eventually to patterns which we are looking for. Each step consists of two layers: a simple cell layer and a complex cell layer. Each cell of a simple cell layer detects a specific feature within its receptive field. Each cell of a complex cell layer absorbs a certain degree of distortion, i.e. size, rotation, location and deformation, by firing itself if any cell within its receptive field of the paired simple cell layer fires. It remains to be seen how accurately the model reflects the real mechanism of the human visual processing. But the important thing is it is based on some of the known principles of the human visual processing and yet works well for many practical problems.

Another idea which we are interested in is a theory that the human visual processing uses multi-resolution channels[5]. In this theory a low resolution channel has a faster signal processing speed and is used to process rough information about the input image while a higher resolution channel has a slower speed but is used to process more detailed information of the input image. There are many different proposals about the number of channels and the specifications of the resolutions. But no consensus has been reached. Anyway, the theory appears to be very persuasive and to have a practical merit.

As mentioned above we are interested in developing a computational model which is based on some of the principles of the human visual processing but is practical enough to be used for a real problem. Therefore our model will be based on our IAM and incorporate those ideas of Necognitron and multi-resolution channels.

In Section 2 we review the IAM model for Korean characters which our proposed model is based on. A description of the new and improved input processing is given in Section 3. We report a preliminary result of the experiment using the proposed model in Section 4. And Section 5 concludes.

2 IAM for Korean Characters

Here we will briefly review the IAM model which was based the original IAM but was modified to recognize Korean characters. (Please see [3] for a more detailed description.) The basic structure of the original IAM is given in (a) of Fig. 1. The feature level detects the visual stimulus from the input and sends the signals to the letter level. The letter level consists of 4 sets of alphabet nodes each of which corresponds to a letter in the given position of the word. And the word level consists of the word nodes which are supposed to be located in the memory and therefore have been acquired through study.

There are many nodes in the letter level and the word level and each node interacts with the other nodes of the same level and the nodes in the other level. There are two types of interactions between nodes: excitation and inhibition, as seen in Fig. 1. The letter level receives inputs from both the feature level and the word level. The input from the feature level corresponds to the bottom-up processing and the input from the word level corresponds to the top-down processing.

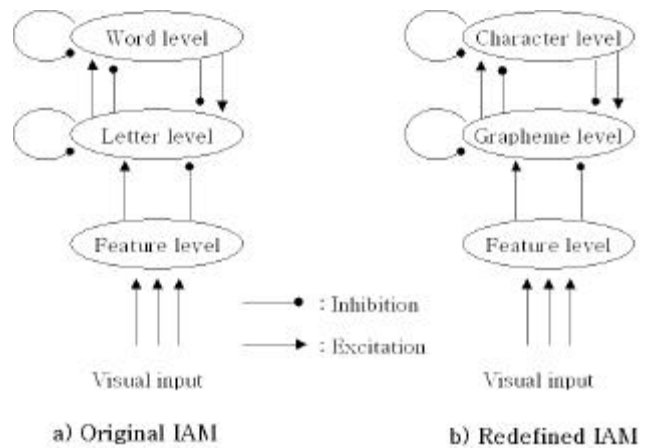


Fig. 1 Basic structure of IAM

The interactions between the levels are both of excitatory and inhibitory types while the interactions between the nodes of the same level only of inhibitory type. It is because only one alphabet of each set of the letter level and only one word of the word level is correct for the given input. Each node in a level tries to get fired for the given input stimulus and all of the nodes compete to win. Through these interactions among the nodes in the same level and between the different levels, eventually one node of the word level and one node of each alphabet set of the letter level will win. We say the model recognizes the input when we identify these winning nodes.

In order to adapt the original IAM for English words to the one for Korean characters, first of all, we have to redefine the meaning of the each level and the nodes in each level. The basic recognition unit of English is a letter. On the other hand that of Korean is a *jaso* (also referred to as a grapheme). The next larger unit of English is a word while that of Korean is a character. When we consider these units of the two languages the word level of the original model should become the character level in the our model and the letter level should become the *jaso* level. By reflecting these changes we obtain the IAM model for Korean characters shown in (b) of Fig. 1.

One more difference exists between the letter level and the *jaso* level. In case of English only four letter words were considered. This means there are four separate networks between the nodes in the feature level and those of the letter level each of which corresponds to one of the four letters in a word. A situation is a slightly more complex in case of Korean characters. There are 6 different compositional structures of Korean characters (which we call character types) as described in the appendix. Since

we do not know the character type until we know the character of the input, we have not only to prepare the six separate input processing structures but also to process all of them at the same time. However, the overall operation of the model remains essentially the same.

3 Extension of Input Processing

3.1 Overall direction

In the original IAM the input stimulus can only be a binary. In our previous model we demonstrated that the input stimulus could be a real value. But the use of a standard template for all input images adopted by the both models is a serious restriction, and even a problem from the point of developing a practical recognition system, as pointed out earlier. In the proposed model, therefore, no assumption is made about the locations and the shapes of strokes of a character except for the assumption that the input has been normalized with respect to the size and the rotation. In the previous models the set of features are fixed for all letters and the detection of these features is done manually. On the other hand in the proposed model we are supposed to define features and extract them automatically. In order to do so we decided to use neural networks.

Furthermore we decided to use features of different resolutions in order to perform the recognition task more effectively. Features of a low resolution are used to extract the overall shape of the input and can be transmitted faster. We send the information to a higher level of the recognition process, i.e. the character level, and use it to squeeze the set of candidate characters. Features of a high resolution are used to extract the detailed information about the shapes of the input so that *jasos* are correctly identified. In either case features are extracted from the input. In terms of IAM we can say that we use a low resolution channel to help the top-down processing while a high resolution channel to help the bottom-up processing.

In our proposed model we use 3 channels: a low resolution, a middle resolution and a high resolution channels. Below we describe the structure and the function of each channel.

3.2 Low resolution channel

We can think of many different meanings of and assign many different roles for a low resolution channel. In our model we use the low resolution

channel to extract the overall shape of the input so that we can reduce the set of the candidate characters. In order to do so we first cluster the characters using the features of a low resolution extracted from the training data images. Then, given an input, we extract its features and determine which cluster it belongs to. The characters which belong to the same cluster as the input's receive the excitatory signals.

In order to obtain the features for the low resolution channel the input image is divided into 5*5 windows without overlap. This means that we extract a feature vector of 36 dimensions if the input is 30*30. For all characters in the character level we collect the training data and cluster them by using a SOM network.

3.3 Middle resolution channel

We use the middle resolution channel to help determine the character type of the input. As mentioned earlier the character type can be determined only when the input character is identified. Practically there is no way to tell the character type of the input without knowing what character the input is. Therefore only we can is to guess the character type by using the rough shape of the character.

In order to extract the features of the channel we cover the input image by small windows allowing overlap. And we use a MLP network for this channel. The input layer is for these features and the output layer consists of the six character types. First we train the network using training data. Then, given an input, we extract the features and feed the features to the network. We send the excitatory signal of the strength which is proportional to the output of each output node of the network to those character nodes of the character type of the output node.

3.4 High resolution channel

The high resolution channel actually consists of 6 separate channels each of which corresponds to a character type. And they process in parallel. Each of the six channels processes each input by assuming that the input is of a its type. As described in the appendix, each character type has a particular composition of *jasos*. Character types are either C-V or C-V-C (C denotes a consonant *jaso* while V denotes a vowel *jaso*) although their structural compositions may vary. Therefore each channel actually consists of either two networks or three networks each of which takes care of a specific *jaso* type. This means we have 15 different networks in total operating separately.

For each network mentioned above we use a Neocognitron-like feed forward network. The structure and the operation of these networks are all the same although the number of inputs and the number of outputs may be different. We can train each network separately since each network operates independently. Each *jaso* recognition network has two hidden layers. Each hidden layer consists of a pair of layers: one simple cell layer and one complex cell layer.

A node of the simple cell layer works like a simple cell of the human visual area while a node of the complex cell layer works like a complex cell. In other words a node of a simple cell layer detects a specific predetermined feature within a given receptive field of the input. And a node of a complex cell layer absorbs distortions in the input by firing itself if any node within the receptive field of the previous layer fires. Therefore a dimensional shrink arises when we go from a simple cell layer to a complex cell layer.

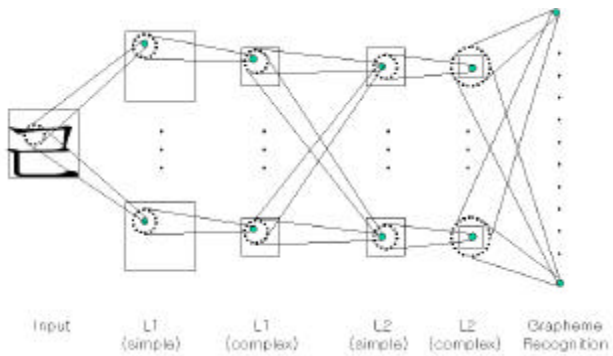


Fig. 2 Structure of a *jaso* recognition network

A typical structure of a *jaso* recognition network is shown in Fig. 2. The size of the input layer is decided by considering the size of the subimage which can cover all possible *jasos* in that location. Note that the exact location of a *jaso* varies from character to character even when it is the same *jaso*. Therefore we usually use a region which is larger than the exact sizes of the *jasos*. Each layer of the hidden layers consists of many plains. Each plain of the first simple cell layer detects a particular micro feature in the input. These 12 different micro features are shown in Fig. 3. Each micro feature corresponds to a short line segment of a particular direction. Plains of the second simple cell layer detect larger features which can be constructed by combining the previous micro features. The nodes of the output layer correspond to a set of

jasos which possibly appear in the location of the input image for this particular character type.

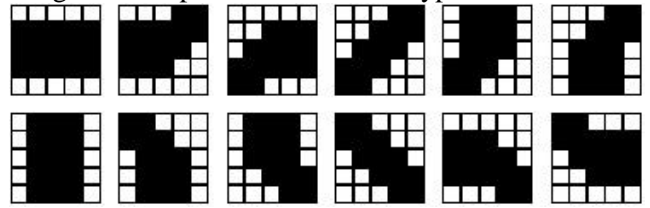


Fig. 3 Micro feature of layer 1

3.4 Operation of the model

The entire model is shown in Fig. 4. The basic mechanism to integrate the output signals from these three different channels is still IAM. The system uses the outputs of the high resolution channel as the pure bottom-up input. The feature level of the original IAM which detects features from the visual input is now replaced by the high resolution channel. The outputs from the low resolution and the middle resolution channels goes directly to the character level to affect the top-down processing. The question is how to control and balance the strengths of the effects from these three different channels. It seems reasonable to give less weights to the outputs from the low and the middle resolution channels since these outputs do not give definite and detail information for the recognition but rather give a sort of guess and only supportive information.

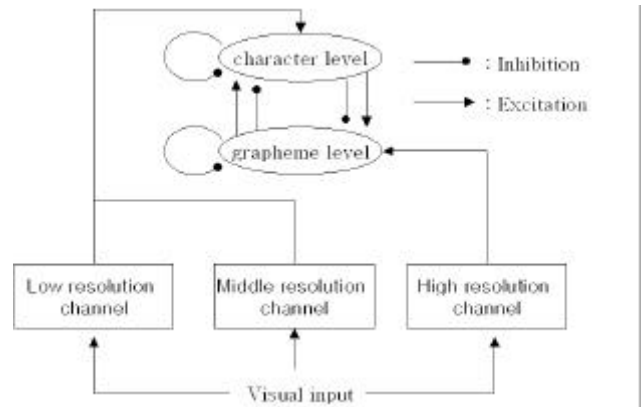


Fig. 4 Overall structure of the model

4 Simulation

We constructed our proposed model and ran simulations by using real data. This is a preliminary report of the simulations. It is preliminary since the simulations we performed only verified the feasibility of such a model. In order to analyze the real potential and point out the pros and cons of the proposed model

we need more simulations using a broader range of data.

	Nodes	Connections
High channel	25,398	1,103,088
Middle channel	106	10,701
Low channel	100	10,000
Grapheme level	158	1,478
Character level	557	154,846

Table 1 Number of nodes and connections

For this simulation we used the most frequently used 600 characters out of the standard set of 2350 Korean characters. This set represents about 95% of accumulated uses of Korean characters. Further we excluded those of character types 3 and 6 from these 600. That left 557 characters for the simulation. This reduction of the number of characters was necessary to make the simulation possible within a reasonable time frame. See Table 1 for the size of the current system. Since this initial simulation also gave us a chance to adjust various parameters in order to run the system successfully, we decided to use printed characters for the simulation.

The size of an input image is 30*30 binary pixels. The low resolution channel uses as its input 6*6 real number pixels which are derived from the original input by using 5*5 windows. The middle resolution channel uses as its input 10*10 real number pixels which are derived from the original input by using overlapped 5*5 windows. As mentioned above the high resolution channel uses the original input as it is. However each network of a high resolution channel uses a different subregion of the input image which covers a particular *jaso* location..

In this simulation we used printed characters of just one font. In order to train the networks of these three channels we prepared 5 different sets of 557 character images by adding 5% random noises to the clean images. Then we prepared the other 5 sets of 557 character images in the same way for the testing use. The results of the simulation are given in Table 2. As seen in the table, the outputs from the low resolution channel and the middle resolution channel helps improve the performance of the system.

Channel	Training data	Test data
Only high	97.8 %	93.4 %
High + Low	98.6 %	95.1 %
High + Middle	98.9 %	98.0 %
All	99.2 %	98.3 %

Table 2 Result of simulations

5 Conclusion

The proposed model is an attempt to develop a practical system based on a solid cognitive model. In order to do so, we started from the IAM model for Korean characters. Then we improved the input part so that we can accept any character image, not a standard image. Further in order to take advantage of multi-resolution channel theory of human visual processing we incorporated three different resolution channels: low, middle and high resolution channels. These channels are realized by various neural network structures.

In order to verify the possibility and the performance of the proposed model we ran simulations by using real data of Korean characters. Through the simulation we confirmed that the system worked as expected. We confirmed that the low and the middle resolution channels help improve the overall performance of the system and are complimentary to the role of the high resolution channel.

Acknowledgment

This research was supported by the Brain Korea 21 Project by Ministry of Education and Brain Science and Engineering Research Program sponsored by Korean Ministry of Science and Technology.

References:

- [1] McClelland J. L. & Rumelhart D. E., An Interactive Activation Model of Context Effects in Letter Perception: Part 1. An Account of Basic Findings, *Psychological Review*, Vol. 88, No. 5, 1981, pp. 375-407.
- [2] Glyn W. Humphreys & Vicki Bruce, *Visual Cognition: Computational, Experimental, and Neuropsychological Perspectives*, Lawrence Erlbaum Associates Ltd., Publishers, 1989.
- [3] Park C. S. & Bang S. Y., Modeling Character Superiority Effect in Korean Characters by Using IAM, *Proceedings of Biologically Motivated Computer Vision*, Vol. 1, 2000, pp. 316-325.
- [4] Kuniyiko Fukushima, Sei Miyake & Takayuki Ito, Neocognitron: A Neural Network Model for a Mechanism of Visual Pattern Recognition, *IEEE Transaction on systems, man and cybernetics*, Vol. 13, No. 5, 1983, pp. 826-834.
- [5] Toshio Inui, A Model of Human Visual Memory: Data Compression with Multi-resolution,

Scandinavian Conference on Image Analysis, Vol 6, 1989, pp. 325-332.

- [6] Kathrayn T. Spoehr & Stephen W. Lehmkuhle, *Visual Information Processing*, W. H. Freeman and Company, 1982.
- [7] Lee S. H., Kim C. H., Hong Ma & Yuan Y. Tang, Mutiresolution Recognition of Unconstrained Handwritten Numerals with Wavelet Transform and Multilayer Cluster Neural Network, *Pattern Recognition*, Vol. 29, No. 12, 1996, pp. 1953-1961.
- [8] Kim J. K. & Kim J. O., Grapheme cognition in Korean character context, *Ph. D. Thesis(in Korean)*, Seoul National University, 1994.

Appendix : A review on the structure of a Korean character

A Korean character consists of two or three components: the first grapheme(*jaso*), the middle grapheme and the last grapheme which is optional. A Korean character is of a two dimensional composition of these three graphemes. The six structural types of Korean characters with their examples are given in Table 3. The first grapheme is a consonant and located on the left, the top or the left upper corner of a character. The middle grapheme is a vowel. There are three types of the vowel graphemes as seen in the Table 3. It is either a horizontal one, a vertical one or the combination of them. The last grapheme is a consonant and always located in the bottom of a character if it exists.

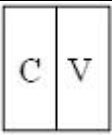
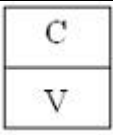
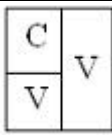
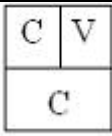
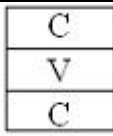
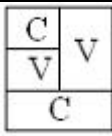
Type	1	2	3
Structure			
Example			
Type	4	5	6
Structure			
Example			

Table 3 Six types of Korean characters