Journal of Korea Multimedia Society Vol. 12, No. 12, DECEMBER 2009(pp. 1718-1727)

A Recognition System for Multi-Form Korean Characters Based on Hierarchical Temporal Memory

Nan Haibao†, Sun-Gap Bae††, Jong-Min Bae†††, Hyun-Syug Kang††††

ABSTRACT

Traditional character recognition systems usually aim at characters with simple variation. With the development of multimedia technology, printed characters may appear more diversely. Existing recognition technologies can't deal with Hangul recognition effectively in diverse environments. This paper presents a recognition system for multi-form Korean characters called RSMFK, which is based on the model of Hierarchical Temporal Memory (HTM). Our system can effectively recognize the printed Korean characters of different fonts, scales, rotation, noise and background. HTM is a model which simulates the neocortex of human brain to recognize and memorize intelligently. Experimental results show that RSMFK performs a good recognition rate of 97.8% on average, which is proved to be obviously improved over the conventional methods.

Key words: Korean characters recognition, neocortex, HTM

1. INTRODUCTION

There have been many methods used in the character recognition domain, such as artificial neural network (ANN) [1], sub-pattern separation algorithm [2] and so on [3-7]. These methods usually aim at black characters on white background and with simple variation. Recently, with the multimedia technology developing, characters appear in different environment by different forms. However, conventional methods can't recognize such characters very well. In this paper, we present a new technology to recognize the characters with variation on font, scale, rotation, noise, and background.

Hangul takes on some unique and scientific characteristics in the composition of structure. A Korean character is made up of three graphemes called the first sound, the middle sound, and the optional last sound. Human can recognize a Korean character according to the three sounds of a character. Enlightened by this idea, we can make a system recognize Korean character by the coding of the three radicals. However, the way to determine and memorize the coding is a critical step.

Recently, the theory of Hierarchical Temporal Memory (HTM) [8] proposed by Jeff Hawkins is one of the researches about modeling human intelligence. HTM simulates the neocortex of human brain to recognize and memorize intelligently. The main advantage of HTM is that it can capture, memorize, and model the structure of the world efficiently. It offers the promise of building machines that approach or exceed human level performance for many cognitive tasks. Therefore,

※ Corresponding Author : Hyun-Syug Kang, Address : (660-701) 900 Gajwa-dong, Jinju, South Korea, TEL : +82-55-751-5998, FAX : +82-55-762-1944, E-mail : hskang@gnu.ac.kr

Receipt date : Feb. 27, 2009, Revision date : July 2, 2009
Approval date : Aug. 28, 2009

† Department of Computer Science, GyeongSang National University
(E-mail : nanhaibao666@hotmail.com)
†† Department of Computer Science, GyeongSang National University
(E-mail : bsgap@hanmail.net)
††† Department of Computer Science, GyeongSang National University
(E-mail : jmbae@gnu.ac.kr)
†††† Department of Computer Science, GyeongSang National University

HTM is thought of as a memory system and it can "learn" how to solve problems.

Based on the HTM technology and the characteristics of Hangul, we develop RSMFK (Recognition System for Multi-Form Korean characters) to recognize characters of different forms. RSMFK is composed of the three subsystems: the segmentation subsystem, the HTM network subsystem and the combination subsystem. We utilize the structural characteristics of Hangul, firstly separate the three radicals of each character using the software - Image Cutter V2.53 in the segmentation subsystem, then identify the separated radicals independently using the corresponding HTM network in the HTM network subsystem, finally combine the three radicals and obtain the recognized character in the combination subsystem. Before the recognition process, the three radical HTM networks should learn the corresponding radicals using the training data. Experimental results show that the recognition is accurate and reasonable when the input data are in different scales, shapes, and background.

This paper consists of 6 chapters as follows: Chapter 2 describes the HTM theory and characteristics of Hangul. Chapter 3 introduces the design of RSMFK. In chapter 4, we show and analyze the experiment. Chapter 5 shows the comparison with other researches. Finally, we discuss the conclusion and the future works in chapter 6.

2. The HTM Theory and Characteristics of Hangul

In this chapter, we will describe the HTM theory and characteristics of Hangul briefly.

2.1 HTM Theory

HTM (Hierarchical Temporal Memory) is a technology that replicates the structural and algorithmic properties of the neocortex [8]. It is a collection of linked nodes, organized in a tree-shaped hierarchy, where each node implements a common learning and memory function. HTM network utilizes Zeta1 nodes in addition to one or more sensors and effectors [9]. The following Fig. 1 shows the structure of HTM network.

A Zeta1 node contains two modules: the spatial pooler and the temporal pooler. The spatial pooler is used to learn a mapping from a potentially infinite number of input patterns to a finite number of quantization centers. The output of spatial pooler is in terms of its quantization centers. The temporal pooler is used to learn temporal groups - groups of quantization centers - according to the temporal proximity of occurrence of the quantization centers of spatial pooler. The output of the temporal pooler is in terms of the temporal groups that it has learned. The input to Zeta1 node is connected to the input of the spatial pooler. The output of the spatial pooler is the input to temporal pooler. The output of the temporal pooler is the output of the node. The node shown in this figure has an input pattern size 6. The spatial pooler within this node has 5 quantization centers marked c1 to c5. The output of the spatial pooler is a vector of length 5. The temporal pooler within this node has 2 groups marked g1 and g2. The output of temporal pooler, and hence of the node, is a vector of size 2.

Through operation of Zeta1 node, HTM network can store information throughout the hierarchy in such a way that models the world. All objects in the world have structures that are hierarchical in both space and time. HTM network can efficiently capture and model the structure of the world.

Zeta1Top Node is usually used for supervised

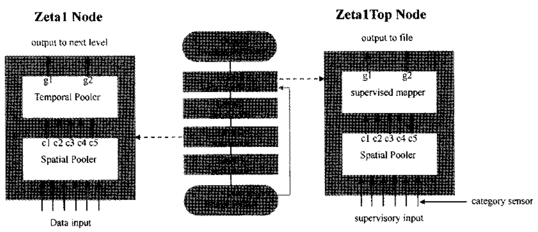


Fig. 1. The structure of an HTM network

learning at the top of a hierarchy. A Zeta1Node can be used in place of a Zeta1Top Node for a fully unsupervised network. As shown in Fig. 1, Zeta1Top Node accepts a second supervisory input in addition to the input signal received by all Zeta1 nodes. Zeta1Top Node uses the supervised mapper in place of the temporal pooler. It outputs a distribution over a set of temporal groups formed via unsupervised learning. Zeta1Top Node outputs a distribution over categories explicitly specified via supervised learning. In short, Zeta1Top Node is a simple classifier used to implement supervised learning at the top of an HTM network.

2.2 Characteristics of Hangul

Hangul is characterized by its large number of characters and its two dimensional composition of three graphemes called the first sound, the middle sound, and the optional last sound. Fig. 2 shows the structure of a Hangul character.

The number of the graphemes which belong to the first sound, the middle sound and the last sound are 19, 21 and 27 respectively. However, the two dimensional composition of these graphemes makes a total of 11,172 characters [10]. It is not easy to recognize such a large number of different characters. But the number of graphemes for each sound is far less than that of characters. Therefore, the recognition task becomes much easer if we recognize a character by identifying the separated graphemes using three HTM networks independently, which is the basic idea of our recognition system.

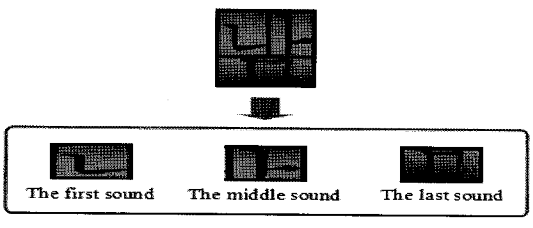


Fig. 2. The structure of a Hangul character

3. RECOGNITION SYSTEM FOR MULTI-FORM KOREAN CHARACTERS

This chapter describes the design of RSMFK (Recognition System for Multi-Form Korean characters). We firstly give an overview of RSMFK, followed by the illustration of the three subsystems of RSMFK.

3.1 Overview of RSMFK

RSMFK is composed of three subcomponents: the segmentation subsystem, the HTM network subsystem and the combination subsystem. Each subsystem has a particular function. Fig. 3 shows the overall structure of RSMFK. The recognition process showed in Fig. 3 is as follows: when a user inputs a Korean character from the user interface, the system will segment the character into three radicals using the segmentation subsystem and classify the radicals into the first sound, the middle sound and the optional last sound. Then these three radicals are separately given as input to the corresponding trained HTM network subsystem. The HTM network subsystem contains three networks with identical architecture. In this subsystem, each of the three HTM networks will infer corresponding radical using the HTM algorithm like human

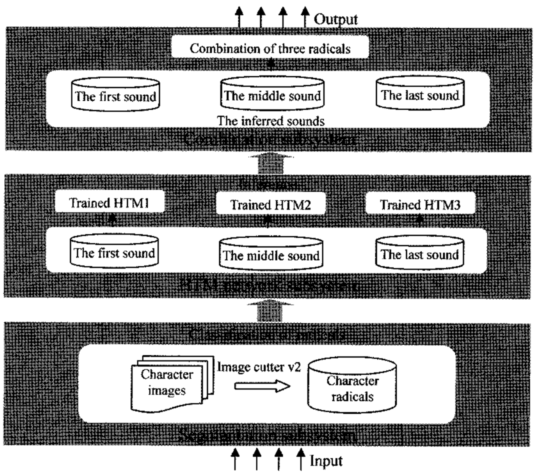


Fig. 3. The structure of RSMFK

brain recognizing an object. After the inference process, the system can obtain the inferred radicals and then combine the radicals using the combination subsystem. Finally, the system outputs the recognized character.

3.2 The Segmentation Subsystem

We can produce Korean character radicals from printed characters using the segmentation subsystem. For separating the three radicals, we utilize the segmentation software - Image Cutter V2.53 [10]. Fig. 4 shows the segmentation process of a character.

The following Fig. 5 shows the flow diagram in the segmentation subsystem. All the training samples will be initially processed into normative images (128 pixels by 128 pixels). Then every character in the data sets will be segmented by Image Cutter V2.53. Each character is usually segmented into three radicals — the first sound, the middle sound and the optional last sound. These radicals are then extracted and post-processed into normative radical images (32 pixels by 32 pixels). Finally, the radicals are classified into three groups. These separated normative radical images will be used as the input sensor vector for the HTM network subsystem.

When the data are clean and regular, the separated radical images are ideal. Since the Image Cutter segments the image of the character only using the horizontal line and vertical line, the segmented results are not very well when the data are seriously distorted, which will exert an influence on the recognition accuracy.

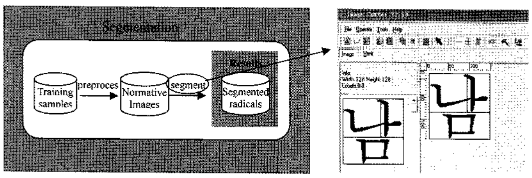


Fig. 4. The segmentation of a character

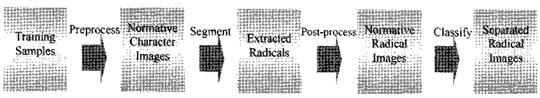


Fig. 5. Flow diagram in the segmentation subsystem

3.3 The HTM network subsystem

The HTM network subsystem is composed of three identical HTM networks which are independently used to determine the first sounds, the middle sounds, and the optional last sounds. We utilize NuPIC [11] to develop the HTM network subsystem.

Fig. 6 is the architecture of the first sound HTM network. In this network, a VectorFileSensor sends input data (normative format of radical) from the segmentation subsystem to Zeta1 node at the bottom level (level 1) which has 16 Zeta1 nodes. In this level, the NuPIC will exhaustively sweep the image using the imageSensor of 4*4 pixels, these HTM nodes will learn the sound in detail. After the inference in the bottom node, the sound is sent to the middle level (level 2) which has 4 Zeta1Nodes. In this level, HTM will further learn the data using the imageSensor of 8*8 pixels and then send it to a single Zeta1TopNnode at the top level. During further inference using the imageSensor of 32*32 pixels in the top node, the sound will be sent to a VectorFileEffector and then to the combination subsystem. The right part of Fig. 6 shows the training process of the first sound HTM network in the level 1 (down) and level 2 (top) using NuPIC.

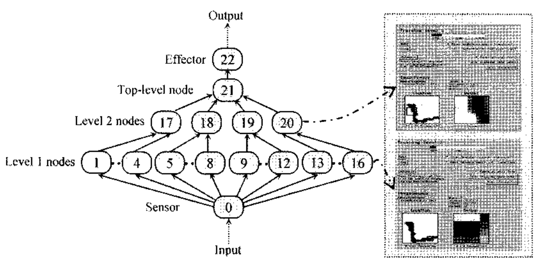


Fig. 6. The architecture of the first sound HTM network

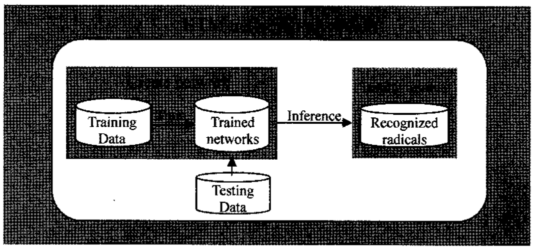


Fig. 7. Flow diagram in HTM network subsystem

Fig. 7 shows the flow diagram of the process in the HTM network subsystem. We initially build three HTM networks, and then train the networks using the training data, after that we can obtain the trained networks. Using the trained HTM networks, we can infer different kinds of similar testing data. The testing results can validate the performance of the trained HTM network.

3.4 The Combination subsystem

After the inference in the HTM network subsystem, the three recognized radicals will be transported to the combination subsystem. In this subsystem, three radicals will be combined into a character by the combination software. As each radical has a unique binary coding, the combination of the three radicals also has a unique binary coding which was one-to-one correspondence to a Korean character. The example of the combining process of a Korean character "남" is as shown in Fig. 8, the binary coding of the sound "ㄴ" is 00101, the coding of sound "ㅏ" is 00011, the coding of sound "ㅁ" is 01000, so the combination of the three coding is 011000010101000 which only represents

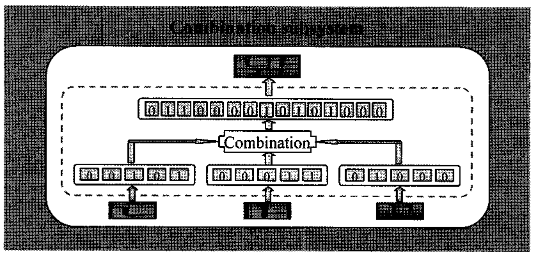


Fig. 8. The combining process of a Korean character

the character "남" in Hangul.

4. EXPERIMENT

RSMFK is in the foundation of series of software and development toolkits. NuPIC 1.6.1 [11] and Vitamin D toolkit [12] are used to implement our system. Its implementation language is Python 2.5. In this chapter, we show the experimental data, optimize several main parameters of HTM algorithm and analyze the effect of these parameters to the experimental results.

4.1 Collection of the Experimental Data

We collect 35 different fonts of Korean characters as the training data to train HTM networks. The data contain all the 19 first sounds, 21 middle sounds and 27 optional last sounds. We use large kinds of characters of variable scales, various rotation, with noises or in different background as the testing data. Fig. 9 enumerates some samples of the training data and testing data.

4.2 Experiment and Optimization of the Parameter Values

We investigate the effect of the parameters maxDistance, sigma of HTM engine on the accuracy in each level. The parameter maxDistance controls the behavior of the spatial pooling algorithm. The parameter sigma controls the behavior of the spatial pooling algorithm in "Gaussian" mode only. We vary the values of maxDistance and set sigma to the square root of maxDistance. The accuracy of level 1 with the variation of the two parameters is shown in table 1. The values of the two parameters in level 2 and level 3 are set to 'defaulted'.

From the above table, we can observe that when maxDistance is set to 6 and sigma is set to 2.45, the system can obtain the highest accuracy. Therefore, the value 6 and 2.45 is separately

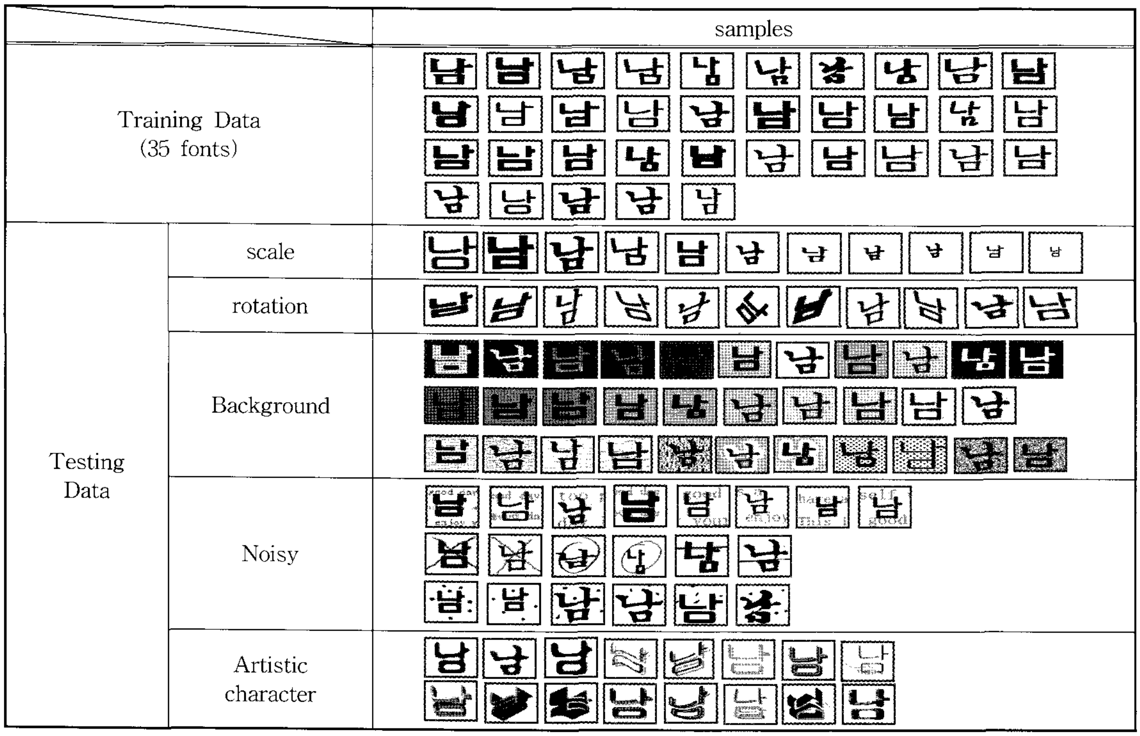


Fig. 9. Example of the sample data

Table 1. The accuracy with different values of two parameters

maxDistance	0.0	1	3	6	9
Sigma	0.408248	1.00	1.73	2.45	3.00
Accuracy (%)	94.28	90.22	95.03	96.10	94.49

regarded as the optimal value of maxDistance and sigma in level 1.

When sigma and maxDistance are set to optimal value in level 1 and defaulted in level 3, we make the same experiment and obtain the optimal value of the two parameters in level 2 is respectively 3 and 1.73.

Likewise, we research the effect of the sigma and maxDistance in level 3 when they are set to optimal values in level 1 and level 2. When the two parameters are separately set to 16 and 4 in level 3, the system can get the best performance.

To optimize other parameters of different levels, we also carry out different experiments and reach a conclusion that the parameters spatialPooler Algorithm and temporalPoolerAlgorithm have an important influence on the recognition results.

The temporalPoolerAlgorithm parameter controls how the distribution over groups is calculated from the distribution over coincidences. The two options for the parameter are "maxProp", "sumProp". If the temporalPoolerAlgorithm is set to "maxProp", the pooler iterates over groups. For each group, it finds the coincidence in that group with the highest value in the belief vector received from the spatial pooler. That maximal belief in the group becomes the value for the group itself, and it is entered into the output vector. During inference, the temporal pooler uses its list of groups to convert incoming belief vectors to distributions over groups. The matrix of weights is also used to calculate output when temporalPoolerAlgorithm is set to "sumProp".

The general behavior of the spatial pooler could be implemented in a variety of ways; in fact, Zetal nodes have three spatial pooling algorithms:

Gaussian, Dot, and Product. The latter two forms require that maxDistance also be specified.

The optimal spatial pooling algorithm depends primarily on the nature of the input vectors. Gaussian spatial pooling should be used for nodes that receive input directly from a sensor. Dot or Product pooling should be used for middle or top nodes that receive input from other ZetalNodes.

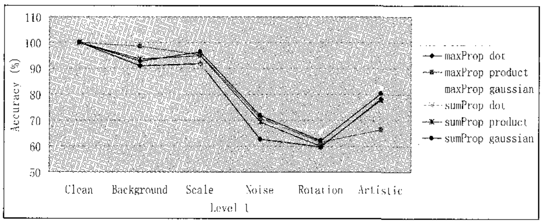
The Gaussian spatial pooling algorithm uses a Euclidean distance metric to compare the input vectors presented during training. The squared Euclidean distance is between an input vector x and an existing coincidence w. This decision is controlled by the parameter maxDistance.

Whereas coincidence matrices generated by the Gaussian algorithm store raw input vectors, those generated using the Dot and Product algorithms store input vectors updated using the winner-take-all approach. Each of these vectors has N *children elements* set to 1— one such element for each child node — and all other elements set to 0.

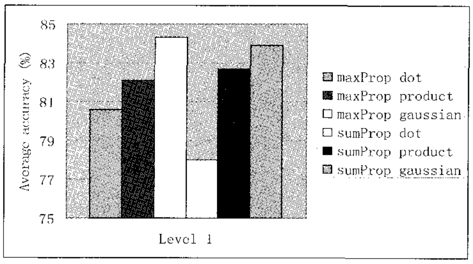
When the Dot inference algorithm receives a concatenated input vector from its child nodes, it generates a belief value for each of its stored coincidences using a dot product operation.

The Product inference algorithm is very similar to the Dot algorithm. We saw that when the Dot algorithm computes a belief value, it sums the support from each of its children. By contrast, the Product algorithm multiplies the support.

When the parameter TemporalPoolerAlgorithm is set to maxProp or sumProp, and the parameter SpatialPoolerAlgotithm is set to dot, product or Gaussian, there are 6 combinations about the two parameters. They are separately maxProp and dot, maxProp and product, maxProp and Gaussian, sumProp and dot, sumProp and product, sumProp and Gaussian. Aim at each combination, we carry out different experiments using different kinds of experimental data. Fig. 10 (a) shows the recognition results of level 1 when the two parameters are set to different options and the data are in different forms. Fig. 10 (b) shows the average accuracy of all the experimental data.



(a) Accuracy with different data and parameters



(b) Average accuracy

Fig. 10. The recognition rate of level 1

From Fig. 10, we can see that, maxProp and gaussion are regarded as the optimal values to the two parameters temporalPoolerAlgorithm and spatialPoolerAlgorithm in level 1. In the same way, we carry out the same experiment in level 2 and 3. According to the experiment, sumProp and product (or maxProp and gaussian) are regarded as the optimal values to the two parameters in level 2. The optimal values for the two parameters in level 3 are sumProp and product.

The above figure also shows that, we can achieve high recognition rate of 100% using clean data. When the data are in different scales or with different background, we can also obtain an ideal recognition result. The average recognition rate for these three cases is 97.8%. When the data are artistic, rotated, or with noise, the results of our experiment are 65%, 75%, and 76% respectively.

5. COMPARISON WITH OTHER RESEARCHES

There are many methods to recognize Korean characters. In this section, we compare the method based on HTM with other methods.

Pyeoung Kee Jim proposed a printed Hangul recognition algorithm on multi-transputer system in paper [2], where he developed a sub-pattern separation algorithm which is based on the vowel. After sub-pattern separation, the system extracts stroke codes for each sub-pattern and recognizes them. He tested the validity of the recognition system for 7,000 printed characters from an ordinary text and showed good results in both speed and recognition rates.

J.koh presents a investigation on recognizing printed Korean characters using neural networks in paper [5]. The approach is based on a variant of the back-propagation algorithm. The results indicate that by transforming the character data into Hough space, the author can achieve a good recognition. The results indicate that the approach can recognize about 81% of the training samples and 73% of the tested samples.

In paper [6], K. Ham introduces a table called Image Abstraction Table, which was constructed by scanning a binary image horizontally as a preprocessing method for the easy stroke extraction. From the table, author's stroke extraction algorithm recognizes and deletes the vowel of a Korean character. As the result of this pre-recognition of vowels, the algorithm segments a Korean character automatically and one or two consonants of a Korean character remains in a character image of the table. Using the stroke extraction algorithm, the consonants in a Korean character are easily recognized and the average recognition rate is 90%.

The above three methods have get certain application in 90's, however, the object of the above three methods is only one type of font (Ming-style printing type).

J. Kim proposes an automatic multilingual character recognition system based on the hierarchical classification neural networks for printed Korean character of 3 fonts [7]. The recognition system is designed by hierarchical nets composed of one type classification net and seven type specific alphabet classification nets. J. Kim proposes some reliability factor decision rules in the type-classification net and alphabet classification nets to recover some classification errors. The testing results for 3000 Korean characters show that the proposed system generated good solution for multilingual character recognition with about 97% accuracy.

The method based on HTM in our system uses the data of different fonts, scales and background as the experimental object. The average recognition rate for these three kinds of data is 97.8%, which is higher than that of other methods. We also make experiments using the data of rotation, with noise or artistic, and the results (about 65%, 75% or 76%) are still acceptable. Table 2 shows different comparison items of different recognition methods.

Table 2. The comparison of different recognition methods

Methods / Items	HTM	Sub-pattern separation algorithm [2]	A variant of back-propagation algorithm [5]	Stroke extraction algorithm [6]	Hierarchical classification neural networks [7]
Number of data	12500	7000	117	700	3000
Number of font	35	1	1	1	3
Scale	5	1	1	1	3
Noisy	Yes	No	No	No	No
Background	Yes	No	No	No	Yes
Recognition Rate	97.8%	95.0%	73%	90%	97%

6. CONCLUSION AND DISCUSSION

In this paper, we develop a Recognition System for Multi-Form Korean characters (RSMFK) using the HTM model. Through the experiment using lots of training and testing data, the system shows a good performance on the multi-form Korean characters such as different backgrounds, scales, artistic recognition, the results are 97.2%, 96.9%, and 76% respectively. Since we use the Image Cutter v2.53 software to segment the character, the recognition accuracy is not very well when the data are seriously distorted and noisy, but the accuracy are also acceptable, they are separately 65%,75%. If we can have some improvement in the segmentation subsystem, the accuracy will be optimized and ideal. In the future, we will continue devoting to the research of segmentation algorithm, execution efficiency of the system using HTM.

REFERENCES

[1] S. Araokar, "Visual Character Recognition using Artificial Neural Networks," *Computer Science - Neural and Evolutionary Computing*, 2005.

[2] K.J.Pyeoung and H.B.Jeen, "Hangul Recognition on Multi-Transputer System [J]," *IEEE Region 10 Conference on Communication, Control and Power Engineering*, Vol.2, pp. 1102-1105, 1993.

[3] K. Kang and J. Kim, "Handwritten Hangul Character Recognition with Hierarchical Stochastic Character Representation [A]," *IEEE Computer Society*, pp. 212-216 Vol.1, 2003.

[4] P. Lee and H. Jang, "A study on printed Hangul recognition with dynamic Jaso segmentation and neural network [J]," *Journal of Korean Institute of Communication*, pp. 2133-2145, 1994(19).11.

[5] J. koh, G. S. Moon, K. G. Mehrota, C. K. Mohan and S. Ranka, "Korean Character Recognition Using Neural Networks [J]," *3rd Symposium on the Frontiers of Massively Parallel Computation*, pp. 162-165, 1990.

[6] K. Ham, "A Study for Printed Korean Character Segmentation and Recognition by Vowel Pre Recognition," *Paper of Conference of Dongguk University*, Vol.10, pp. 555-567, 1991.

[7] J. Kim, K. Kim and S. Chien, "Korean and English Character Recognition System Using Hierarchical classification Neural Network," *IEEE International Conference on Digital Object Identifier*, pp. 759-764 Vol.1, 1995.

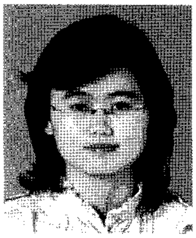
[8] J. Hawkins and S. Blakeslee, *On Intelligence*, New York, Henry Holt, 2004. http://www.onintelligence.com

[9] D. George and B. Jaros. "The HTM Learning Algorithms," Numenta Inc., 2007. http://www.numenta.com/for-developers/education/Numenta_HTM_Learning_Algos.pdf

[10] R.Xu and C.P.Liu, "Grapheme Segmentation and Recognition in Machine Printed Hangul Characters," *Journal of Chinese Information Processing*, pp. 66-71, 2006(26).

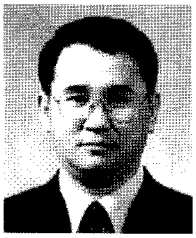
[11] Numenta Inc., "NuPIC Getting Started," http://www.numenta.com/

[12] Vitamin D Inc., "Vitamin D Toolkit Reference Guide," http://www.numenta.com/



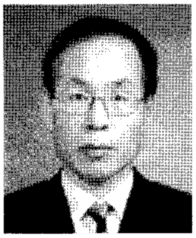
Nan Haibao

Nan Haibao received the B.Sc. degree in computer science from BeiHua National University, China, in 2007 and the M.Sc. degree in computer science from Gyeongsang National University, Korea, in 2009. Her main research interest includes Neural Network, and Artificial Intelligence.



Sun-Gap Bae

Sun-Gap Bae received the B.Sc. degree in physics from Kyungpook National University, Korea, in 1987 and the M.Sc. degree in computer science from Gyeongsang National University, Korea, in 2004. He is currently pursuing a Ph.D. in Gyeongsang National University, Korea. His main research interest includes Embedded DB, Neural Network, and Artificial Intelligence.



Jong-Min Bae

Jong-Min Bae received the B.Sc. degrees in Mathematics Education from Seoul National University, Korea, in 1980, and the M.S. degree, Ph.D. degree in Computer Science from Seoul National University, Korea, in 1983 and 1995, respectively. Currently, he is a professor of Gyeongsang National University, Korea. His research interest includes DB, XML, and Artificial Intelligence.

Hyun-Syug Kang

Hyun-Syug Kang received the B.Sc. degrees in Computer Science from Dongguk University, Korea, in 1981, and the M.Sc. degree, Ph.D. degree in Computer Science from Seoul National University, Korea, in 1983 and 1989, respectively. Currently, he is a professor of Gyeongsang National University, Korea. He is a research fellow of Research Institute of Computer and Information Communication. His research interest includes Embedded DB, XML, MPEG-7, Neural Network, and Artificial Intelligence.