Recognition of Virtual Written Characters Based on Convolutional Neural Network

[1]Seungmin Leem, [*2]Sungyoung Kim
[1]*Consulting Group, AIW Co., mapsosa215@naver.com*
[*2]*Dept. of Computer Engineering, Kumoh National Institute of Technology, sykim@kumoh.ac.kr*

Abstract

This paper proposes a technique for recognizing online handwritten cursive data obtained by tracing a motion trajectory while a user is in the 3D space based on a convolution neural network (CNN) algorithm. There is a difficulty in recognizing the virtual character input by the user in the 3D space because it includes both the character stroke and the movement stroke. In this paper, we divide syllable into consonant and vowel units by using labeling technique in addition to the result of localizing letter stroke and movement stroke in the previous study. The coordinate information of the separated consonants and vowels are converted into image data, and Korean handwriting recognition was performed using a convolutional neural network. After learning the neural network using 1,680 syllables written by five hand writers, the accuracy is calculated by using the new hand writers who did not participate in the writing of training data. The accuracy of phoneme-based recognition is 98.9% based on convolutional neural network. The proposed method has the advantage of drastically reducing learning data compared to syllable-based learning.

Keywords*: Virtual Writing, Hand Written Character, Character Recognition, Convolutional Neural Network, Motion Recognition*

1. Introduction

In the information age, most of the data is being converted from analog to digital. Recently, the use of computers, smart phones, and tablets has become commonplace, and the distance between analogue and digital is becoming more and more intense. As a result, people are becoming less and less able to make handwriting and as a result, paper-based handwriting materials become harder to find around. However, in the educational environment, there are some cases where handwriting is performed through a digital input device. Typically, an electronic whiteboard is an example. Most of them are types of touch-screen electronic boards. The touch-based interface has become a common interface that is already commonplace in our daily lives.

Touch screens account for more than 50% of the price of electronic whiteboards and are expensive to maintain and repair. The reason is that the cost increases because the touch panel is attached to the existing display by using tempered glass and iron frame. In this study, we searched for ways to drastically reduce the price by interacting with the device in other ways than the touch method.

Generally, a user interacts with devices through voice, motion, or touch. In the case of speech recognition, recurrent neural networks are used in recent studies. However, it is not suitable for use in a classroom where there is a lot of noise. Motion is a useful technology used in various fields such as game, smart TV, and CCTV. In addition, recently, VR / AR technology has been attracting attention, and our lives will become more convenient if we can make a command or handwriting to a computer through motion analysis in a virtual environment. Therefore, in this paper, it is aimed to recognize the virtual handwriting by receiving the handwritten data in the three-dimensional space with the movement of the hand and analyzing it.

---

[*] Corresponding Author
Received: Mar. 2, 2018, Revised: Mar. 22, 2018, Accepted: Mar. 30, 2018

In this paper, we propose a recognition method for virtual handwritten Hangul input by user 's hand movements in 3D space where electronic whiteboard exists. For this purpose, consonant and vowel are distinguished by using the result of character stroke and moving stroke used in previous research [1], and convolutional neural network (CNN) is trained by using the results to enable accurate recognition. Compared to the syllable-based learning method, it is possible to reduce the amount of training data required when the training is performed based on the unit of the consonants and vowels. Finally, the recognized consonants and vowels are combined to provide the syllable information to the user.

2. Related works

Recognition on handwritten characters can be divided into off-line methods [2-4] that receive written handwriting characters as input and on-line methods [5-9] that use real-time and positional information obtained through motion. In the case of the off-line methods, there are advantages against the on-line methods that the recognition rate can be improved through the normalization process of various problems occurring in handwriting. However, there is a disadvantage that the results can not be verified in real time. The online method is a method of recognizing using data input in real time. Unlike the offline method, additional temporal information can be obtained.

In the past, for character recognition various features such as SVM and HMM were extracted and used for classification. It was very difficult process to define and extract effective and suitable features. In recent years, however, deep learning techniques have been developed and applied to handwriting recognition due to its superior performance. In particular, Korean handwritten characters have a complicated pattern and the number of classes is very large. Therefore, it is difficult to recognize than Arabic numerals and Latin alphabets, and the recognition accuracies are relatively low. However, it can be seen that the performance is improved by applying deep learning technology [10, 11]. In the case of Korean handwriting recognition, it is generally necessary to prioritize the character segmentation. For example, we can decompose a Korean syllable into some consonants and vowels, or into initial consonant, medial, final consonants, or even strokes. The reason for the preprocessing is from the structural characteristic of Hangul. Hangul syllables are composed of a combination of initial consonant, medial and final consonants in the computer. The number of possible combination of syllables with initial consonant, medial and final consonants is 11,172. Therefore, it is very difficult to obtain class-specific data for all syllables. Separating letters makes it possible to construct effective training and testing data sets without considering all possible syllables.

3. Localization and recognition of consonants/vowels

In case of motion-based writing in the virtual writing environment, both the character stroke and the movement stroke appear as shown in Fig 1. This phenomenon occurs because it does not include the process of specifying start and endpoints. Because the moving strokes can increase the difficulty of recognition step, we should remove the movement stroke in advance.

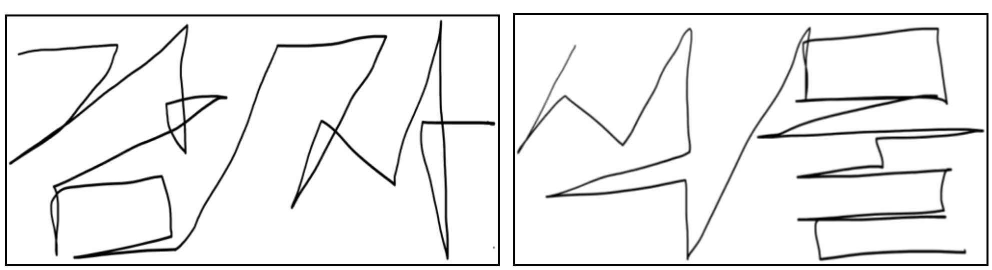


Figure 1. Examples of virtual handwritten characters that contain some character strokes and movement strokes

3.1. Detection on endpoints of strokes

The reason for detecting the endpoints is to divide a syllable into stroke units, which is the minimum unit constituting the syllable. The separated strokes are classified into movement strokes and character strokes. The detection method uses a characteristic that the direction changes from the current stroke to the next stroke. At that point, the value of the curvature becomes higher, which is defined as the endpoint. If we disconnect the continuous strokes based on the endpoint candidates, a syllable will be broken down into stroke units. In order to extract the endpoints, we apply the method as in the previous research [1].

3.2. Classification of stroke patterns

The characteristics of the character stroke and the movement stroke are as follows: direction patterns of '→', '↘', '↓', and '↙' occurs in case of character stroke. '↑', '↗', '↙', '←' and '↖' patterns in movement strokes [1]. Most of the patterns occur in just one of character strokes or movement strokes, but the '↙' pattern can occur in both type of strokes. The '↙' pattern can be classified according to context of adjacent patterns [1]. As you can see, there are all eight possible strokes pattern in all syllables of Hangul.

In the previous study [1], we classified stroke patterns based on chain code. After calculating 8-direction chain codes, a pattern of a stroke is defined according to main direction of the chain codes. However, it may be difficult to classify 'ㅇ' and 'ㅎ' since the direction of strokes in them changes continuously. Therefore, there are many points recognized as endpoints and all possible stroke patterns can occur. As a result, it was impossible to classify 'ㅇ' and 'ㅎ' with previous method.

Endpoints of 'ㅇ' and 'ㅎ' are characterized by very short distances between them. Based on this characteristic, we analyze the many handwritten data of 'ㅇ' and 'ㅎ'. We can recognize the consonants if the mis-detected endpoints are within a certain distance.

3.3. Localization of consonants and vowels

It is possible to classify strokes constituting the character stroke and the movement stroke through the preceding steps. However, for syllable recognition, it is necessary to discriminate consonants and vowels by grouping several separated strokes.

To localize consonants and vowels of a syllable, a binary image is generated by using the coordinate information of the separated strokes from virtual handwritten characters. When the binary image is created, some strokes can be treated as a single object in the case where the endpoints of them are connected. Next, labels are assigned to each object (part of a Hangul character) using the labeling technique. Then, a syllable is constructed using the positional relationship of the initial consonant, medial, and final consonants. There are some difficulties in the case of 'ㅊ' and 'ㅎ', because they are consist of several strokes in a consonant. So we separate these consonants with additional processing.

In the process of localizing consonants and vowels from separated strokes, various problems may arise. The biggest problem of them is that a single consonant or vowel can be separated to two or more subregions. In Fig. 2, the right consonant 'ㄹ' is erroneously detected as two partial regions. This is a case where some of the character strokes are misclassified as a moving stroke in the classification step, and so connection between two character strokes of 'ㄹ' is broken.

Figure 2. An example of a consonant with subdivided fragments

To solve this problem, some consonants and vowels are subdivided into several parts. Then, we define some possible combinations of them and train a convolutional neural network with the new patterns from the combinations. For example, 'ㅁ' is a consonant that can be combined with 'ㄱ' + 'ㄴ'. Except consonants and vowels that can be stroked at one go, we define separable combinations for the remaining syllables. Table 1 shows some examples of possible combinations.

Table 1. Some examples of possible combinations of strokes that make up a phoneme

phoneme	*Possible combinations*
'ㄷ'	('ㅡ' + 'ㄴ')
'ㄹ'	('ㄱ' + 'ㅡ' + 'ㄴ'), ('コ ' + 'ㄴ'), ('ㄱ','ㄷ')
'ㅏ'	('ㅣ'+'-')

Through the training of the combination of the additional defined patterns, it is possible to solve the degradation of the recognition accuracy by subdividing a phoneme.

3.4. Recognition based on convolutional neural network

In this paper, CNN with the structure as shown in Fig. 3 is used for training and recognition. As shown in Fig. 3, there are five convolution layers and a fully connected layers, and pooling layers are arranged in each convolution layer. This structure is very similar with LeNet-5 [12] that provides typical CNN architecture in character recognition. The dropout process is performed on the fully connected layer to minimize overfitting [13] of the network. The network uses ReLU [14] as the activation function. Although it is possible to improve recognition accuracy through delicate design of the structure, it is possible to confirm that satisfactory accuracy can be obtained even by using the proposed structure.

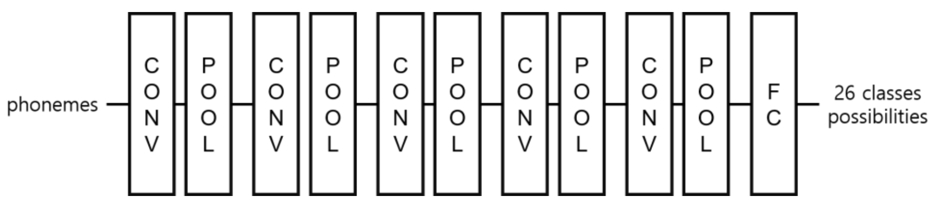
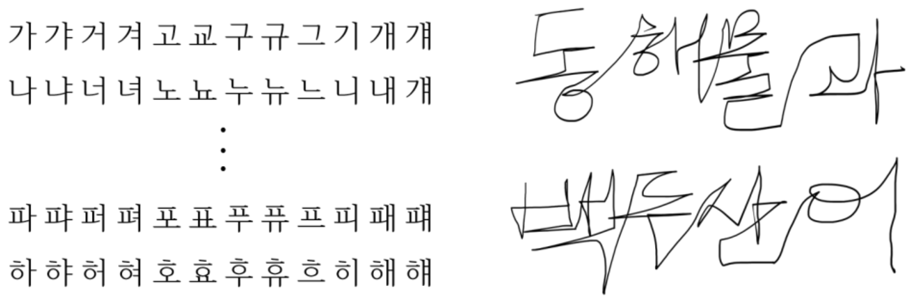


Figure 3. An architecture of proposed CNN

4. Experimental results

In order to construct the training data for the experiment in this paper, 168 syllables were defined by combining 14 consonants (except double consonants) and 12 vowels (10 vowels and 2 diphthongs). For this syllable, five writers wrote each syllable twice and so 1,680 syllables. Some examples of syllables used in training can be seen in Fig. 4(a).

The syllable shown in Fig. 4(a) are examples of used syllables in training. The actual data used for training is a handwritten syllable including a character stroke and a movement stroke as shown in Fig. 1. For training set, the movement strokes are removed and training step is performed after localizing consonants and vowels. For the test data set, we used virtual handwritten characters of Sections 1 to 4 of the national anthem. Fig. 4(b) shows parts of data set. Since writing area is limited in real space, we input test data into the network in unit of word.



(a) (b)

Figure 4. An architecture of proposed CNN

The total number of consonants and vowels to be separated in the first to fourth verse of the national anthem is 433. Table 2 shows the result of measuring accuracy by separating consonants and vowels from the method used in the previous study [1].

Table 2. Accuracies for localization of consonants and vowels

	# of phoneme	***accuracies***
1st verse	104	94%
2nd verse	110	97%
3rd verse	116	99%
4th verse	103	98%
Total	433	97%

In order to recognize, we train the proposed CNN using the consonants and vowels separated from the previous step. At this time, we have defined additional combinations for some consonants and vowels that cannot be drawn with one stroke, and they are used for training together. The CNN recognizes phonemes and some diphthongs are discriminated by combining with several vowels in post-processing step. For example, 'ㅔ' can be detected with sequence of 'ㅓ' and 'ㅣ'. The improvement in accuracy with the use of the new combination pattern was about 2.1%. Table 3 shows recognition results based on phonemes according to the use types of the training data set. A syllable can be composed from recognized consonants and vowels by using a composition table of Johab Hangul code. At this time, the accuracy can be improved by using the ensemble technique in CNN.

Table 3. Recognition accuracies of proposed methods

	Without additional patters	Using additional patters
CNN	96.3%	98.4%
Ensemble CNN	-	98.9%

5. Conclusion

In this paper, we proposed a recognition method for virtual handwritten characters input by hand motion in 3D space where camera is located. A virtual cursive character consists of a character stroke and a movement stroke. Since the movement strokes are an element that interferes with the character recognition, they should be removed. A character was separated into consonants and vowels using the pattern and positional relationship of the stroke after the removal. Using this result as an input to the convolutional neural network, a recognition model for virtual handwritten characters was created. Since the number of Hangul syllables is too many, syllable-based training is too difficult. The syllable was separated into consonants and vowels, and the amount of learning data was remarkably reduced.

6. Acknowledgments

This paper was supported by Kumoh National Institute of Technology

7. References

[1] S. L. Leem, H, S. Jeong, S. Y. Kim, “Remote Drawing Technology Based on Motion Trajectories Analysis”, Journal of KIIECT(domestic journal), Vol. 9, No. 2, pp.229-236, 2016.
[2] U.V. Marti, H. Bunke, “Using a statistical language model to improve the performance of an HMM-based cursive handwriting recognition system”, IJPRAI, Vol.15, No.1, pp.65-90, 2001.
[3] R. Kala, H. Vazirani, A. Shukla, R. Tiwari, “Offline handwriting recognition using genetic algorithm”, IJCSI, Vol.7, No.2, pp.16-25, 2010.
[4] B. Chacko, V. Krishnan, G. Raju, P. Anto, “Handwritten character recognition using wavelet energy and extreme learning machine”, Int. J. Mach. Learn. Cyber, Vol.3, No.2, pp.149-161, 2012.
[5] A. Graves, M. Liwicki, S. Fernandez, “A novel connectionist system for unconstrained handwriting recognition”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.31, No.5, 2009.
[6] M.K. Brown, W. Turin, J. Hu, “HMM based online handwriting recognition”, TPAMI, Vol.18, No.10, pp.1039-1045, 1996.
[7] X. Li, D. Y., Yeung, “On-line handwritten alphanumeric character recognition using dominant points in strokes”, Pattern recognition on 1997, Vol.30, No.1, pp.31-44, 1997.
[8] N. Arica, F. T. Yarman, “An overview of character recognition focused on off-line handwriting”, IEEE Transactions on Systems Man and Cybernetics, Vol.31, No.2, pp.216-233, 2001.
[9] R. Plamondon, S. N. Srihari, “Online and off-line handwriting recognition: a comprehensive survey, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.22, No.1, pp.63-84, 2000.
[10] I.J. Kim, X. Xie, “Handwritten Hangul recognition using deep convolutional neural networks”, IJDAR, Vol.18, No.1, pp.1-13, 2015.
[11] I.J. Kim, C. Choi, S.H. Lee, “Improving discrimination ability of convolutional neural networks by hybrid learning”, IJDAR, Vol.19, No.1, pp.1-9, 2016.
[12] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition” Proc. of the IEEE, Vol. 86, No. 11, pp.2278~2324, 1998.
[13] N. Srivastava, G. E. Hinton, A. Krizhevsky, “Dropout: a simple way to prevent neural networks from overfitting”, JMLR. 2014.
[14] V. Nair and G. Hinton “Rectified linear units improve restricted Boltzmann machines”, Proc. of International Conference on Machine Learning, pp. 807-814, 2010.