'91 韓國自動制御學術會議論文集 1991. 10.22~24

A Recognition of Hand Written Hangul by Fuzzy Inference

*Jeong Young Song, **Hee Hyol Lee, *Kageo Akizuki

* Faculty of Science and Engineering, Waseda University
** Faculty of Engineering, Fukuoka Institute of Technology

Abstract

Unlike printed character, the reccgnition of Hand written one has various kinds of difficulties due to the existence of the huge pattern associated with the person who writes. Therefore, in general, recognition of Hand written characters requires an algorithm which takes into consideration of the individual differences. Hangul characters are basically made of straight lines and circles. They can be represented in terms of feature parameters such as the end point of the straight line, the length and the angle. Then all Hangul characters can be represented by the number of basic segments(－, ／, ＼, ｜, ○) multiplied by the feature parameters respectively. In this study we propose a method for recognizing Hand written Hangul characters in terms of fuzzy inference.

1. Introduction

A character recognition as a branch of pattern recognition occupies an important place accompanied by the development of office automation as the method of man-machine interface. In general, the character recognition is classified into three large groups. These are printed character recognition, Hand Written character recognition and On-line Hand Written character recognition. The On-line Hand Written character recognition is a method recognizing by successive processing while it is writing. The Hand Written recognition is classified into the limited characters, the common Hand Written characters and the free Hand Written characters. In this study we deal with the common Hand Written characters, written in imitation of the sample characters.

On the other hand, Hangul character recognition has been developed. The recognition by means of the tree grammar[1] and the Hangul structure and the filtering[4] are reported. An advantage of these recognition methods makes the best use of one's Hangul structure analysis, but the disadvantage is in difficulty to analysis and assemble of the Hangul characters.

Therefore, in this study we take notice of the Hangul structure which is basically made of straight lines and circles. We extract basic segments(／, ＼, ｜, －, ○) for the Hangul characters, and then we extract the feature parameters from each segment. In particular, we take consideration of the fuzziness in the direction and the length of each segment, because the feature parameters of the segments differ from sample ones. Lastly, we propose a method for recognizing Hand Written Hangul character in terms of fuzzy inferrence.

2. The Structure of the System

The structure of the system for this study is composed from a camera as hardware transforming Hand Written Hangul characters into digital data, personal computer and software as shown in Figure 1.

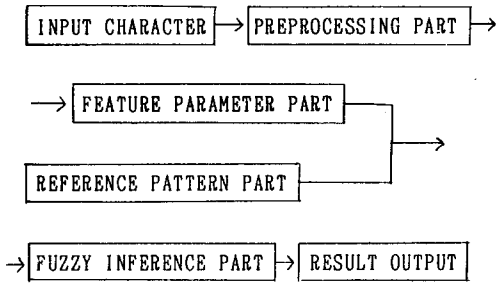


Fig. 1 The structure of the system

The explanation for each part of Figure 1 is as follows: Input character part is written as a Hand Written Hangul character on 6 by 6 cm paper and that changes into digital data by a camera. Preprocessing part centers the character and performs thinning algorithm. We use the Hildith's algorithm for thinning. The feature parameter part seeks the number of segments, starting and ending coordinates of each segment, the length of the segment and the angle as the feature parameters. Reference part is registered as a fuzzy data for each segment of the character through preparatory experiment. The fuzzy inference part calculates the grade of feature parameters by means of fuzzy production rule, and let the maximum value of average of the grades be the inferenced result.

3. Preprocessing

The character data is changed into dots(1 or 0) of multiply 640 by 378. Furthermore, if '1' among 25 dots is greater than 13 dots then we change a mass which is 25 dots into '1' or else we change a mass into '0', because this character information has too much. A system compresses the character information into coordinates of 82 by 82 of 4 by 4 bits. It performs Hilditch's thinning algorithm because it is difficult to extract directly feature parameters from the character information. The process, from input character to end of preprocessing for Hand written Hangul character'영'(Young) as a sample, is shown below.

Fig. 2 Input character'영'(Young)

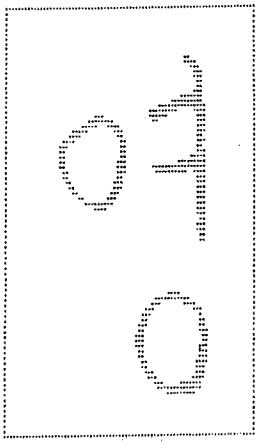


Fig. 3 Before performing thinning algorithm

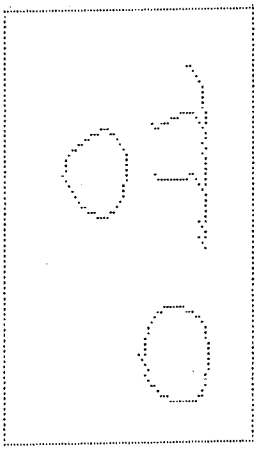


Fig. 4 After performing thinning algorithm

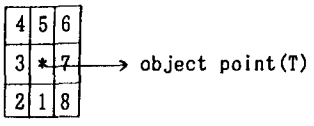
4. Extraction of Feature Parameters

The character information of the end of preprocessing is coordinates of 82 by 82 and it

extracts the feature parameters from the character information. The feature parameters are the number of segments, the starting and ending coordinates of each segments, the length of the segment and the angle. It seeks the feature parameters as follows:

Algorithm

Step 1. Search at X direction with 3 by 3 mask.



3 by 3 mask

Step 2. If T is '1' then go to step 3 or else go to step 1.

Step 3. If the '1' is only one around T then go to step 4 or else go to step 5.

Step 4. If the number of '1's is greater than 5 in the same direction then the coordinate becomes the starting coordinate and then search coordinates by permiting 2 of right and left of direction code in proceed direction until ending condition erasing the T. Go to step 8.

Step 5. If mask has two '1's around T then search in small number of direction of 3 by 3 mask. Now, T is not erased and if starting coordinate equal to ending coordinate then go to step 6 or else it searches until it matches ending condition. If the coordinate becomes the starting coordinate, go to step 7.

Step 6. In this case, that is a circle or a square. If the number of '1's is greater than 6, a square(SIKA) is formed. If there are 2 or more SIKAs, then go to step 1 or else erase the circle. Go to step 1.

Step 7. Search coordinates by permiting 2 of right and left of the direction code in proceed direction and erase T.

Step 8. If the coordinate is ending condition, then extract feature parameters using the starting and ending coordinates. But if the length of the segment is less than 5, abandon the segment.

Step 9. If Y is 82 then stop or else go to step 1.

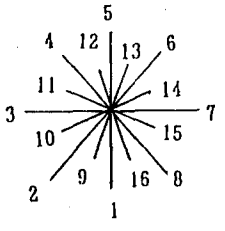


Fig. 5 Direction code

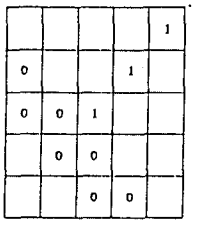
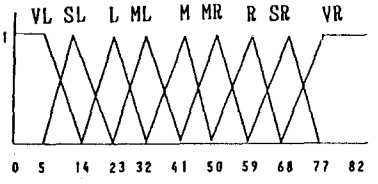


Fig. 6 A sample of ending condition for 2 direction

5. Reference Pattern

The feature parameters are starting coordinates XO, YO, the length L, the angle θ. The membership function of the fuzzy representation is decided through a preparatory experiment.



VL: furthest left SL: far left
L : left ML: left from the center
M : center MR: right from the center
R : rignt SR: far right
VR: furthest right

Fig. 7 Membership function of feature parameter(XO)

The reference pattern is registered as a form of fuzzy data. In general, the feature parameters depend on writers, for example, disparities between man and woman, and on their age, and on the written situation. In this study we will recognize Hand Written Hangul character in terms of fuzzy inference. The fuzzy representation of

XO as a sample is shown in Table 1.

XO	value of fuzzy	representation of fuzzy
0-14	VL	location on the furthest left side
5-23	SL	location on the far left side
14-32	L	location on the left side
23-41	ML	location on the left side from the center
32-50	M	location on the center
41-59	MR	location on the right side from the center
50-68	R	location on the right side
59-77	SR	location on the far right side
68-82	VR	location on the furthest right side

Table 1. The fuzzy representation of XO

The recognition is done by production rules using fuzziness data of the reference pattern part for each segment. The rule is a form of IF (condition) THEN (result). A rule for Hand Written Hangul character '영'(Young)as a sample is shown in Figure 8.

IF XO1=SR AND YO1=VU AND L1=SLG AND θ1=100 AND
XO2=L AND YO2=U AND L2=0 AND θ2=360 AND
XO3=MR AND YO3=U AND L3=S AND θ3=180 AND
XO4=MR AND YO4=MU AND L4=S AND θ4=180 AND
XO5=MR AND YO5=MD AND L5=0 AND θ5=360 AND
THEN '영'(Young)

Fig.8 The production rule for Hand Written Hangul character' 영 '(Young)

6. Fuzzy Inference

The system extracts feature parameters from input character for each segment. The feature parmeters are changed to fuzzy data. The fuzzy data for input data '영'(Young) as a sample is shown in Figure 9.

For the first segment;
XO1 is located on the right side and,
YO1 is located on the furthest up side and,
L1 is the longer distance and,
θ1 is an angle of about 100.

For the second segment;
XO2 is located on the left side and,
YO2 is located on the up side and,
L2 is the shortest distance and,
θ2 is an angle of 360.

For the third segment;
XO3 is located on the right side from the center and,
YO3 is located on the up side and,
L3 is a short distance and,
θ3 is an angle of about 180.

For the fourth segment;
XO4 is located on the right side from the center and,
YO4 is located on the up side and,
L4 is a short distance and,
θ4 is an angle of about 180.

For the fifth segment;
XO5 is located on the right side from the center and,
YO5 is located the down side and,
L5 is the shortest distance and,
θ5 is an angle of 360.

Fig.9 The fuzzy data for the input data'영'(Young)

The obtained fuzzy data is send to the reference part by the group of segment number. It performs the fuzzy inference by the production rule. The groups of Hangul characters by the number of segments are shown in Figure 10.

2 : 이(i) , 으(eu)
3 : 오(o) , 기(ki), 우(u) , 의(eui) , 응(eung). . .
4 : 인(in), 노(no), 소(so), 사(sa), 나(na), 익(ik). . .
5 : 주(ju), 강(kang), 자(ja), 조(jo), 식(sik) , 호(ho). .
6 : 곤(kon) , 정(jeong) , 동(dong), 진(jin) , 윤(yun) . .
7 : 배(be), 김(kim) , 민(min) , 한(han) , 창(chang) . .
8 : 황(whang) , 관(kwan), 현(hyun), 남(nam) , 박(park). .
9 : 열(yeul), 백(bek) , 택(tek) , 란(ran) , 록(rok) . .
10 : 범(beum), 련(ryun), 필(pil) , 림(rim)

11 : 췰(cheul) , 팰(pal)
13 : 뮬(ryul), 렐(ryel)

Fig.10 Groups of Hangul characters by the number of segments

We recognize Hangul characters before and after of the number of the segments. And then we calculate the grade of feature parameters by means of fuzzy production rule, and let the maximum value of average of grades be the inferenced result[3]. We calculate the grade in terms of minmax operation as in Figure 11.

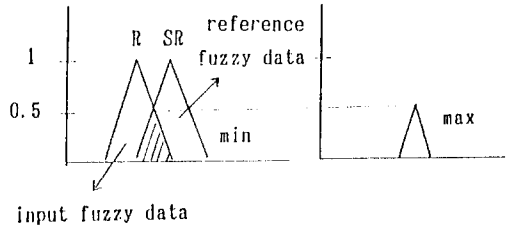


Fig.11 Minmax operation

$$Grade = \frac{(WX1+WY1+WL1+W\theta 1)+\ldots+(WXN+WYN+WLN+W\theta N)}{4*N}$$

(N:number of segment)

7. Experimental Result

The random names taken from Korean spiritual and cultural research institute published in 1983 are written by some people. We made production rule for 154 Hangul characters. We applied the recognition method with the Hangul characters written by four men and three women, and got the 80 percent results. The effectiveness of this method has been confirmed through simulation studies. The influence of results of thinning algorithm and of the segments which come in contact with another one was a major cause of its mistake. Hereafter we will intend to improve algorithm as being able to recognize a character in the running style.

References

1. Takeshi AGUI, Masayuki NAKAJIMA, KIM. A Hangul recognition by tree grammar, The Transaction of the Institute of Electronics and Communication Engineering of JAPAN, Vol. J61-D. No. 11, pp. 866-873, 1978. 11
2. Kenji OHMORI, Online Handprinted Character Recognition by Fuzzy Inference, The Journal of the Institute of Electronics, Information and Communication Engineers, Vol. J72-D-II, No. 3, pp. 369-379, 1989. 3
3. CHOI, LEE, AKIZUKI. A Korean Speech Recognition by Fuzzy Rule Base, The Institute of Electrical Engineers of JAPAN, Vol. 111C, No. 5, 1990
4. OH, SHIN, CHIN, A Recognition Algorithm of Hangul Alphabet Using 2-D Digital Filtering, Vol. 21, No. 3, pp. 55-59, 1984. 2