Accuracy Improvement for Handwritten Japanese Word Recognition by Combination of Character and Word Recognizer

Yoshinobu Hotta, Hiroaki Takebe, Misako Suwa, Satoshi Naoi
Fujitsu Laboratories Ltd.
4-1-1, Kamikodanaka, Nakahara-ku, Kawasaki, 211-8588, Japan
E-mail: {y.hotta, takebe.hiroaki, suwa, naoi.satoshi}@jp.fujitsu.com

Abstract

This paper proposes a combination method of character recognizer (CR) and word recognizer (WR) for Japanese handwritten words. CR is composed of character segmentation, isolated character recognition, and postprocessing. On the other hand, WR recognizes word images holistically based on a given word lexicon. These methods have been used as key modules in handwritten address recognition but have not been compared quantitatively up to now. In this paper, CR and WR are first empirically compared with each other using synthesized words generated from character images. Based on the comparison, a combination method is then proposed. Experimental results show that the combination approach can achieve better recognition performance than the other two methods.

1. Introduction

There are still a lot of cases where people fill in information such as the address or the name by hand in writing mail, registration sheets of bank card, etc. To utilize user peculiar information effectively, it is needed to recognize handwritten words by OCR (Optical Character Reader) with high accuracy. However, the number of categories of Japanese Kanji character is huge (typically, 4,000), and character shapes are very complicated compared with alphabets. Furthermore, there are many types of touching patterns between adjacent characters. Segmentation and recognition of Japanese handwritten words is therefore a very difficult problem. Though several segmentation methods for handwritten Japanese words have been proposed [1]-[3], these methods are not always effective especially when adjacent characters touch at multiple points or overlap heavily. In particular, when a writing box is small, heavy touching appears to occur frequently. On the other hand, the holistic word recognition approach is proposed to avoid such difficult segmentation problem in which the word image is recognized holistically without segmentation [4]-[7]. But character recognition based approaches and the word recognition based approaches have not been compared quantitatively up to now.

In this paper, CR and WR are first empirically compared with each other using synthesized words generated from character images. A combination method of CR and WR is then proposed based on the comparison. Finally, effectiveness of the combination method is demonstrated in real word images.

2. Character Recognizer

2.1. Character Segmentation

First connected components (CCs) are extracted by labeling. Next if the bounding box of each CC overlaps vertically, then these CCs are integrated. After that, the aspect ratio of each CC's bounding box is calculated. If the ratio is less than a predetermined threshold ($height/width < Th1$), the vertical projection of black pixels of the CC is counted and if the histogram value is less than another predetermined threshold at a certain point, then the CC is cut at that point. $Th1$ is set to 1.0 so that touching parts between adjacent characters can be cut. Although projection analysis is a very simple segmentation method, the result can still serve as a guidepost [2,3].

2.2. Recognizer

Not only each CC, but also the combination of adjacent CCs can be a character candidate. A candidate image is first size-normalized to 64*64. Then the normalized image is divided into 13*13 blocks and local chain code histograms of the character contour with 16 directions are extracted

Proceedings of the 2005 Eight International Conference on Document Analysis and Recognition (ICDAR'05)
1520-5263/05 $20.00 © 2005 IEEE

from each block. Next the 13*13 blocks are down sampled with Gaussian filter into 7*7 blocks, and the 16 directions are also down sampled into 8 directions. Finally, dimension of the feature becomes 392(=7*7*8) [8]. The extracted feature and reference feature vectors in character feature dictionary are matched by Modified Quadratic Discriminant Functions (MQDF) [8,9]. For each candidate, the recognized character code and the distance are output. At that time, a given word lexicon is utilized so that the characters within the lexicon are output.

2.3. Postprocessing

From all segmentation paths, the optimal path, which minimizes the summation of character distances, is selected by DP search. This method is called the candidate character lattice method [10]. Figure 1 illustrates the candidate character lattice method. The printed characters are the recognition results of CR, and the number under each character describes the minimum distance of CR. In Figure 1, "平塚" is selected because the path minimize the summation of the distances.

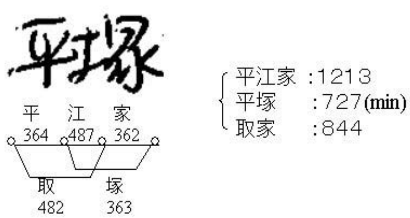

Figure 1. Candidate Character Lattice Method

3. Word Recognizer

3.1. Overview

As for the holistic word recognition approach, we adopted the method in which word features are synthesized from character features [6]. Figure 2 shows the whole flow of the word recognizer. For the input word image, size-normalization is conducted, and then the feature is extracted directly from the normalized image (Fig.3). Here, the feature is the same as the character feature in Section 2.2. As for reference feature vectors, word features are synthesized from character features based on list of words. Then the synthesized word feature (***s***) and the extracted word feature from the normalized image (***x***) are matched by Euclidean distance ($d(\boldsymbol{x},\boldsymbol{s}) = (\boldsymbol{x}-\boldsymbol{s})^2$).

This method is effective especially when the total number of words is too much or the word names such as city names change frequently because of merging.

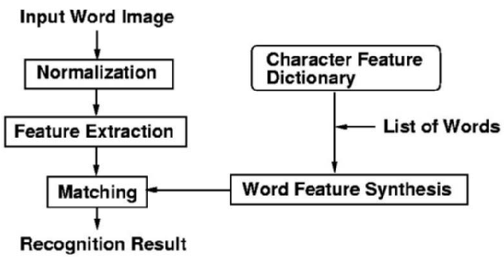

Figure 2. Whole Flow of Word Recognizer

The basic concept of this approach is presented as follows. If only the local kind of feature is used, a word feature extracted from the normalized word image (Fig.3) can be approximately synthesized from character features (Fig.4). Therefore character features extracted from shrunken character images are stored in the dictionary for word feature synthesis.

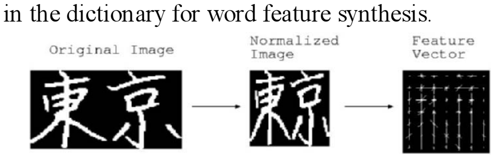

Figure 3. Word Feature From Word Image

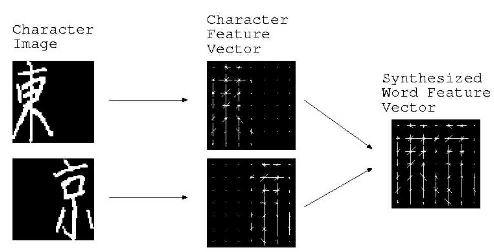

Figure 4. Synthesized Word Feature from Character Feature

3.2. Character Features for Word Feature Synthesis

As explained in Section 3.1, character features for word feature synthesis are extracted from horizontally shrunken character images. The procedures are described in the following.

1) Shrink the character image so that it occupies n/m ($n=1,2,.., m$) of the width with respect to the word normalization size. Here m is predetermined and fixed to 6, and the normalization size is 64*64.

2) Generate plural character features according to the position of the image on word image size. When *m* is fixed, there are *m-n+1* kinds of variations. The number, *n,* changes from *1* to *m*, thus totally *m*(m+1)/2* kinds of character features are obtained per category. When *m* is *6*, total number of character features per category is 21.

3) When *m* and *n* are fixed, mean feature of all samples belonging to the same category becomes the character feature.

An illustration is described in Figure 5.

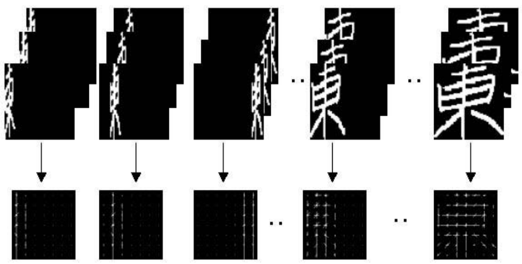

Figure 5. Character Features for Word Feature Synthesis

3.3. Word Feature Synthesis

Word features are synthesized from character features according to a given list of words. For example, in recognizing city names, a lexicon of cities is given. The character features used for the synthesis are selected according to the complexity degree of the characters included in the word. Here, the complexity degree of the character is measured by the vertical line density of the character. For example, when the word feature " 東京" is synthesized, the character feature of " 東 " extracted from *3/6* width image, and the feature of "京 " extracted from *3/6* width image are used. To adjust the fluctuation of character width in input word image, other character features are also used. That is, the combination of *4/6* width of "東 " and *2/6* width of "京", or that of *2/6* width of "東" and *4/6* width of "京" is used. Figure 6 shows the example of three types of synthesized word features of " 東京".

4. Word Image Generator (WIG)

To evaluate the performance of CR and WR quantitatively, test word images are generated automatically by using the character images acquired from the character database. In generating word images, several parameters can be considered such as each character size, the distance between characters, the center of each character etc. Here, the size of each character is the same as that in the character database, and the center position of each character is also set to be the same.

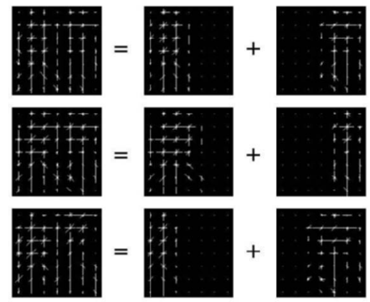

Figure 6. Example of Three Types of Synthesized Word Features of " 東京 "

The distances between characters are set to be the same as well, but they are gradually changed in order to control the touching degree between characters. Figure 7 shows an example when the number of character is three. The rectangle means the virtual bounding box of each character. The distances (*d*) between characters (C1-C3) are the same, and the center position of them is also the same. The upper figure shows an example when $d > 0$, while the lower one shows the case when $d < 0$.

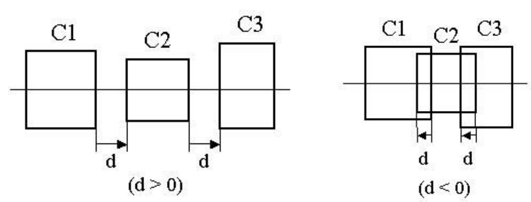

Figure 7. How to Generate Word Image

Examples of synthesized word images are shown in Figure 8. The distance is gradually changed with d=20, 0, -5, and -10 dot, because the size of the character image is about 30-60dot. The smaller the distance is, the heavier the touching is.

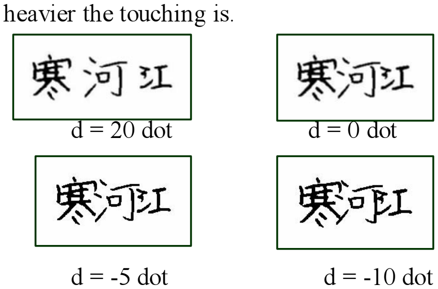

Figure 8. Generated Word Images

Proceedings of the 2005 Eight International Conference on Document Analysis and Recognition (ICDAR'05)
1520-5263/05 $20.00 © 2005 **IEEE**

5. Preliminary Experiment

5.1. Performance Comparison of CR and WR

City names are used for test and WIG generates the words. The total number of categories is 671. The distance between characters is changed with *d*=20, 0, -5, -10. According to the distance, there are four datasets and one word image is generated for each category. Thus there are totally 671 word images when the distance is fixed. The word recognition rate is defined as follows.

*Word Recog. Rate = (number of correctly recognized word) / (total number of words) * 100.0*

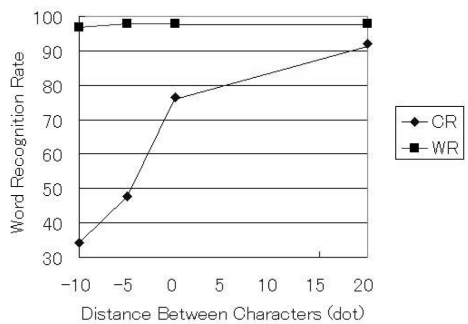

Figure 9. Comparison of Character and Word Recognizer

When the distance between characters is 20, the word recognition rate of CR is about 92.1% and that of WR is about 97.8%. When the distance is minus, the word recognition rate of CR drops significantly whereas that of WR is still higher than 95.0% (Fig.9).

5.2. Error Analysis

An error analysis of CR and WR is examined in detail in this section. There are distinct differences between the error tendencies of CR and WR.

1) CR:incorrect and WR:correct

When adjacent characters touch heavily and their strokes overlap or touch at multiple points, CR cannot cut each character, and fails to recognize the word. WR is relatively robust to these touching cases.

Figure 10. Examples of (CR,WR)=(NG,OK)

2) CR: correct and WR: incorrect

Even when touching is not so heavy, WR sometimes fails to recognize the word if some other words contain similar shapes. The Euclidean distances of 1st and 2nd candidates are very close then. This is an essential problem of WR, and does not depend on the degree of touching. However, CR can correctly recognize the words. In Figure 11, the left words under the word image mean the answer and the right words mean the recognition result of WR.

Figure 11. Examples of (CR,WR)=(OK,NG)

6. Combination of CR and WR

6.1. Algorithm

Although a combination method using the *Borda Count* is proposed [11], our strategy is different because the error tendencies and the recognition performance of CR and WR are different with each other. Here the recognition result of WR is partly complemented by CR. Figure 12 illustrates the whole flow.

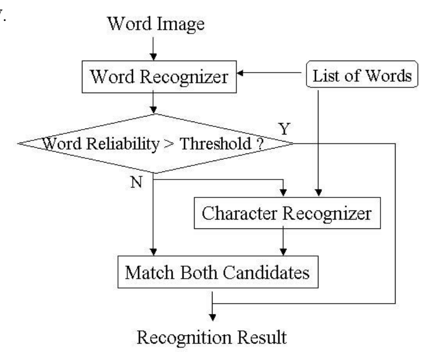

Figure 12. The Whole Flow of Combination Method

(1) First, WR is conducted for the recognition of the input word image.

(2) If the recognition reliability of WR is higher than a predetermined threshold, the word recognition result is output; otherwise, CR is performed in order to discriminate among those similar words. As for the recognition reliability, difference of the Euclidean distance between 1st and 2nd candidates are adopted.

(3) The recognition results of WR and CR are matched if CR is performed. WR outputs several recognition candidates with higher reliabilities, and

CR also outputs several character candidates with higher reliabilities. If the character candidate from CR is included in WR candidates, the reliability of WR is replaced with that of CR. If several characters in a word are included, the average reliability of each character is set.

(4) Sort the word candidate reliabilities based on new reliabilities obtained by (3).

Figure 13 illustrates the combination process.

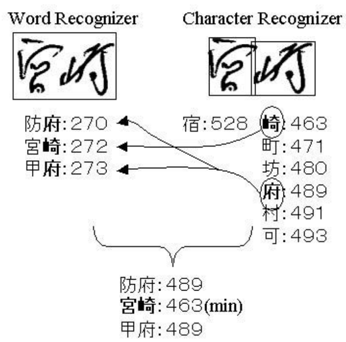

Figure 13. A Sample of Combination Process

6.2. Experiment

Experiments are conducted to compare the combination method with CR and WR. City names are used for test and real word images are gathered from real mails. Total number of categories of city names is 671. Total number of the word images is 217, and that of word categories included in the images is 109. Writing quality of the images is low and touching among characters occurs frequently. The definition of the word recognition rate is the same as that in Section 5.1.

Table 1. Experimental Results.

	CR	WR	Combination
Word Recog. Rate	59.9%	86.2%	91.7%

By combining CR with WR, the word recognition rate is increased from 86.2% to 91.7%.

7. Conclusion

A character recognizer and a word recognizer are compared for Japanese word recognition using synthesized words generated by the word image generator. When the distance between characters is big, the word recognition rate of CR is higher than 90.0 %, but when the distance is minus and touching occurs frequently, the rate drops less than 35%. On the other hand, WR is relatively robust for multiple touching, overlapping of characters. Even when the distance is minus, the word recognition rate of WR is higher than 95.0 %. But if the input word contains other words with similar shapes, WR may not make the correct recognition. To solve these problems, a combination method of CR and WR is proposed. In this method, CR can solve the problem of WR when the distance is not so small. Experiments are performed using real word images. The results show that the combination method can achieve the highest recognition rate.

8. References

[1] H.Nishimura, Y.Nakano, and H.Ikeda, "A Segmentation Method for Touching Handwritten Japanese Characters", Proc. of DAS'98 ,pp.252-261, 1998.

[2] H.Ikeda, Y.Ogawa, M.Koga, H.Nishimura, H.Sako, and H.Fujisawa, "A Recognition Method for Touching Japanese Handwritten Characters", Proc. of ICDAR'99, pp.641-644, 1999.

[3] T.Yamaguchi, T.Yoshikawa, T.Shinogi, and S.Tsuruoka, "A Segmentation Method for Touching Japanese Handwritten Characters Based on Connecting Condition of Lines", Proc. of ICDAR'01, pp.837-841, 2001.

[4] O.Shiku, A.Nakamura, H.Kuroda, and S.Miyahara, "A Method for Handwritten Japanese Word Recognition Based on Holistic Strategy", IPSJ, Vol.41, No.4, pp.1086-1095, 2000.

[5] S.Naoi, M.Suwa and Y.Hotta, "Recognition of Handwritten Japanese Addresses based on Key Character Extraction and Holistic Word Matching," Proc. of DAS'98, pp.149-152, 1998.

[6] Y.Hotta, H.Takebe and S.Naoi, "Holistic Word Recognition based on Synthesis of Character Features," Proc. of DAS'00, pp.313-324, 2000.

[7] C.H.Wang, Y.Hotta, M.Suwa and S.Naoi, "Handwritten Chinese Address Recognition," Proc. of IWFHR'04, pp.539-544, 2004.

[8] T.Wakabayashi, Y.Deng, S.Tsuruoka, F.Kimura and Y.Miyake, "Accuracy Improvement by Nonlinear Normalization and Feature Compression in Handwritten Chinese Character Recognition," IEICE Trans., Vol.J79-D-II, No.5, pp.765-774, 1996.

[9] F.Kimura, K.Takashina, S.Tsuruoka, and Y.Miyake, " Modified Quadratic Discriminant Functions and the Application to Chinese Character Recognition", IEEE Trans. PAMI 9, No. 1, pp.149-153, 1987.

[10] Y.Murase, T.Wakahara, and M.Umeda, "Online Writing-Box Free Character String Recognition by Candidate Character Lattice Method", IEICE Trans., Vol.J68-D, No.4, pp.765-772, 1985.

[11] J.J.Hull, T.K.Ho, J.Favata, V.Govindaraju and S.N.Srihari, "Combination of Segmentation-based and Wholistic Handwritten Word Recognition Algorithm", Proc. of 2nd IWFHR, pp.229-240, 1991.

Proceedings of the 2005 Eight International Conference on Document Analysis and Recognition (ICDAR'05)
1520-5263/05 $20.00 © 2005 IEEE