Online Recognition of Unconstrained Handwritten Japanese Text Using Statistical Information

Katsumi Marukawa, Takeshi Nagasaki, Kazuyoshi Kikuta

Online Recognition of Unconstrained Handwritten Japanese Text Using Statistical Information

Katsumi Marukawa and Takeshi Nagasaki
Central Research Laboratory, Hitachi Ltd.
1-280 Higashi-koigakubo, Kokubunji-shi,
Tokyo 185-8601, Japan
{marukawa, naga-t}@crl.hitachi.co.jp

Kazuyoshi Kikuta
Government & Public Corporation Information Systems Division, Hitachi Ltd.
1-6-27 Shinsuna Koto-ku, Tokyo 136-8632, Japan
kazuyoshi.kikuta.jf@hitachi.com

Abstract

We developed an online recognition system to read unconstrained handwritten Japanese texts using statistical information. The substantial problems of reading handwritten Japanese text are how to correctly segment freely written characters and how to correct errors of character classification. Our method searches for the best interpretation by integrating the likelihoods of character segmentation, character classification and language processing. For the language processing, candidates of words in any position are extracted from a dictionary including about 240,000 words, and the extracted words are evaluated using their grammatical connective probability and word bi-gram probability as the context. Experiments using 467 texts showed that our complete method is more accurate than any partial method. The rate of recognition per text improved from 10 to 38%, and the rate of recognition per character improved from 72.2 to 82.1%. The effectiveness of our method was proved.

Keywords: online recognition, unconstrained handwritten Japanese text, language processing, grammatical connective probability, word bi-gram probability

1. Introduction

With the coming of the ubiquitous information age, information freely handwritten using devices such as digital pens, tablet PCs, digital boards, etc. needs to be input as text data to computers. The technology to read handwritten strings such as addresses and names has been developed. However, the technology to read unconstrained handwritten text such as documents and memos has not been developed yet. Therefore, products, services, and solutions that use handwritten information are difficult to create. To solve this problem, the technology to read unconstrained handwritten text is necessary.

In this paper, we described a technology for reading the unconstrained handwritten Japanese text, as shown in Figure 1. Japanese text includes about 5,000 characters, many of which are quite similar. Japanese text does not use splitters such as the spaces in English.

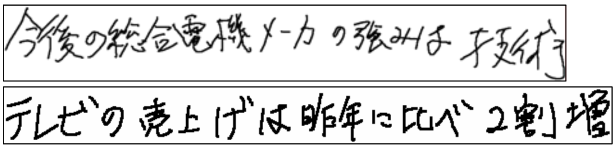


Figure 1. Examples of handwritten Japanese text.

Several methods have been developed for the online reading of Japanese text. One [1] is an initial attempt. This method finds the best interpretation of a character pattern sequence by using DP-matching. However this method does not consider the likelihood of character segmentation; it only considers the likelihood of the context. The two methods [2,3] consider the likelihoods of character segmentations, character classifications and character bi-gram probability as the context. These approaches are similar and formulate the problem as a search for the most probable interpretation.

Our method has three features. First, it expresses the result of pre-segmentation by using a network. We call this the segmentation network (S-network). Thus, our method can control as units of several connected components. Japanese text written in various pitches, sizes, and aspect ratios can be read. The patterns in the S-network are classified, and the results are expressed by using another network. We call this network the character network (C-network). Second, our method uses recursive transition network (RTN)-type word matching and devise a control of this word matching. Words in any text position are extracted from the C-network by using about 240,000 words. This word matching can easily deal with conjugated (verbs, adjectives, etc.) or unconjugated words (noun, etc.) in all. Finally, our method uses the likelihoods of character segmentation, character classification and language processing. It searches for the most probable interpretation. As the likelihood in language processing, grammatical connective probability and word bi-gram probability are used. In particular, using grammatical connectivity removes many false words, and using of word bi-gram improves the ratio of right words. We evaluated our method by using 467 samples and obtained much higher accuracy than a conventional method.

The rest of this paper is organized as follows. Section2 gives an overview of previous work. Section 3 explains the problems of reading handwritten Japanese text. Section 4 presents the language processing and how we integrated the likelihood of results on character segmentation, character classification, and language processing. Section 5 reports our experiments and results. Finally, in Section 6, we draw some conclusions.

2. Previous work

2.1. Reading printed Japanese text

To read pre-determined content such as addresses and names, a method of using a relevant knowledge base has been developed. However, no particular knowledge base is used for reading Japanese text. In this case, a general word dictionary and grammatical rules are used as a knowledge base. Methods for reading printed Japanese text can be classified into three main groups as follows [4].

(1) Correction of rejected character [5]

First, the first character candidates with a high likelihood are regarded as characters without missing. Others characters are regarded as rejected characters. Second, rejected characters are corrected to suitable characters by using character bi-gram probability. This method cannot correct to suitable characters when wrong characters are not detected as rejected characters.

(2) Correction using spelling correction function [6,7]

First, the spelling correction function detects unsuitable characters from the text consisting of the first character candidates. After the second step, the same processes as in method (1) are carried out.

(3) Correction by using word matching [8-10]

This method uses a general word dictionary and grammatical rules. First, words in any position are extracted by referring to the word dictionary. A suitable interpretation is searched for from the extracted words by referring to the grammatical rules.

Method (3) has a higher accuracy than the others. However, Method (3) needs much more CPU power than the others. Readers for printing Japanese text are mostly successful at character segmentation. The above methods are designed under this assumption. Therefore, we cannot obtain high accuracy on handwritten Japanese text even if method (3) is applied.

2.2. Reading handwritten Japanese text

In general, the processing for handwritten Japanese text readers consists of the three processes shown in Figure 2.

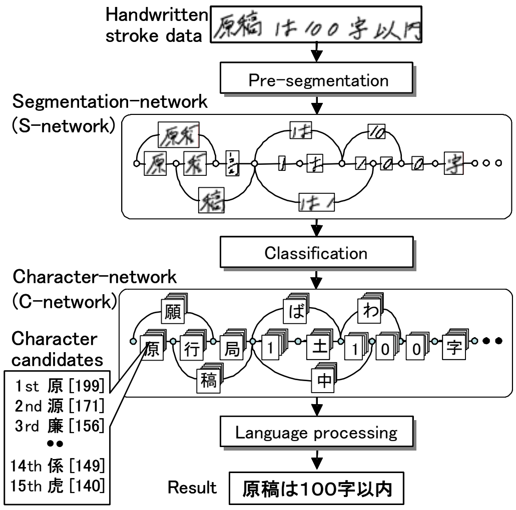


Figure 2. Processes of reading handwritten Japanese text.

The processes work as follows.

(1) Pre-segmentation

This process extracts possible patterns as characters from handwritten stroke data. An S-network is made.

(2) Classification

This process classifies patterns included in the S-network. A C-network is made.

(3) Language processing

This process searches for the best interpretation using linguistic information such as character bi-gram probability of Japanese text.

We now explain the conventional methods described above in detail. The first method [1] uses a C-network and a morphological analyzer. First, this method makes possible strings to combine character candidates in a C-network. Second, the morphological analyzer processes each string. Finally, a suitable string is determined. This method does not consider the likelihood of pre-segmentation. Therefore, this method is not adequate for to resolve the ambiguity of pre-segmentation. The other two methods [2,3] consider the likelihood of pre-segmentation, character classification, and character bi-gram as the context. One of them [2] makes S-networks using limited possible patterns. Therefore, the right patterns might be not in the S-network. The other [3,11] considers character size variations. This method uses the likelihood of size variations as a part of likelihood. However, this method pre-segments by using histogram projections. Therefore, this method does not have adequate accuracy. These two methods also use character bi-gram probability in language processing. Therefore, they cannot obtain higher accuracy than a method using grammatical rules and word bi-gram probability. To obtain higher accuracy, such information should be considered. We propose a method that does so.

3. Problems of reading handwritten Japanese text

3.1. Pre-segmentation

One problem of pre-segmentation is that a pattern sequence cannot be uniquely determined using only the form of a segmented pattern. That is, possible patterns should be extracted without missing the right patterns. This is because we cannot obtain the right interpretation if the right patterns are not in the S-network. On the other hand, restraining the number of possible patterns as much as possible is necessary to avoid creating a large search space. Calculating the likelihood of the segmentation of each pattern is also necessary. We adopted the method in Ref. [12] to calculate the likelihood of the segmentation.

3.2. Classification

A classification module that outputs the right character as the first character candidate would be desirable. However, this is difficult in really. Therefore, the right characters should be as high ranking as possible and should always be among the candidates. We need to research classification steadily.

3.3. Language processing

Obtaining the right interpretation by using only the results of pre-segmentation and classification is difficult. Language processing must use a Japanese knowledge base: grammatical rule, word bi-gram probability, etc. The problems of this process are extracting the right words from the C-network and finding the right word sequence from the extracted words. We describe this process in detail in the next section.

4. Language processing

4.1. Outline of processing

Our proposed language processing mainly consists of three modules, as shown in Figure 3. First, complex word matching is carried out. This module inputs a C-network and outputs possible words with matching costs and grammatical attributes. This module matches character candidates in a C-network with a word dictionary including about 240,000 words. The matching process is carried out at each node of a C-network to extract words at any node. The extracted words are expressed in a network called a word network (W-network). Second, the word connectivity is evaluated. This module inputs a W-network and outputs possible word pairs with evaluation values. These outputs are expressed in a network called an evaluation network (E-network). This module creates word pairs by using a W-network and gets the connective possibility of each word pair using grammatical rules. Possible word pairs are selected. Then, the evaluation values of the possible word pairs are calculated. The values are attached to E-network. Finally, a path search is carried out. This module searches for the most suitable path from an E-network by applying the Viterbi algorithm.

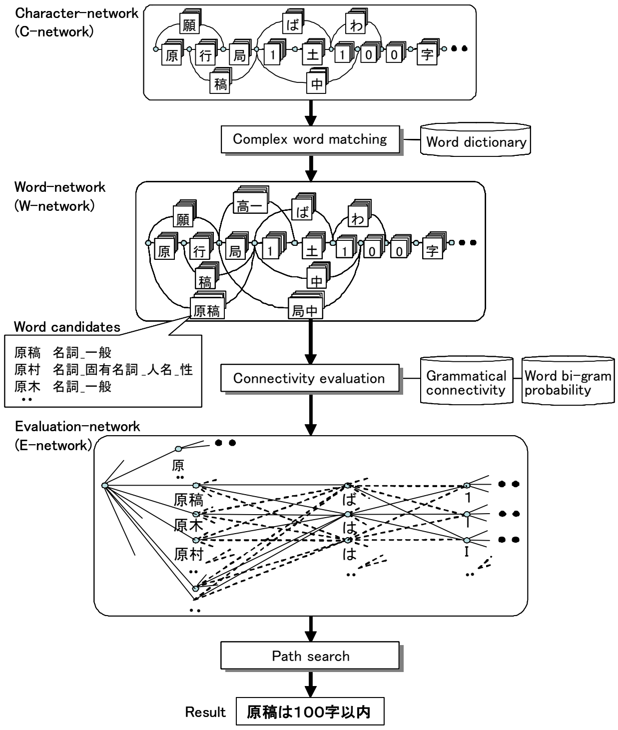


Figure 3. Outline of language processing.

4.2. Data for language processing

In making the word dictionary for the complex word matching and the grammatical connectivity data as shown in Figure 3, we used the morphological analysis dictionary “ipadic” version 2.7.0. This dictionary was developed in the laboratory of Prof. Matsumoto at the Nara Institute of Science and Technology. It includes about 240,000 entries and five related types of information, as follows.

(1) Morpheme dictionary
Entries, parts of speech (POSs), morpheme generation costs, etc.

(2) Parts of speech (e.g., verbs, adjectives)
POSs and the presence of conjugation.

(3) Conjugation type (e.g., 五段, カ変)
Types of POSs with conjugations.

(4) Conjugated form (e.g., 未然形, 連用形)
Forms of each conjugation type and their conjugative suffixes.

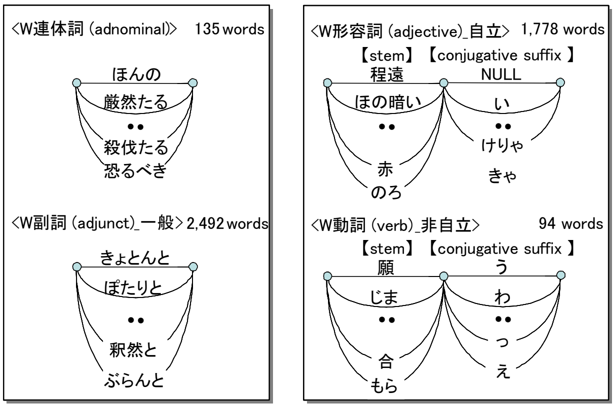
(5) Connection table with grammatical rules
Possibilities of connections between two conjugated forms.

Our word dictionary was made from the information of types (1), (2), (3), and (4). The grammatical connectivity data was made from the information of types (5). The word bi-gram data, as shown in Figure 3, was made from one year of articles of Mainichi Newspapers.

4.3. Complex word matching

Our method uses RTN-type word matching. Character candidates in a C-network are matched with a word dictionary. Each character candidate plays the role of an access key to the dictionary. The matching process is carried out at each node of a C-network to extract words at any node.

Our word dictionaries are of two kinds: with and without conjugation, as shown in Figure 4. In dictionaries with conjugation, conjugative suffixes of words with the same conjugated form are the common words. Therefore, conjugative suffixes of words that have the same conjugated form are matched only once. Thus, the processing time is reduced by avoiding redundant matching. A dictionary is made for each POS. There are about 70 dictionaries.



(a) Example of dictionary without conjugated form.

(b) Example of dictionary with conjugated form.

Figure 4. Structures of RTN-type word dictionaries.

The word matching is carried out under two conditions. First, the number of character candidates is restricted according to Equation 1. We set the threshold to 20. The value of the similarity of a classification is from 0 to 255. This means that only candidates that satisfy Equation 1 are matched. The second condition is that the number of extracted words at each node of a C-network using each word dictionary is restricted according to Equation 2. We set the threshold to 35.

$$\textit{1st simirality} - \textit{mth simirality} \leqq \alpha \qquad (1)$$

$$\textit{the number of extracted words} \leqq \beta \qquad (2)$$

4.4. Connectivity evaluation

In the connectivity evaluation, possible word pairs are selected, and their evaluation values are calculated. The result is outputs as an E-network. Input data (a W-network) and the processed result are shown in Figure 5. In particular, the lower part of the figure shows possible word pairs linked by solid lines and impossible word pairs linked by broken lines. The likelihoods of segmentations, classifications, grammatical connectivity, and word bi-gram are attached for each possible word pair.

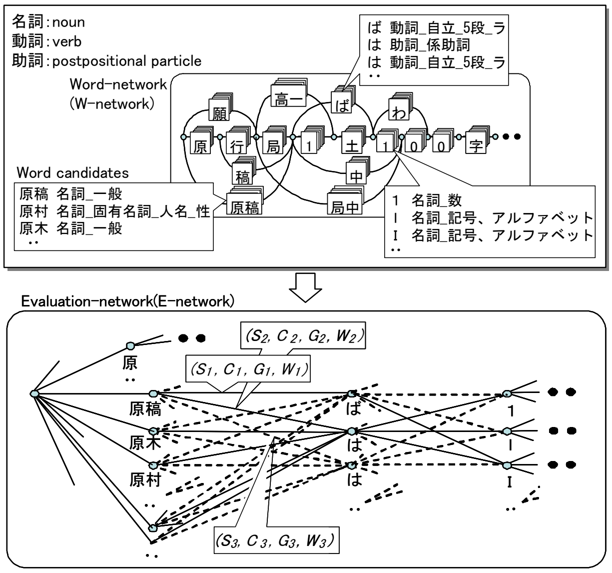


Figure 5. Evaluation of connectivity of words.

The likelihood of grammatical connectivity is calculated from the connective probability in the connection table by logarithmic transformation. The likelihood of a word bi-gram is calculated from the word bi-gram probability as described above. We used 24,576 grammatical rules and 2,010,016 word bi-gram items.

In making an E-network, the values of grammatical connectivity in the connection table must be accessed. The input data for this access is a word pair. If the value is not zero, the word pair is regarded as grammatically possible. If the value is zero, the word pair is regarded as impossible. Impossible word pairs are not used to make an E-network.

We now explain the preparation of an example E-network by using Figure 5. A group of front word of pairs are "原稿", "原木", "原村", etc. These are extracted from the same path. A front word "原稿" has "ば", "１", "中", etc. as group of back word of pairs. These are extracted from several different paths. For example, "ば", "は (verb, 自立５段ラ活用)", and "は (postpositional particle, 係助詞)" are part of a group of back word of pairs. They are extracted from the same path. Here, some different morphemes have as same notation. Because these words are different grammatically, this method treats them as different words. First, back words to connect to the front word "原稿" are searched. Second, pair words are made. Next, the value in the connection table is accessed. The possibility of the word pair is checked. In the case of "原稿" and "は (postpositional particle, 係助詞)", the probability of connectivity is not zero. Therefore, the link between "原稿" and "は (postpositional particle)" is stretched on the E-network. In the case of "原稿" and "は (verb, 自立５段ラ活用)", the probability of connectivity is zero. Therefore, the link between "原稿" and "は (verb)" is not stretched on the E-network. In this way, the E-network is made to stretch links among pair words referring the probability of connectivity. The evaluation value, E, attached to each possible word pair is calculated using in Equation 3.

$$E_{ij} = \sum L_1(W_i, W_j) + \sum L_2(W_i, W_j) + L_3(W_i, W_j) + L_4(W_i, W_j) \quad (3)$$

The term E_{ij} is the evaluation value for word$_i$ and word$_j$, L_1 is the summation of the likelihood of the pre-segmentation consisting of word$_i$ and word$_j$, L_2 is the summation of the likelihood of the classification consisting of word$_i$ and word$_j$, L_3 is the likelihood of the grammatical connectivity, and L_4 is the likelihood of the word bi-gram.

4.5. Path search

The most suitable paths are searched for in the E-network. We applied the Viterbi algorithm. This algorithm sums the evaluation values on the E-network and outputs the word sequence of the route with the highest value as the best interpretation.

5. Experiment and discussion

5.1. Experiment sample

We used 467 Japanese texts of about 10 characters each. The total number of characters was 4,384. These were entered by 98 people using Anoto Pens. Examples of entry images are shown in Figure 6.

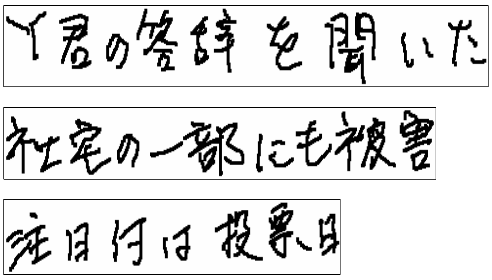


Figure 6. Examples of entry images.

5.2. Experimental results

We measured accuracies on texts and characters using six experiments to learn the effectiveness of our Japanese language processing.

(1) Basic performance: using pre-segmentation and classification.

(2) Basic performance and complex word matching: carried out word matching in addition to (1).

(3) Basic performance, complex word matching, and grammatical connectivity: considered grammatical connectivity in addition to (2).

(4) Basic performance, complex word matching, and word bi-gram: considered word bi-gram in addition to (2).

(5) Basic performance, complex word matching, grammatical connectivity, and word bi-gram: considered grammatical connectivity and word bi-gram in addition to (2).

(6) Performance of a conventional method: considered character bi-gram in addition to (1).

The experimental results are shown in Figure 7. The recognition rate per text using the basic performance (experiment (1)) was 10%. By adding language processing (experiment (5)), the recognition rate improved to 38%. The recognition rate of the conventional method (experiment (6)) was 22%. The recognition rate per character using the basic performance (experiment (1)) was 72.2%. The recognition rate when language processing was added (experiment (5)) improved to 82.1%. The recognition rate of the conventional method (experiment (6)) was 76.0%. Thus our method is effective.

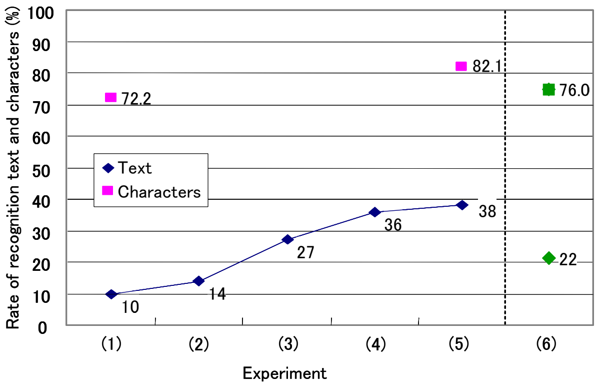


Figure 7. Rates of recognition of text and characters.

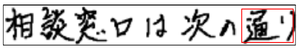
5.3. Discussion

5.3.1. Effectiveness of using complex word matching

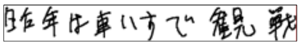
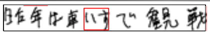
The recognition rate per text was 14%. The accuracy improved 4 percentage points by adding the complex word matching. An example of a result corrected by using the word matching is shown in Figure 8(a). The segmentation error, “副”, was corrected to the right word, “通り”, by using word information. The recognition rate of the conventional method is higher by 8 percentage points than that of the method using the word matching. This is because the conventional method partially includes information on the connectivity between words.

5.3.2. Effectiveness of using grammatical connectivity and word bi-gram

An example of a result corrected by using grammatical connectivity and word bi-gram is shown in Figure 8(b). In this case, two parts were misread. One is a character incorrectly classified as “牛”. The right character is “は”. The other error was two characters incorrectly segmented as “吋”. The right characters (word) are “いす”. The latter error was corrected to the right word by using the likelihood of the grammatical connectivity. However, the former error was not corrected this way. This is because the combination of "牛 (noun)" and "車 (noun)" is normal grammatically. Therefore, this error was corrected by using the likelihood of the word bi-gram.

Stroke data	
Experiment (1)	相談窓口は次の副
Experiment (2)	相談窓口は次の通り

(a) Effectiveness of word matching

Stroke data	
Experiment (1)	 昨年牛車吋で観戦
Experiment (3)	昨年牛車いすで観戦
Experiment (5)	昨年は車いすで観戦

(b) Effectiveness of grammatical connectivity and word bi-grams

Figure 8. Examples of reading results (1).

The recognition rate per text of experiment (4) is only 2 percentage points lower than that of experiment (5), and 9 percentage points higher than that of experiment (3). Therefore, using the likelihood of word bi-gram is more effective than using the likelihood of the grammatical connectivity.

5.3.3. Analysis of error factor

We analyzed the misread samples and found two main types of errors. One type is that the right word is not extracted as a word candidate. Two reasons account for this type of error. One reason is that the right character is not included in the character candidates. The other reason is that Equation (1) or (2) is not satisfied. The other type of error is that a wrong word candidate is selected as a better word, as shown in Figure 9, although the right word is extracted as a word candidate. In this case, the evaluation value of the misread word, "殿 (noun)", was higher than that of the right word, "震災 (noun)".

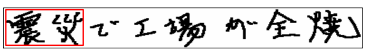
Stroke data	
Result	殿で工場が全焼

Figure 9. Example of reading result (2).

6. Conclusion

We proposed a new error correction method for reading unconstrained handwritten Japanese text. This method is applicable to offline and online recognition. The substantial problems are correcting segmentation errors and classification errors. Our method searches for the best interpretation by integrating the likelihoods of character segmentation, character classification and language processing with grammatical connectivity and word bi-gram. Experiments using 467 texts showed that our complete method is more accurate than any partial method. The recognition rate per text improved from 10 to 38%, and the recognition rate per character improved from 72.2 to 82.1%. The recognition rate of the conventional method was 22% on text and 76.0% on characters. Thus our method is effective.

References

[1] H. Murase, T. Wakahara and M. Umeda, "Online Writing-Box Free Character String Recognition by Candidate Character Lattice Method", Trans. of IEICE, Vol. J69-D, No. 9, pp. 1292-1301 (1986) (in Japanese).

[2] S. Senda, M. Hamanaka and K. Yamada, "Box-free Online Character Recognition Integrating Confidence Values of Segmentation, Recognition and Language Processing", Technical report of IEICE PRMU98-138, pp. 17-24 (1998) (in Japanese).

[3] T. Fukushima and M. Nakagawa, "On-line Writing Box Free Recognition of Handwritten Text based on Stochastic Models", Technical report of IEICE PRMU 98-139, pp. 25-30 (1998) (in Japanese).

[4] F. Nishino, "Natural Language Processing in Text Recognition", Vol. 34, No. 10, pp. 1274-1280 (1993) (in Japanese).

[5] T. Sugimura and T. Saito, "A Study of Reject Correction for Character Recognition Based on Binary n-Gram", Trans. of IEICE, Vol. J68-D, No. 1, pp. 64-71 (1985) (in Japanese).

[6] M. Hatada and H. Endoh, "Spelling Correction Method for English and Katakana in Japanese OCR text", Trans. of IPSJ, Vol. 38, No.7, pp. 1317-1327 (1997) (in Japanese).

[7] K. Takeuchi and Y. Matsumoto, "OCR Error Correction Using Stochastic Language Models", Trans. of IPSJ, Vol. 40, No. 6, pp. 2679-2689 (1999) (in Japanese).

[8] T. Takao and F. Nishino, "Implementation and Evaluation of Post-processing for Japanese Document Reader", Trans. of IPSJ, Vol.30, No.11, pp.1394-1401 (1989) (in Japanese).

[9] N. Itoh, "A Method of Post-processing for Online Japanese Character Recognition based on Bigrams", Technical report of IPSJ NP 97-6, pp. 37-44 (1993) (in Japanese).

[10] M. Nagata, "A Japanese OCR Error Correction Method Using Character Shape Similarity and Statistical Language Model", Trans. of IEICE, Vol. J81-D-II, No. 11, pp. 2624-2634 (1998) (in Japanese).

[11] T. Fukushima and M. Nakagawa, "On-line Writing-box-free Recognition of Handwritten Japanese Text Considering Character Size Variations", Proc. of ICPR2000 (2004).

[12] N. Furukawa, J. Tokuno and H, Ikeda, "Development of Character Segmentation Method for Recognition of Unconstrained Handwritten String", Proc. of FIT2005 (2005) (in Japanese).