A Discriminative Model for On-line Handwritten Japanese Text Retrieval

Cheng Cheng , Bilan Zhu, and Masaki Nakagawa
Dept. of Computer and Information Sciences
Tokyo University of Agriculture and Technology
2-24-16 Naka-cho, Koganei-shi 184-8588 Tokyo, Japan
50009834702@st.tuat.ac.jp {Zhubilan, nakagawa}@cc.tuat.ac.jp

Abstract— **This paper describes an unconstrained on-line handwritten Japanese text retrieval system from character recognition candidates. The system is based on a discriminative model which integrates the scores of character recognition, segmentation and geometric context in search and retrieval, and the parameters are trained by supervised learning. Experiments on TUAT Kuchibue database show that the proposed method can effectively improve the system performance. When the search method with the optimal threshold retrieves for a keyword consisting of two, three or four characters, its *f*-measure is 0.720, 0.868 or 0.923, respectively.**

Keywords- text retrieval; discriminative model; geometric context; character recognition

I. INTRODUCTION

Handwriting input interfaces has been employed in environments where a keyboard is not suitable. Portable devices like PDA, iPod, interactive electronic whiteboards and tablet PC's are examples of such environments where a keyboard is too large for mobile systems or it is not suitable for annotations.

With the development of the Internet, and the availability of pen input devices, the size of handwritten (pen input) data is increasing rapidly. Efficient handwritten searching, browsing and retrieval tools are required by users from various domains. For this purpose, many general purpose handwritten retrieval systems have been developed during the past decade.

According to the word similarity scoring technique, there are two frameworks used. With the first method, handwritten text is recognized with a handwriting word recognizer (HWR), and the results are used for search and retrieval of the text. Marukawa et al. proposed a method, which can reduce errors when searching from incorrect recognition results by using two or more character recognition candidates and a confusion matrix [1]. Ota et al. extended the above idea by generating search terms considering mis-segmentation and mis-recognition with corresponding confidence and bi-gram probabilities [2]. Imagawa et al. investigated the reliability of recognition results with a neural network and improved both the recall and precision rates [3]. Lopresti et al. examined how OCR noises affect the performance of common information search models [4]. Cao et al describes a keyword spotting method by modeling imperfect word segmentation as probabilities and integrating these probabilities into the word spotting algorithm [5]. Liu et al presents a text query-based method for keyword spotting from online Chinese handwritten documents [6]. The similarity between a text word and handwriting is obtained by combining the character similarity scores given by a character classifier.

However, some handwritten databases such as George Washington's handwritten papers [7] and old Greek Christian manuscripts [8] are too degraded to utilizing a handwriting word recognizer. Thus, the anther method is used. The method evaluates the similarity of an input keyword and handwritten text. During retrieval, the input keyword is converted to a handwriting pattern, and then the features are extracted. The similarity between the input keyword and handwritten text is measured as a distance between the two feature vectors. Manmatha et al. computed the similarity by dynamic time warping (DTW) technique [7]. Konidaris et al. proposed a technique for keyword spotting in Christian manuscripts [8]. The aim is to search for typed keywords in a large collection of digitized historical printed documents in which the retrieval result is optimized by user feedback. Zhang et al present an effective and efficient approach for word image matching by using gradient-based binary features [9]. The disadvantage of the method is that it requires on-line matching which is relatively time consuming.

We have implemented a keyword retrieval system for on-line handwritten Japanese documents using a handwriting word recognizer (HWR) in our previous work [10] where the similarity evaluation between the input keyword and handwritten text is obtained by combining the character similarity scores given by a character classifier.

In this paper, we propose a discriminative model to improve similarity evaluation between the input keyword and handwritten text, being inspired by the Conditional Random Field model (CRF) [11]. The similarity evaluation is computed using the scores of character recognition, segmentation, and geometric context with weighting parameters. The parameters are trained by the discriminative model which uses the results of HWR for search and retrieval.

Section 2 summarizes the retrieval system. Section 3 describes the evaluation method for retrieval results. Section 4 describes the parameter learning method. Section 5 details the experiments, and section 6 draws the conclusion.

1520-5363/11 $26.00 © 2011 IEEE
DOI 10.1109/ICDAR.2011.306

II. Overview of a Retrieval System

Our keyword retrieval system for on-line handwritten Japanese text using a handwriting word recognizer (HWR). First, on-line handwritten Japanese text is over-segmented into primitive segments according to the features such as spatial information between adjacent strokes. Then one or more consecutive primitive segments form a candidate character pattern, and each pattern is associated with several candidate classes with scores assigned by character classification. The combination of all candidate patterns and character classes is represented by a candidate lattice. Last, the system searches the candidate lattice, and obtains several retrieval results that match with the input keyword. The retrieval results are evaluated by character recognition, segmentation and geometric context. In order to increase the precision of its output, we prune results whose evaluated scores are below a threshold score ***Ts***.

III. Evaluation for Retrieval Results

Given a keyword, the system searches the candidate lattice to find stroke sequences matching with the input keyword where a stroke is a sequence of pen-tip coordinates from pen-down to pen-up. The score between the input keyword and each candidate is evaluated by

$$f(S, Keyword) = \sum_{i=1}^{6} \lambda_i f_i \qquad (1)$$

where S is a sequence of strokes matching with the keyword, f_j is the feature function for the score of character recognition, segmentation and geometric context, and λ_j is the weighting parameter.

The six feature functions in Eq. (1) depict the characteristic of character shape and geometric context, and the details are shown in Table 1.

TABLE 1. Summary of feature functions

$f_1 = \sum_{i=1}^{n-1} D_1(s_i, s_{i+1}, c_i, c_{i+1})$	$f_2 = \sum_{i=1}^{n} HW(s_i, c_i)$
$f_3 = \sum_{i=1}^{n} Inner(s_i, c_i)$	$f_4 = \sum_{i=1}^{n} D_2(s_i, c_i)$
$f_5 = \sum_{i=1}^{n-1} Seg(s_i, s_{i+1})$	$f_6 = \sum_{i=1}^{n} Rec(s_i, c_i)$

In table 1, ***n*** is the length of the input keyword, c_i is the *i-th* character of the input keyword; s_i is the *i-th* character patterns matching with c_i.

The term $D_1(s_i, s_{i+1}, c_i, c_{i+1})$ is a QDF function for a binary geometric feature vector p^b_i and characters c_i, c_{i+1}. The binary geometric feature vector p^b_i has two elements measured from the bounding boxes of two adjacent character patterns s_i,s_{i+1}: the vertical distances between the upper bounds and between the lower bounds as shown in Fig.1.

The term $HW(s_i, c_i)$ is a QDF function for a geometric feature vector and characters c_i. The geometric feature vector comprises the height and width of bounding box of character pattern s_i.

The term $Inner(s_i, c_i)$ is a QDF function for a geometric feature vector q_i and characters c_i. The geometric feature vector q_i comprises six values as shown in Fig.2 where ***acs*** is the average character size. The first three values represent the horizontal gaps of three vertical slits (partitioned from vertical projection), and the last three ones represent the vertical gaps of three horizontal slits (from horizontal projection).

The term $D_2(s_i, c_i)$ is a QDF function for a geometric feature vector p^u_i and characters c_i. The geometric feature vector p^u_i comprises two elements, the first element represents the length from the top to the center line of the text line and the second element represents that from the bottom to the line as shown in Fig.1.

The values of geometric features are normalized with respect to the average character size *acs* for scale invariance.

The term $Seg(s_i, s_{i+1})$ is a function of a SVM classifier which measures the plausibility of segmentation between two adjacent candidate character patterns s_i, s_{i+1} [12].

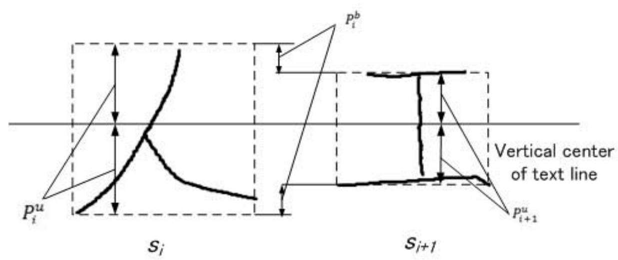

Fig. 1. Some geometric features.

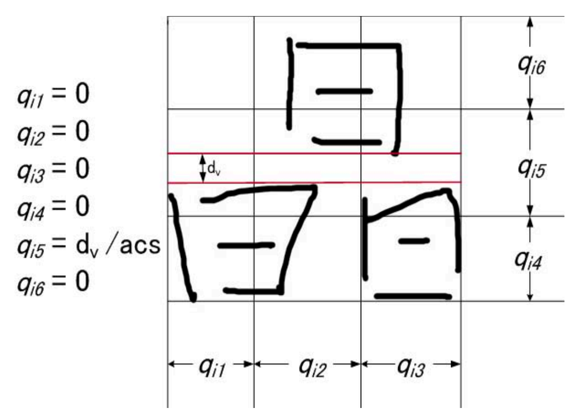

Fig. 2. Feature values of character pattern inner gap.

IV. Parameter Estimation

We are inspired with Conditional Random Field to construct a parameter estimation method for evaluating the retrieval results as follows:

Given the training data $D=\{x^i, y^i\}_{i=1}^N$, wher N is the number of input keywords, $x^i=\{x_1^i, x_2^i, \ldots x_T^i, x_{T+1}^i, x_{T+2}^i, \ldots x_S^i\}$ denotes the sequence of retrieval results in a candidate lattice ($x_1^i \ldots x_T^i$ are correct, and the others are incorrect), and y^i is the input keyword, the conditional probability for correct retrieval is formulated as

$$P(x^i|y^i) \approx \frac{\exp\left[\sum_{j=1}^{T} f\left(x_j^i, y^i\right)\right]}{\sum_{j=1}^{S} \exp f\left(x_j^i, y^i\right)} \quad (2)$$

where $f(x_j^i, y^i)$ is defined in Eq. (1).

As many output candidates were incorrect, we pruned the stroke candidates in which the evaluated score were below a threshold score ***Ts***. In order to let the scores of correct stroke candidates become as high as possible and the scores of the incorrect stroke candidates become as low as possible, the loss function is defined as the summation of the negative log-likelihoods:

$$f(\theta) = -\sum_{i=1}^{N} logP(x^i|y^i)$$
$$= \sum_{i=1}^{N} log\left[\sum_{j=1}^{S} exp\left(f(x_j^i, y^i)\right)\right] - \sum_{i=1}^{N}\sum_{j=1}^{T} f(x_j^i, y^i) \quad (3)$$

The stochastic gradient decent algorithm is used to optimize the weighting parameters:

$$\lambda(t+1) = \lambda(t) - \varepsilon(t)\nabla f(\theta) \quad (4)$$

where $\varepsilon(t)$ is a learning step and $\nabla f(\theta)$ denotes the partial derivative with respect to the parameters:

$$\frac{\partial f(\theta)}{\partial \lambda_k} = \sum_{i=1}^{N} \frac{\sum_{j=1}^{S}\left[exp\left(f\left(x_j^i, y^j\right)\right) * f_k\right]}{\sum_{j=1}^{S} exp\left(f\left(x_j^i, y^j\right)\right)} - \sum_{i=1}^{N}\sum_{j=1}^{T} f_k \quad (5)$$

It is necessary to point out that the well-known CRFs infer and maximize the posteriori probability while the model proposed by us infers and maximizes the conditional probability as shown in Eq. (2).

V. Experiments and Evaluation

We evaluate the performance of the retrieval system with *f-measure*

$$f = \frac{2}{1/r + 1/p} \quad (6)$$

where ***r*** is recall and ***p*** is precision defined in Eq.(7) and Eq.(8), respectively. The recall rate measures the tolerance of the system to search errors, while the precision rate measures the tolerance to noises. The *f-measure* is an overall performance of the retrieval system.

$$r = \frac{\textit{Number of correct search}}{\textit{Number of search keywords in target data}} \quad (7)$$

$$p = \frac{\textit{Number of correct search}}{\textit{Number of searched items (include noise)}} \quad (8)$$

We employ the database "HANDS-Nakayosi_t-98-09" (in brief, Nakayosi)[14] to train our character classifier and geometric scoring functions.

We employ the database "HANDS-Kondate bf-2001-11" (in brief, Kondate) to train our SVM classifier for the candidate segmentation point probability.

We employ the database "HANDS-kuchibue_d-97-06" (in brief, Kuchibue) [14] in our experiments, by concatenating single characters as text lines. Each page has at most 10 lines and each line has at most 10 characters with the gap between adjacent characters randomly ranging from 1 to 10 pixels. The data of 60 writers are used for training parameters and the rest 60 for testing.

We compared the retrieval performance of three parameter learning methods with the same environment. Table 2 lists the results of our previous work [10] in which $\lambda_{1\sim5}=0$ and $\lambda_6=1$, . Table 3 lists the results of the parameter learning method proposed by *Zhu et al* [15] for Japanese text recognition that trains the weighting parameters by the genetic algorithm to optimize the recognition rate. Table 4 lists the results of the system proposed in this paper.

From the results, we can see that our proposed model has brought the best retrieval performance. Introducing the adjustable weighting parameters to evaluate the similarity between the input keyword and handwritten text has improved the retrieval accuracy; the method for estimating the weighting parameters to optimize the conditional probability for correct retrieval has brought better retrieval accuracy than the method in [15] that estimates the weighting parameters to optimize the recognition rate.

TABLE 2. Results of our previous work [10]

Length	*Recall*	*Precision*	*f-measure*
2	71.8%	50.2%	0.582
3	90.0 %	71.3%	0.791
4	89.8%	87.0%	0.883

TABLE 3. Using the parameters of character string recognition [15]

Length	*Recall*	*Precision*	*f-measure*
2	84.2%	63.6%	0.719
3	89.6%	84.6%	0.866
4	91.9%	92.6%	0.920

TABLE 4. Results of proposed system

Length	*Recall*	*Precision*	*f-measure*
2	84.2%	63.8%	0.720
3	90.9%	83.2%	0.868
4	91.9%	93.1%	0.923

VI. Conclusion

In this paper, we described a discriminative model for unconstrained handwritten Japanese text retrieval inspired by the Conditional Random Field model, which incorporates the scores of character recognition, segmentation and geometric context. Experiments on Kuchibue database demonstrate the effectiveness of our proposed method

Acknowledgment

This work is being partially supported by the R&D fund for "development of pen & paper based user interaction" under Japan Science and Technology Agency.

References

[1] K. Marukawa, H. Fujisawa, Y.Shima. "Evaluation of Information Retrieval Methods with Output of Character Recognition Based on Characteristic of Recognition Error (Japanese)," Trans IEICE, vol.J79-D-II, no.5, pp.785-794, 1996.

[2] M. Ota, A. Takasu, J. Adachi. "Full-text search Methods for OCR-recognized Japanese Text with Misrecognized Characters (Japanese)," Trans. IPSJ, vol.39, no.3, pp.625-635, 1998.

[3] T. Imagawa, Y. Matsukawa, K. Kondo, T. Mekata. "A document image retrieval technique using each character recognition reliability (Japanese)," Technical report of IEICE PRMU99-72, pp.63-68, 1999.

[4] D. Lopresti and J. Zhou, "Retrieval Strategies for Noisy Text," Proc. 15th Annual Symposium on Document Analysis and Information Retrieval, pp. 255-269, 1996.

[5] H. Cao, et al, "A probabilistic method for keyword retrieval in handwritten document images", Pattern Recognition, pp42 (12):3374-3382, 2009.

[6] H. Zhang, Da-Han Wang, Cheng-Lin Liu. "Keyword Spotting from Online Chinese Handwritten Documents Using One-Vs-All Trained Character Classifier" Proc. 12th ICFHR, pp.271-276, 2010

[7] R. Rath and R. Manmatha. "Word spotting for historical documents".IJDAR, pp 9(2):139–152, 2007

[8] T. Konidaris · B. Gatos · K. Ntzios · I. Pratikakis · S. Theodoridis · S. J. Perantonis, "Keyword-guided word spotting in historical printed documents using synthetic data and user feedback" IJDAR, pp9 (2), 167–177, 2007.

[9] B. Zhang, S.N. Srihari, C. Huang, Word image retrieval using binary features, Document Recognition and Retrieval XI, vol. 5296, SPIE, Greenbelt, MD, pp. 45–53, 2004.

[10] C. Cheng. B. Zhu and M. Nakagawa." Improvements in keyword search Japanese Characters within handwritten digital ink" Proc. 10th ICDAR, pp 863-866, 2009.

[11] J. Lafferty, A. McCallum, F. Pereira, Conditional random fields: probabilistic models for segmenting and labeling sequence data, Proc 18th ICML, pp.282-289, 2001.

[12] B. Zhu and M. Nakagawa, "Segmentation of on-line freely written Japanese text using SVM for improving text recognition," IEICE Trans. Inf. & Sys., E91-D(1), pp.105-113, 2008.

[13] H. Oda, B. Zhu, J. Tokuno , M. Onuma , A. Kitadai. and M. Nakagawa, "A compact on-line and off-line combined recognizer," Proc.10th IWFHR, pp.133-138, 2006M.

[14] M. Nakagawa and K. Matsumoto, "Collection of on-line handwritten Japanese character pattern databases and their analysis", IJDAR, pp.69-81, 2004.

[15] B. Zhu, M. Nakagawa., "On-line handwritten Japanese text recognition by improving segmentation quality". Proc. 11th IWFHR, pp. 379–384, 2008