Segmentation-free recognizer based on enhanced four plane feature for realistic Chinese handwriting

Tong-Hua Su[a], Tian-Wen Zhang[a], Hu-Jie Huang[a], and Cheng-Lin Liu[b]
[a]School of Computer Science and Technology, Harbin Institute of Technology
[b]National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences
{tonghuasu, twzhang}@hit.edu.cn, liucl@nlpr.ia.ac.cn

Abstract

Directional features are preferred in off-line Chinese character recognition due to their superior performance. This paper proposes an enhanced four plane feature (en-FPF) within a segmentation-free recognition framework. First, the directional planes are strengthened by replenishing salient pixels. Second, the method to count perpendicular strokes are renewed. In experiments of realistic Chinese handwriting recognition, the proposed enhancement yields desirable improvements of recognition rates, especially to punctuation marks and digits. Compared with four-orientation Gradient feature and Gabor feature, the superiority of en-FPF is also observed.

1. Introduction

The off-line recognition of realistic Chinese handwriting seems one of the most challenging problems. Mostly, it is due to the great writing variability in real-life Chinese handwritten documents. Generally, there exist three hierarchies of complexities. At the document level, there are overlapping, touching, and crossing between adjacent textlines. Moreover, at the textline level, besides undulation and skewness of textline, overlapping, touching, and crossing among character neighbors may also present. In addition, at the character level, the variable size of the characters, deformation of strokes and even erasure of characters may raise.

The state-of-the-art methods can be dichotomized into segmentation-based and segmentation-free systems [1]. In the former, a character segmentation stage attempts to separate each character first, then the isolated handwritten Chinese character recognition (HCCR) is triggered to identify the character block [2]. In literature, feature extraction methods have been intensively studied and among them, directional ones (e.g. Gradient feature [3], Gabor feature [4]) have demonstrated steady superior performance when embedded into different classifies [5].

The latter incorporates the character segmentation and recognition in one step, and optimizes the process with a kind of expectation maximization (EM) algorithms [1][6]. In [7][8], two typical systems for English textline recognition are described. Both of them intend to improve their systems using statistical language models. Recently, we have developed a segmentation-free recognizer to transcribe realistic Chinese handwriting. The study stresses on the feature extraction and representation techniques. The results suggest that directional features such as four plane feature (FPF) should be integrated as part of the final feature. This paper explores the FPF further. After identifying its weakness, we construct an enhanced FPF (en-FPF) through strengthening directional planes in an effective and low-cost manner. The experiments are conducted on HIT-MW database [9][10] and the results verify the discriminative ability of the en-FPF.

The paper is organized as follows. In the Sect. 2, the segmentation-free recognizer is briefly summarized. The enhanced feature extraction method is detailed in Sect. 3. Experiments and concluding remarks are provided in Sect. 4 and 5, respectively.

2 System Description

The architecture of the segmentation-free recognizer is shown in Fig. 1. When a textline image is fed as input, it is converted to a sequence of feature vectors (or observations) $\mathbf{O} = \mathbf{o}_1, ..., \mathbf{o}_m$ using the sliding window based feature extraction method. Sliding window is used to draw an interested zone, following the writing

978-1-4244-2175-6/08/$25.00 ©2008 IEEE

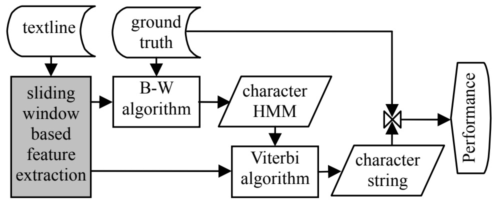


Figure 1. System architecture.

orientation. At each step, the zone is mapped from the image space to the feature space. Generally, the height of the sliding window is the same as that of textline. The other two parameters, the width W and the shift step S, should be tuned carefully. Supposing f_i is the window at the i-th shift, the feature vector $\mathbf{o}_i$, in observation sequence $\mathbf{O} = \mathbf{o}_1, ..., \mathbf{o}_m$ is calculated as follows:

$$\mathbf{o}_i = \psi(f_i), \tag{1}$$

where ψ is the feature mapping function. To resist the undulation of textlines, only the body zone instead of the whole window is used when extracting directional features. The body zone is separated by the topmost and bottommost foreground pixels in vertical direction.

The underlying character string $\hat{S} = s_1, ..., s_n$ is identified by maximizing the a posteriori (MAP) $P(S|\mathbf{O})$, once the character HMMs have been generated by embedded Baum-Welch algorithm (B-W algorithm) on training and validation sets. There is no attempt to segment the textline into characters, though soft segmentation is delivered as a byproduct in the recognition phase. On the other hand, the ground truth database is the reference transcription of textline database, and it is used not only in the training stage but also in the performance evaluation stage. More details are available in [1].

3 Enhanced Four Plane Feature

FPF is originally used in the isolated HCCR [11]. It generates directional planes using stroke runlength analysis and unfortunately, small strokes (especially to punctuation marks and digits) and salient pixels on the flange of strokes are sometimes discarded. We enhance the planes by integrating most of above strokes and pixels. Supposing the foreground pixel is “1” and background one “0”, the process works as follows:

Step 1 Generate the initial directional planes. Each textline image (it is denoted as OT) is scanned in four directions (horizontal, vertical, right diagonal, and left diagonal) and only the strokes are retained which are longer than a threshold $c \times SW$ (SW denotes the average stroke width and c is a constant). SW is estimated on the whole textline by an analysis of stroke histogram (as in [12]). The four planes are marked as HT, VT, RT and LT, respectively. Currently, the parameter c is set to 1.5.

Step 2 Enhance above planes. Planes in Step 1 comprise line elements. Compared with the actual strokes, certain portion may be unreasonably discarded, especially in the stroke ends. Such finding motivates us to add the discarded portions upon the initial planes:

$$VT \leftarrow VT \cup (OT \cap \sim (HT \cup RT \cup LT)), \tag{2}$$

where $\cap$, $\cup$ and $\sim$ are logic AND, OR and NOT operators, respectively. The planes in the right side are the initial directional planes. HT, RT and LT can be updated similarly. The evolution of VT is exemplified in Fig. 2 (added pixels are rendered with blue color). Obviously, the following equation is true:

$$HT \cup VT \cup RT \cup LT = OT. \tag{3}$$

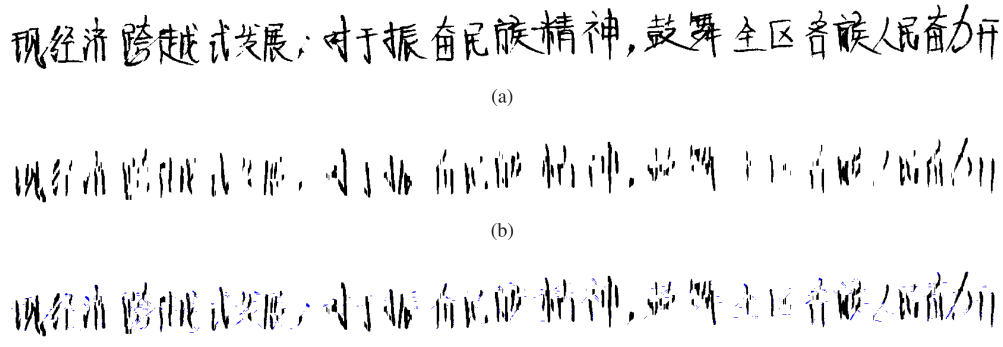
This means that the original image can be completely reconstructed by the four planes.

Step 3 Partition sliding window at each plane into smaller cells. Currently, 8×2 uniform cells are drawn per window.

Step 4 Extract directional features. To each plane, the scanning lines are perpendicular to the stroke direction. For instance, as regards the vertical plane, the scanning line runs horizontally. Previously, the stroke count in each cell is increased by one every transition from foreground to background. Herein, both transient from foreground and that into foreground can invoke the addition. Eventually, all counts ($8 \times 2 \times 4$) in current window are concatenated together as the feature vector.

4 Experiments

The benchmark data used in this paper come from the HIT-MW database [9][10]. It is collected from more than 780 participants with a systematic way. The HIT-MW database can be seen as a representative subset of real Chinese handwriting (refer [10] for more details). Currently, the HIT-MW database provides 5,667 textlines which can be used freely. These textlines are partitioned into training set (953 textlines), validation set (189 textlines), and test set (383 textlines) according to random sampling theory (we recommend readers to see [1]) (the HIT-MW database and the experimental data in this paper are available at: `http://hitmwdb.googlepages.com`).



(a) (b) (c)

Figure 2. Image of (a) the original textline, (b) the initial VT, (c) the enhanced VT.

In this paper, the training set and validation set are used to design the HMM-based recognizer. The validation set here is used to tune the parameters, such as the number of Gaussian components in mixture density and training iterations, which will be used in the final recognizer.

The output of certain recognizer is compared with the reference transcription and two metrics, the correct rate (CR) and accurate rate (AR), are calculated to evaluate the results. Supposing the number of substitution errors (S_e), deletion errors (D_e), and insertion errors (I_e) are known, CR and AR are defined respectively as:

$$\begin{cases} \mathrm{CR} = (N_t - D_e - S_e)/N_t \\ \mathrm{AR} = (N_t - D_e - S_e - I_e)/N_t \end{cases}, \tag{4}$$

where N_t is the total characters in the reference transcription. In general, CR is nonnegative while AR may be negative if there are overmany insertion errors.

The recognition rates of FPF and en-FPF are summarized as the first two rows in Table 1. The higher rates under each character type are highlighted with bold font. Since there are merely 8 instances of English letters in test set, their results are not considered in this paper. Note: 1) the FPF has different value of c than the paper [1]; 2) during the recognition stage, a punishment is given proportional to the length of the output string. It can be seen that the enhancement of directional plane averagely increase the CR and AR with 2.73 percent and 1.11 percent, respectively. Among digits, punctuation marks and Chinese characters, the performance of punctuation marks is the most beneficial one. More than 10 percent of improvement in CR and AR is obtained. Another noticeable improvement lies in the CR of digit. The reason underlying above superiority is that the Step 3 of en-FPF integrates the small strokes of punctuation mark and digit which are often discarded by FPF (as in Step 2). The smallest improvement to FPF is found in AR of Chinese character. To explore its confidence, we first calculate the AR of each textline by FPF and en-FPF, respectively. As a result, 383 pairs of ARs are obtained. Then we can conduct Wilcoxon signed-rank test [13]. Using above method, their statistical significance holds at 0.0001 level.

We also investigate the performance of Gabor feature and Gradient feature. These two features are preferred in the isolated HCCR. In our previous work [6], Gabor feature is successfully adapted into segmentation-free framework and best recognition rates (both CR and AR) are achieved among three examined features when tested on 200 handwritten textlines produced by a single writer. If we perform Sobel operator on a whole textline instead of a character, the Gradient feature can also be used in segmentation-free framework. We first decompose the gradient vectors into 8 standard directions. Next, we merge each pair of directional planes of opposite directions into orientation plane to give an identical length of feature vector. In addition, an 11 by 11 Gaussian filter (with the standard deviation of 3.6) is used to reduce its dependence on the stroke position.

The results of Gabor feature and Gradient feature are also shown in Table 1 (the last two rows). However, their performance is inferior to both FPF and en-FPF. As for Gradient feature, our experiments on the isolated HCCR have show that it is much dependent on the shape normalization process. Unfortunately, no shape normalization has been applied in segmentation-free framework yet. As for Gabor feature, it is more sensitive to the variety in stroke width. In multiple-writer handwriting, stroke widths are frequently varied though uniformly distributed in single-writer case.

The number of training samples has a significant effect on the final recognition rates. We plot the average

Table 1. The comparisons of recognition rates among several systems (%).

	Digit		Punctuation		Chinese character		Average	
	CR	AR	CR	AR	CR	AR	CR	AR
FPF	42.17	37.83	32.32	28.66	35.97	32.98	35.79	32.71
en-FPF	**51.30**	**39.13**	**44.95**	**38.76**	**37.45**	**33.13**	**38.52**	**33.82**
Gradient	38.26	29.13	39.39	27.40	33.80	29.52	34.45	29.31
Gabor	27.39	25.65	31.19	24.87	33.87	29.40	33.44	28.87

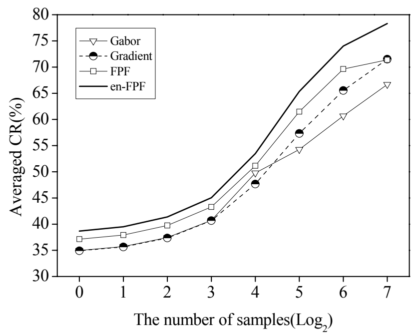


Figure 3. Relationship between the number of samples and CR.

CRs of Chinese characters using above features versus the sample size of the character class in Fig. 3. It is obvious that en-FPF is superior to others.

5 Conclusion

This paper presents an enhanced four plane feature (en-FPF) within a segmentation-free recognition framework. The original FPF is generated using runlength histogram analysis, whereby small strokes and salient pixels on the verge of strokes are often discarded. We hence strengthen the directional planes by replenishing salient pixels. Also, the method to count perpendicular strokes are renewed to improve the robustness. Experiments are conducted on a realistic Chinese handwriting database and the proposed enhancement yields desirable improvements of recognition rates, especially to punctuation marks and digits. The en-FPF is also compared with four-orientation Gradient feature and Gabor feature and shows obvious superiority.

References

[1] T.-H. Su, T.-W. Zhang, H.-J. Huang, and Y. Zhou. HMM-based recognizer with segmentation-free strategy for unconstrained Chinese handwritten text. *Proceedings of the 9th ICDAR*, pages 133–137, 2007.

[2] S. Srihari, X. Yang, and G. Ball. Offline Chinese handwriting recognition: an assessment of current technology. *Frontiers of Computer Science in China*, 1(2): 137–155, 2007.

[3] C.-L. Liu. Normalization-cooperated gradient feature extraction for handwritten character recognition. *IEEE Trans PAMI*, 29(8): 1465–1469, 2007.

[4] X. Wang, X. Ding, and C. Liu. Gabor filters-based feature extraction for character recognition. *Pattern Recognition*, 38(3): 369–379, 2005.

[5] C.-L. Liu, K. Nakashima, H. Sako, and H. Fujisawa. Handwritten digit recognition: benchmarking of state of the art techniques. *Pattern Recognition*, 36(10): 2271–2285, 2003.

[6] T.-H. Su, T.-W. Zhang, D.-J. Guan, and H.-J. Huang. Gabor-based recognizer for Chinese handwriting from segmentation–free strategy. *Proceedings of the 12th International Conference on Computer Analysis of Images and Patterns*, pages 539–546, 2007.

[7] U.V. Marti, and H. Bunke. Using a statistical language model to improve the performance of an HMM-based cursive handwriting recognition system. *International Journal of Pattern Recognition and Artificial Intelligence*, 15(1): 65–90, 2001.

[8] A. Vinciarelli, S. Bengio, and H. Bunke. Offline recognition of unconstrained handwritten texts using HMMs and statistical language models. *IEEE Trans PAMI*, 26(6): 709–720, 2004.

[9] T.-H. Su, T.-W. Zhang, and D.-J. Guan. HIT-MW dataset for offline Chinese handwritten text recognition. *Proceedings of the 10th IWFHR*, 2006.

[10] T.-H. Su, T.-W. Zhang, and D.-J. Guan. Corpus-based HIT-MW database for offline recognition of general-purpose Chinese handwritten text. *International Journal of Document Analysis and Recognition*, 10(1): 27–38, 2007.

[11] Y. Chen, X. Ding, and Y. Wu. Off-line handwritten Chinese character recognition based on crossing line feature. *Proceedings of the 4th ICDAR*, pages 206–210, 1997.

[12] T.-H. Su, T.-W. Zhang, H.-J. Huang, and Y. Zhou. Skew detection for Chinese handwriting by horizontal stroke histogram. *Proceedings of the 9th ICDAR*, pages 899–903, 2007.

[13] J.E. Freund. *Modern Elementary Statistics*. Prentice-Hall, New Jersey, 1984.