Gesture Recognition Techniques in Handwriting Recognition Application

Fengjun Guo, Shijie Chen
Motorola Software Solution, Shanghai
1257 Ming Yue Rd. Bldg 11 Pu Dong District, Shanghai, P.R. China 201206
Feng-JunGuo@Motorola.com

Abstract—Handwriting-gesture recognition has been widely implemented in handwriting input application. Usually, gestures are used to conduct edit operations or be set as short-cut of an application. In this paper, we compare several handwriting-gesture recognition methods, and address their different user cases. These methods include pixel-matching method, rule based method and discriminant-function based method. For discriminant-function based method, we describe 2 sub-methods. They are prototypes based method and training based method. We not only analyze recognition accuracy of gestures for these methods, but also analyze their distinguished capability when recognizing gestures and alphanumeric in same recognizing mode. Experiments results show that, if the gesture-samples are enough, training based method achieves the highest accuracy. Furthermore, when recognizing mixed input of gestures and other handwriting symbols, training based method almost doesn't degrade accuracy of these symbols.

Keywords-gesture recogntion; handwriting recognition; Rule-based method; Discriminant-function

I. INTRODUCTION

With touch-sensor is widely implemented in mobile and other electronic devices, many researchers pay attention to topic about handwriting gesture recognition. Gesture-recognition becomes a hot human-interactive research domain. In several popular electronic products, handwriting gestures based applications become selling points, i.e. the multi-touch gestures of iPhone.

In generally, handwriting-gesture recognition application can be divided into 2 categories. One is pre-defined gesture application, and the other is user-defined gesture application. For the 1st user-case (pre-defined gesture application), system provides a pre-defined gesture-list. Each gesture in the list has be defined a fixed function, i.e. editing operator or short-cut etc. For the 2nd user-case (user-defined gesture application), system allows users to define their own gestures. Then let user set functions of these gestures. Normally, system provides a graphic user-interface to let user define these gestures and their correspondent functions.

In this paper, we describe our recent works in handwriting-gesture recognition. In handwriting input method, an important function of gestures is to conduct editing operations. For example, insert a space, enter, or delete a character etc. For such kind of application, in China market, there are Chinese gesture standard, GB/T18790-2002. Figure 1 shows 5 editing gestures that defined in GB/T18790-2002. For each gesture in the figure 1, a black dot shows the start point of handwriting stroke.

In the paper, we describe following 3 methods to recognize gestures.

Method 1, system directly calculates similarity between handwriting pixels of input and stored pixel-templates of gestures. The pixel-template of a gesture is constructed based on the normalized and re-sampled stroke of a typical gesture-sample. Pixel-templates of all gestures have same dimension. Then a normalized correlation between input-pixels and pixel-templates is calculated as the similarity. This is method is utilized as the baseline method in the paper. We name it pixel-matching method. Because its theory is easy to be understood, we will not describe its detail techniques again in later section.

Method 2, we call it decision-tree based method. It extracts geometry features of input, and based on pre-defined rules to distinguish gestures. The method is fast and very effective for simple-shape gesture. But for complex shape gesture, its accuracy is not good enough.

Method 3, we name it as discriminant-function based method. It implements techniques that are used in handwriting recognition. In the method, there are 2 different sub-methods. They implement different methods to construct gesture-templates. We call them prototype based method and training based method. .Section 3 will describe them in detail.

This remainder paper is organized as follows 4 sections. Section Ⅱ introduces the rule based recognition method, which implements geometry feature, Section Ⅲ describes discriminant function based method, and Section Ⅳ shows experimental result. Finally, Section Ⅴ draws conclusions.

II. RULE BASED METHOD

In the method, its recognizer is based on several geometry features, i.e, the count of stroke-segmentations, the direction of these stroke-segmentations etc.

For the rule based method of this paper, we assume that gestures are one kind of polylines. So some complex gestures are not supported by this method.

In the method, corner finding algorithm is used to split the gesture stroke into primitives, such as lines. Then the

978-0-7695-4221-8/10 $26.00 © 2010 IEEE
DOI 10.1109/ICFHR.2010.29

feature of the primitives is extracted to extract geometry features. In this experiment, the count and the direction of lines (primitives) are used. For each gesture, system set a distinguished rule. If an input handwriting meets the setting rule, then it will be set as the correspondent gesture. Otherwise, it will be passed to following step to tell whether it is other gestures or not. Recognition procedure is to search candidate-gesture in a pre-defined logic tree. Figure 2 shows the workflow of the rule based method.

In figure 2, we can find that the rule based method has a reject scheme for non-gesture input. The reject scheme makes the method easy to be combined with handwriting recognition system to recognize gesture and other handwriting symbols in one recognition mode, and need not switching mode back and forth.

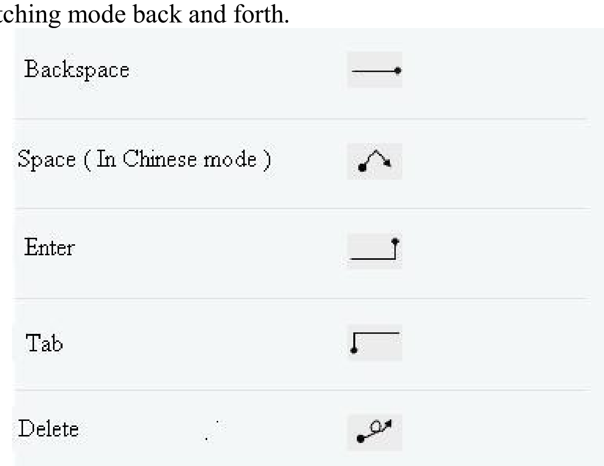


Figure 1. Handwriting gesture defined in GB/T18790-2002

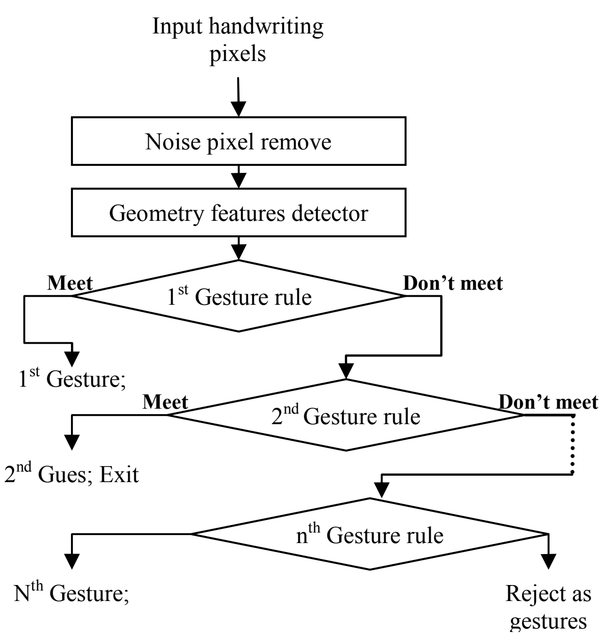


Figure 2. Workflow of rule based method

III. DISCRIMINANT-FUNCTION BASED METHOD

In the method, we implement the techniques that are successfully used in handwriting recognition.

Firstly, we use preprocessing method to remove variation of the same gesture. Normalization [5] and a re-sampling technique are useful to remove the variation of input samples.

For the feature extraction, we use several different features, e.g., directional feature and regional stroke count (RSC) feature vector.

To extract these features, the handwriting inks of an input sample are firstly converted to a binary image. The image is then normalized to compensate the shape distortion.

For the extraction of directional feature, the normalized image is uniformly divided into $n \times m$ grids. A feature vector is extracted from each grid, to characterize the image's local directions. For the whole image, we can derive a $n \times m$ dimensional directional feature vector. Readers are referred to [1] for the details.

The regional stroke count (RSC) feature vector was addressed in [2][6]. Its extraction-method is as following.

Let $\{f(x,y)|_{x,y=0}^{W-1,H-1}\}$ be the normalized image, where W and H are the image width and height respectively, $f(x,y)$ takes the value of either 1 or 0 based on whether current pixel locates in handwriting stroke or not. The 16-dimensional RSC feature vector is actually the combination of the 8-dimensional horizontal RSC feature vector (HRSC) and the 8-dimensional vertical RSC feature vector (VRSC). To extract the HRSC feature vector, the image is uniformly separated into 8 horizontal rectangle regions. From i-th horizontal rectangle, we compute i-th horizontal RSC (HRSC) feature Γ_{HRSC}^{i} by:

$$\Gamma_{HRSC}^{i} = \sum_{y=0}^{H-1} \sum_{x=(i-1)W/8}^{iW/8-1} f(x,y)\bar{f}(x+1,y), \quad i = 1,2,...8$$

Where $\bar{f}(x,y)$ is the logical negation of $f(x,y)$.

To extract the VRSC feature vector, the image is uniformly separated in to 8 vertical rectangle regions. From i-th vertical rectangle, i-th vertical RSC (VRSC) feature Γ_{VRSC}^{i} is computed by:

$$\Gamma_{VRSC}^{i} = \sum_{x=0}^{W-1} \sum_{y=(i-1)H/8}^{iH/8-1} f(x,y)\bar{f}(x,y+1), \quad i = 1,2,...8$$

In classifier selection, we just implement Euler distance to calculate the distance between input feature vector and stored feature template of each category. The Euler distance based classifier is used in 2 following methods too(Prototype based method and Training based method). It is a computation-saving method. Of curse, you can select more complex classifier to enhance the accuracy of classifier, i.e. the modified quadratic discriminant functions [7]. It will need more computation-resource.

Based on the different construction method of gesture-templates, we category the discriminant-function based into 2 sub-methods. The 1st method, we name it prototype based method. It extracts feature of few typical samples of a gesture, and set them as templates of the gesture. The 2nd method, we name it training based method. It implements training technique of handwriting-recognition to work out templates of gestures. We will describe them in detail in section A and section B of the section.

A. Prototype based method

The method just extracts features of few typical samples of a gesture, and set them as templates of the gesture. When conducting recognition, system calculates similarity (in the paper, we use distance) between features of input and stored templates of gestures.

The method is useful for user-defined gesture application. For the application, system doesn't know that user will define which kind of gesture. Normally, system provides a graphic user-interface, and require user to draw several samples of the gesture that he define. Then system could use these samples as templates of the gestures.

B. Training based method

Training based method implements training technique of handwriting-recognition to work out the templates of gestures. Based on these templates, system implements distance-classifier to recognize gestures.

In order to obtain better accuracy, in training procedure, we implement MCE training to enhance accuracy.

MCE training was first introduced by Katagiri and Juang[9]. More helpful works about MCE can be find in [3][4]. MCE training introduces a misclassification measure, and minimizes the measure to enhance accuracy. Normally, the misclassification measure is a empirical average cost as follows [9][10],

$$\min L(\Lambda) = \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{M} l(d_k(X_i);\Lambda) I(\mathrm{X}_i \in C_k)) \quad (1)$$

where M and N are the total number of classes and training data, and classes. X is a vector in feature space and $\Lambda = \{\lambda_k\}_{k=1}^{M}$ is parameter set of a discriminant function $g_i(X;\Lambda)$. Function $I(\cdots)$ is an indicator function which has value of one when the argument is true and zero otherwise.

In (1), function $d_k(\cdots)$ denote a misclassification measure function, and $l(\cdots)$ is loss function equation. They are respectively defined in (2) and (3). Loss function $l(\cdots)$ is a sigmoid function. Figure 3 shows the shape of the loss function.

$$d_k(X) = -g_k(X;\Lambda_k) + \max_{i \neq k} g_i(X;\Lambda_i)$$

(2)

$$l(d_k((X_i);\Lambda)) = \frac{1}{1+e^{(-\gamma d_k(X|\Lambda)+\theta)}}$$

(3)

In (2), $g_i(x \mid \Lambda)$ is discriminant function as follows,

$$g_i(X;\Lambda),\ i = 1,2,\cdots,M$$

(4)

In a recognition-task, if $g(X;\Lambda_s) = \arg\max_i g_i(x)$. Then class s is the recognition-result.

In (3), θ is used to set the shift of sigmoid function. It is usually set to 0. γ is related to the slope of the sigmoid function. For different data distributions, different values of γ should be selected. Normally set to $\gamma \geq 1$. In the task of this paper, we set γ as a value from 10 to 100.

In this paper, we use Euler distance between class-template and input feature as discrimnant function $g_i(X;\Lambda)$.

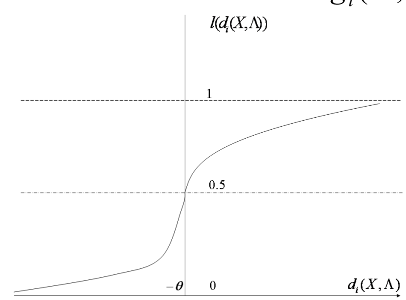


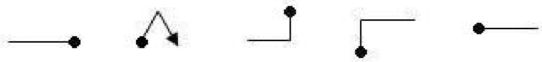
Figure 3. Shape of loss function

The training based method is very suitable for pre-defined gestures application. In the case, we could collect enough training handwriting gesture-samples to training a template for each gesture. Especially, the training based method achieves very good accuracy for complex shape gestures.

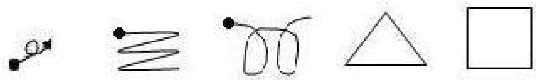
IV. EXPERIMENTS

To evaluate different gesture recognition methods mentioned above, we conduct serial experiments. We will not only analyze their recognition accuracy for gestures, but also will compare their distinguished capability when using them to recognize gestures and alphanumeric without switching recognizing mode. We call the recognition mode as mixed recognition mode. The accuracy in the mixed recognition mode is very important for handwriting recognition application. Users always wish to draw gestures without switching to a special gesture model. Mixed input gestures and alphabet without switching mode could enhance the usability of a handwriting input method.

For gesture database, we asked more than 15 peoples to write handwriting data of 10 gestures shown in figure 4. These gestures covers both simple shape gestures (Listed in (1) of figure 4) and complex shape gestures (Listed in (2) of figure 4). We split these gesture data into training and testing dataset. In training set, each gesture has more than 150 samples, and each gesture in testing set has more than 40 samples.



(1) Simple shape gestures



(2) Complex shape gestures

Figure 4. Experiment used gestures

A. *Just recognizing gestures*

In the section, we will compare the accuracy of above methods when they are used to just recognize gestures

In the 1st experiment, we just compare accuracies of 5 simple gestures listed in (1) of figure 3. The reason is that the rule-based method is difficult to recognize gestures with complex geometry shape, i.e, loops. For 5 simple shape gestures listed in (1) of figure 3, they are composed of one or several stroke-segments. Their decision-rules can be well defined.

The results of the 1st experiment are shown in the 2nd row of table 1. In order to easily describe experiment results, we name above mentioned methods as follows,

Method Pixel -- Stand for "Pixel matching based method"

Method Rule -- Stand for "Rule based method"

Method Proto -- Stand for "Prototype based method"

Method Comb -- Stand for "Combine rule based method (to recognize simple shape gestures) with prototyped based method (to recognize complex shape gestures)"

Method Training -- Stand for "Training based method"

The 2nd row of table 1 shows the experiment for 5 simple-shape gestures. Result shows that training based method achieves the highest accuracy. The accuracy of rule based method gets the lowest accuracy.

We conduct the 2nd experiment to compares the accuracy of different method when we use them to recognize all gestures in figure 3.These gestures include simple shape gestures and complex shape ones. In order to do that, we design a combination-method. The combination-method implements rule based method to recognize 5 gestures with simple shape, and uses method proto to recognize other 5 gestures with complex shape. We name the method as Method Comb.

The 3rd row of table 1 shows the accuracies comparison when we use different to recognize 10 gestures in figure 3. The results show that method training gets the highest accuracy too.

B. *Recognizing gestures and alphanumeric in a mixed recognition mode*

In the section, we will compare the distinguish-capability of different methods in a mixed recognition mode. As mentioned above, in the mixed recognition mode, system will recognize gestures and alphanumeric without switching recognizing mode.

Because pixel-matching based method is difficult to combine with handwriting recognizer, it can't be implemented in the mixed recognition mode. We don't compare its result.

In following experiments, the test set of digit and English alphabet are written by more than 1000 different writers. The test set of digit handwriting covers more than 35,000 samples. The test set of English alphabet covers more than 80,000 samples.

In tables 2, we compare the accuracy of gestures in the mixed recognition mode. It shows that the training based method could get the best accuracy.

In tables 3, we compare the accuracy of alphanumeric in mixed gesture and handwriting recognizer mode and just handwriting recognizer mode.

In the table 3, we can find that the accuracies of digit and English alphabet are about 90%. The reason is that there are several confused pairs among digit, lowercase alphabet, and uppercase alphabet. For example, digit '0', letter 'o' and 'O' can not be distinguished. Digit '1' and letter 'l' can not be distinguished. If remove such kind of false-recognition, the accuracy will much higher. The digit accuracy is higher than 99%, and the accuracy of English alphabet is higher than 97%.

Table 3 shows, in mixed recognition mode, training based method just degrade little bit accuracy for alphanumeric.

V. CONCLUSION

In the paper, we describe and compare several gesture recognition methods.

Among them, pixel-matching based method and rule based method need not update the database of handwriting recognition engine. But their recognition rate is not very high. Rule based method need provide a rule for each gesture. For gestures with complex geometry shape, i.e. loops etc, the precise rule is very difficult to be worked out. So the rule based method is useful for the gestures with simple-shape, but it is not fit for gestures with complex shape. The advantage of rule based method has reject scheme for un-gesture handwriting input. It is easily combined with handwriting recognition system to realize mixed handwriting input for gesture and other symbols. For pixel-matching method, it is difficult to design a reject-scheme for un-gesture input. So it is not easy to be implemented in mixed gestures and handwriting symbols input application.

Discriminant-function based method achieves higher accuracy for gestures. Especially, they can recognize gestures with complex-shape. But they need update database of handwriting recognition, and spend more computation consumption.

In discriminant-function based method, the advantage of the prototype based method is that it only needs few typical samples of gestures, and doesn't need a long-time training procedure. It is very useful for a user-defined gesture application. In the application, system allows users to define their own gestures. System can't pre-define them. So we can't collect enough training sample for a user-defined gesture. In the use-case, the prototype based method is a good solution.

The advantage of training based method can achieve the highest accuracy. However, it needs lots of training samples and a long time training procedure. It can be used into pre-defined gesture recognition.

TABLE I. RECOGNITION RATE COMPARISON OF GESTURES WHEN USING DIFFERENT METHODS

Recognition Method	Method Pixel	Method Rule	Method Comb	Method Proto	Method Training
Accuracy of 5 simple-shape gestures	***97.97%***	***93.91%***		***99.5%***	***100%***
Accuracy of 10 gestures in figure 3 (Includes simple and complex shape gestures)	***95.44%***		***94.18%***	***97.5%***	***99.49%***

TABLE II. RECOGNITION RATE COMPARISON OF GESTURES IN FIGURE 3 IN MIXED RECOGNITION MODE

Method	Method Comb+ handwriting recognizer	Method Proto+ handwriting recognizer	Method Training+ handwriting recognizer
Accuracy of 10 gestures in figure 3	***85%***	***91.5%***	***98.48%***

TABLE III. COMPARISON OF RECOGNITION RATE OF ALPHANUMERIC

Methods in different modes	Just handwriting recognizer (Doesn't combine with any gesture-recognizer)	Mixed recognition mode		
		Method Comb + handwriting recognizer	Method Proto + handwriting recognizer	Method Training + handwriting recognizer
Digit	***90.16%***	***89.38%***	***90.15%***	***90.12%***
English alphabet	***88.11%***	***87.07%***	***88.07%***	***88.02%***

REFERENCES

[1] G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. *(references)*

[2] N. Kato, S. Omachi, H. Aso, Y. Nemoto, "A Handwritten Character Recognition System Using Directional Element Feature and Asymmetric Mahalanobis Distance", IEEE PAMI, March 1999. 99 (3), pp. 258–262,

[3] Y. Y. Tang, L. T. Tu, S. W. Lee, W. W. Lin, I. S. Shyu, "Offline Recognition of Chinese Handwriting by Multifeature and Multilevel Classification", IEEE Trans. PAMI , May 1998 , 98(5), pp. 556–561,

[4] B.-H. Juang, W. Chou, and C.-H. Lee. "Minimum classification error rate methods for speech recognition", IEEE Trans. on SAP, 1997, 3(5), pp. 257–265

[5] Wu Chou, B.H. Juang, "Pattern recognition in speech and language processing", CRC Press LLC, 2003

[6] S.W. Lee, J.S Park, "Nonlinear Shape Normalization Methods for The Recognition of Large-Set Handwritten Characters", Pattern Recognition, 1994, Vol.27, (7), pp. 895–902,

[7] Y.Ge, Fengjun Guo, etc. "Online Chinese Character Recognition System with Handwriting Pinyin Input", ICDAR-2005, Soul, Korean, 2005, pp. 1265–1269,

[8] F. Kimura etc. "Modified quadratic discriminant functions and its application to Chinese character recognition". IEEE Trans. PAMI, 1987, 9(1): pp149–153

[9] B-H. Juang and S. Katagiri. "Discriminative learning for minimum error classification", IEEE Trans. Signal Processing, 40(12), 1992: :3043-3054

[10] S. Katagiri, C.H. Lee, and B.H. Juang, "New discriminative algorithm based on the generalized probabilistic descent method", in Proceedings of IEEE Wordshop on Neural Network for Signal Processing, Sept. 1991, pp. 299-309