Signature Verification using Integrated Classifiers

Wan Liang(1), Lin Zhouchen(2) and Zhao Rongchun(1)

(1)Provincial Key Lab of SIIP, Dept. of CSE, Northwestern Polytechnical University, Xi' an 710072, P.R. China

(2)Multimodal User Interface Group, Microsoft Research, Asia, Beijing 100080, P.R. China

E-mail: liangwan@mail.nwpu.edu.cn, zhoulin@microsoft.com, rczhao@nwpu.edu.cn

Abstract: This paper presents a new approach for off-line signature verification. The proposed system is based on global, grid, ink distribution and texture features. The Boosting algorithm is applied to train and integrate multiple classifiers, and the distance-based classifier used as the base classifier corresponding to each feature set. Adaptive threshold is associated with individuality. Experimental results show the system is insensitive to the order of base-classifiers and gets a high verification ratio.

Keywords: Signature verification, boosting algorithm, adaptive threshold

1 Introduction

Signature verification is to evaluate whether a suspected signature is genuine or forgery. It's widely used in the fields of finance and security. Usually three kinds of forgery can happen in signature verification. Random forgery is taking the genuine signature of others for that of the current user. Skilled forgery is produced with close imitations. It is hard to be discriminated from the genuine one only by shape variations. Simple forgery is produced with the knowledge of content but without close imitations. For example, the forger signs out of his/her memory on the genuine signature.

Many systems for HSV have been proposed in the literature. Sabourin and Drouhard[7] presented an method based on directional probability density functions together with BP neural networks to detect random forgery. Qi and Hunt[6] used global and grid features with a simple Euclidean distance classifier. Bajaj and Chaudhury[4] proposed a system consisting of sub-classifiers that are based on three sets of global features. Sansone and Vento[2] proposed a sequential three-stage multi-expert system, in which the first expert eliminates random and simple forgeries, the second isolates skilled forgeries, and the third gives the final decision by combining decisions of the previous stages together with reliability estimations. Baltzakis and Papamarkos[1] developed a two-stage neural network, in which the first stage gets the decisions from neural networks and Euclidean distance classifiers supplied by the global, grid and texture features, and the second combines the four decisions using a radial-base function (RBF) neural network.

In this paper, multiple classifiers integration using the Boosting algorithm is proposed. This system is designed to detect both random and simple forgeries. In the rest part of this paper, section 2 discusses features extracted from a signature image. Section 3 presents the details of the base classifiers and integrated classifier. Section 4 gives our experimental results. Finally, section 5 concludes this paper.

2 Feature extraction

For different applications, appropriate features should be designed. In our system, four groups of features are extracted as global, grid, ink distribution and texture features.

We choose a subset of global features proposed by Baltzakis[1], which includes pure width, image area, centers of the signature, maximum vertical and horizontal projection, vertical and horizontal projection peaks, local slant angle, number of edge points, and number of cross points.

To provide the structural information on local scales, we define the grid gray feature as the average gray values in multi-resolution grids superposed on the preprocessed image. Considering variances among signatures of the same person, grid coordinates are determined adaptively based on horizontal (and vertical) pixel projection histograms. Denote $g\left(i, j\right)$ as the signature image.

$$\begin{aligned} & Hist_h\left(j\right) = \sum\nolimits_i \delta_{i,j},\ j = 1,...,width \\ & in\ which\ \delta_{i,j} = \begin{cases} 1 & \mathrm{g}\left(i, j\right) > 0 \\ 0 & else \end{cases} \\ & GridW\left(m\right) = k : \sum\nolimits_{j=1}^{k} Hist_h\left(j\right) \geq m * \Delta, m = 1,...,WNum, \\ & in\ which\ \ \Delta = \sum\nolimits_j Hist_h\left(j\right) / WNum \end{aligned} \tag{1}$$

$GirdW$ records grid horizontal coordinates. $WNum$ is the number of horizontal grid. $Hist_h$ is the project histogram along vertical direction. Calculate vertical coordinates as $GridH$.

As a pseudo-dynamic descriptor, ink distribution feature[6] gives out gradual changes of gray values on the local scale. Define high pressure area as

$$g_{hpr} = \begin{cases} 1 & g\left(i, j\right) > T_{hpr} \\ 0 & else \end{cases} \tag{2}$$

where T_{hpr} is a predefined threshold. Still use grids $GridW \times GridH$, and calculate the amount of high pressure pixels falling in each grid as ink distribution features.

As a popular approach to texture analysis, each element in the co-occurrence matrix[8] $M\left(d,\theta\right)$ represents the probability of the combination of gray values at pairs of points separated by the vector $\delta = \left(d,\theta\right)$, where d is the distance and θ the angle. Here four matrices $M\left(1,0°\right), M\left(1,45°\right), M\left(1,90°\right), M\left(1,135°\right)$ are obtained at first. Then from each matrix, four second-order statistic measurements[8] (matrix energy, difference matrix energy, difference matrix mean, and relevance variance) are evaluated. Finally group all those values into a 16-dimension feature vector.

3 Integration of multiple classifiers

3.1 Base classifier

Logically the "distance" between two signature samples of the same person will be smaller than that of two different persons. Taking the standard mean of genuine samples as the reference, we calculate distances between genuine samples and the reference, and those between forgeries and the reference. A suspected signature is evaluated to be genuine if its distance to the reference isn't larger than a threshold.

We define the adaptive threshold for each owner as

$$T^k = \frac{D_{gmax}^k + D_{gmin}^k}{2 * \sigma}, k = 1,..,4 \tag{3}$$

where D_{gmax} denotes the maximum distance between genuine training samples and the reference. D_{gmin} is the minimum distance between forgeries and the reference, and σ is the standard deviation of genuine training samples.

The definition of the above threshold considers the balance between type I error (false rejection rate, or FRR) and type II error (false acceptance rate, or FAR). With the plot between the threshold and the recognition rate, we can choose the near-optimal feature set through experiments.

3.2 Integrated classifier

In general, there are two modes of integrating multiple classifiers: parallel[1] and sequential[2]. In the parallel model, the classifier on the second stage takes the recognition results from the first stage as the input. Lacking of direct knowledge of sample distributions becomes the major defect of this model. The sequential model, however, can not decrease error rates of two types at the same time.

We find that the Boosting algorithm can combine the merits of both modes. In the training stage, it works like the sequential model; and in the evaluation stage, it works as the parallel model (See Fig.1). Generally speaking, Boosting algorithm can change many weak learners into a strong learner. Here we adopt the Adaboost algorithm[9], which is the basis of many Boosting algorithms. One of its highlights is resampling idea, by keeping the sample distribution, or weighting over the sample set. Initially the weights are set equal. After each iteration step, the weights of misclassified samples are increased and the weights of correct cases decreased accordingly, which makes the next classifier focus more on the "wrong" examples.

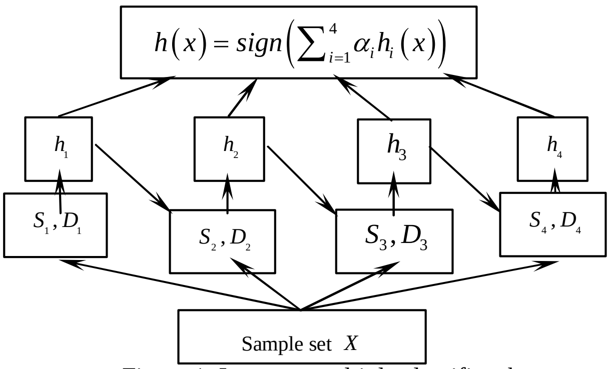

Figure 1: Integrate multiple classifiers by Adaboost algorithm

The integration of multiple classifiers using the Adaboost algorithm is illustrated in Figure 1, where h_i refers to an individual classifier for i-th feature set S_k . D_k is the sample distribution, and α_k is the weight associated with each classifier. For the current user, sub-classifiers are trained one after another. While in testing stage, the final result is obtained using the linear combination of recognition results of sub-classifiers.

Experimental tests show the order of sub-classifiers has little effects on the performance of the integrated classifier.

4 Experimental results

4.1 The signature database

Two subject sets are invited to provide Chinese signatures. One set O^1 consists of 8 owners, each one providing 12 genuine signatures, and 1 simple forgery each for the rest 7 owners. The other set O^2 consists of 20 owners, each providing 24 genuine signatures, and 1 simple forgery each for the rest 19 owners. Many factors, such as the present mood of the owner, paper quality, and stencil type, may affect the signature. Here all users use gel pens of the same type, and write on the same forms containing 3*4 writing cells.

4.2 Preprocessing

Use forms containing 12 writing cells printed on the A4 paper to collect samples. Acquire digital data by a flat bed scanner with a resolution of 300 dpi and 256 gray levels. Since the data sheet is produced in the same structure, simple rules can be used to detect the form lines and extract each individual signature.

The scanning process introduces homogenous background into signature images. To eliminate the background, local contrast enhancement is first employed. Here we convolve the signature image with a 3*3 Laplacian parametric mask. Then local binarization[10] proceeds.

Compared to the original tracks, the binarized signature shrinks across the cross-section. If taking such foreground signature track as a mask image, the gray-level signature extracted will lose much edge information. So a dilation morphological operator with 3*3 cross element is applied to make strokes wider. In the processed image, small ruptures still exist in very thin strokes. A bridge operator[10] is employed to join two connected components that are one pixel apart. Finally we extract the gray-level signature tracks by using the resultant image as a mask image over the original signature image.

4.3 Test on orders of sub-classifiers

Since one sub-classifier in the system corresponds to one feature set, the order of sub-classifiers is indeed the order of feature sets selected. The sub-classifiers used in the Adaboost algorithm is actually trained only one time, and they depends on the recognition results of the previous sub-classifier, if such a sub-classifier exists. Then whether the order of sub-classifiers may affect the performance of the integrated classifier needs to be validated.

There are all together 4!=24 different orderings in total. This test is conducted on the user set O^1. For the *i*-th user, his/her training dataset is composed of 7 genuine samples, 7 simple forgeries, and 3*8 random forgeries randomly selected from sample sets of the 8 owners in O^2. His/her testing dataset is composed of all 12 genuine samples, 9 simple forgeries, and 12*7 random forgeries selected from the sample sets of the rest 7 owners in O^1. The testing results are shown in Table 1, where E_1 is the total number (for 8 onwers) of genuine samples which are false rejected. E_2 is the total number of simple forgeries which are false accepted. E_3 is the total number of random forgeries which are false accepted, and numbers '1,2,3,4' stand for texture features, grid gray features, ink distribution features and global features respectively. Table 1 answers the above-mentioned question, that the specific order of sub-classifiers has little effects on the integrated classifier.

Table 1: System verification results

	E1	E2	E3		E1	E2	E3		E1	E2	E3
1-2-3-4	5	6	3	3-4-1-2	6	6	4	2-3-4-1	3	5	4
1-3-2-4	4	4	4	4-1-2-3	6	7	6	2-4-3-1	6	8	5
1-4-2-3	4	4	2	4-2-1-3	3	9	4	3-1-4-2	5	8	6
2-1-3-4	5	8	5	4-3-1-2	2	8	5	3-2-4-1	2	5	1
2-3-1-4	5	6	4	1-2-4-3	4	7	3	3-4-2-1	1	9	2
2-4-1-3	5	7	1	1-3-4-2	3	7	2	4-1-3-2	4	5	1
3-1-2-4	6	2	2	1-4-3-2	4	5	2	4-2-3-1	3	6	3
3-2-1-4	3	5	5	2-1-4-3	3	7	2	4-3-2-1	4	4	4

4.4 System testing and results

The proposed method is tested on the user set O^2. For the *j*-th user, his/her training dataset $D_2(j)$ consists of 12 genuine samples, 3 simple forgeries, and 3*8 random forgeries randomly selected from O^1. Denote the integrated classifier as H. We use the remaining 12 genuine signatures, 9 simple forgeries, and 12*19 random forgeries randomly selected from O^2, as the testing set for the *j*-th user.

Type I error and Type II error are calculated. We record error rates for individual feature set and the integrated method.

Compared with single feature set, the integrated classifier reaches much better performance on random forgeries (refer to the fourth column in Table 1). Among the feature sets, texture feature has the discrimination ability comparable to the integrated method.

We also find the number of simple forgeries used in the training process will affect the performance of individual classifiers, and the integrated classifier as well. It may due to the fact that simple forgeries have smaller distances to the reference mean than random forgeries. Their introduction will decrease the relevant threshold, which in turn increases FRR will increase a bit accordingly. Based on the tests, we finally include 3 simple forgeries in the training dataset.

Table 1: System verification results

Error rate	FRR (%)	FAR(%) Simple Forgery	FAR(%) Random Forgery	Total Error (%)
Texture features	17.92	11.11	1.71	2.83
Grid gray features	14.17	35.0	6.01	7.45
Ink distribution features	15.83	22.22	4.54	5.72
Global features	17.92	41.67	10.48	11.97
The integreated classifier	10.00	16.67	0.66	1.69

4.5 Comparison test

The present approach is also compared with that proposed by Baltzakis and Papamarkos[1]. Table 2 records error rates of individual feature set in combination with neural network, results of the two-stage system with RBF used as the second stage classifier (proposed by Baltzakis), and results of the two-stage system with BP used in the second stage.

It is clear that using different classifiers in the second stage results in quite different performance. That the second classifier is trained only by using the results from the first stage may explain this. In other words, no knowledge of sample distributions is taken into account in this two-stage system.

Table 2: Results of Baltzakis and Papamarkos' method based on user set O^2

Error rate	FRR (%)	FAR(%) Simple forgery	FAR(%) Random forgery	Total error (%)
Texture feature NN	25.00	30.56	0.86	3.09
Grid skeleton NN	25.42	22.78	1.34	3.27
Global feature NN	42.08	27.22	2.26	5.08
Two-stage system with RBF	77.92	7.78	0.81	4.78
Two-stage system with BP	14.58	33.33	0.24	2.13

5 Conclusions

This paper presents a new signature verification system that uses the Adaboost algorithm to integrate multiple classifiers. The effectiveness of our system has been testified on a small user sets. The experimental results show that the integrated classifier can reach a better performance than individual classifiers and the performance is insensitive to the order of the base classifiers. Moreover, the user-adaptive thresholds free the system from manual parameter tuning. Finally, its superiority over some existing algorithms is also testified.

References

[1] Baltzkis H, Papamarkos N, 2001, A new signature verification technique based on a two stage neural network classifier, *Engineering Application of Artificial Interlligence*, 14: 95-103

[2] Sansone C, Vento M, 2000, Signature verification: increasing performance by a multistage system, *Pattern Analysis & Application*, 3:169-181

[3] Kashi R, Nelson W, 2002, Signature verification: benefits of multiple tries, *Proceedings of the 8th international workshop on frontiers in handwriting recognition*, 424-427

[4] Bajaj R, Chaudhury S, 1997, Signature verification using multiple neural classifiers, *Pattern Recognition*, 30(1):1-7

[5] Sabourin R, 1997, Off-line signature verification: Recent advances and perspectives, *BSDIA'97*, Curtiba, 84-98

[6] Qi Y Y, Hunt B R, 1994, Signature verification using global and grid features, *Pattern Recognition*, 22(12): 1621-1629

[7] Sabourin R, Drouhard J P, 1992, Offline signature verification using directional PDF and neural networks, *Proceedings 11th international conference on pattern recognition*, 2:321-325

[8] Franke K, Bunnemyer O, Sy T, 2002, Ink texture analysis for writer identification, *Proceedings of 8th international workshop on frontiers in handwriting recognition*, 268-273

[9] Freund Y, Schapire R E, 1999, A short introduction to boosting, *Journal of Japanese Society for Artificial Intelligence*, 14(5):771-780

[10] Franke K, Koppen M, 2001, A computer-based system to support forensic studies on handwriting documents, International Journal on Document Analysis and Recognition, 3:218-231

摘 要：本文提出一种新的脱机签名鉴定方法。采用全局、网格、墨迹分布以及纹理四种特征，本系统以基于距离准则的分类器作为子分类器，进而用 Boosting 算法训练、整合多分类器。其中采纳自适应的阈值选择方法训练各子分类器。实验表明本文提出的算法对子分类器的顺序不敏感，达到很高的识别率。

关键词：笔迹鉴 Boosting 算法 自适应阈值