Off-line Signature Verification Using Linear Regression Classifier

Bo Xu^{1,a}, Daozhi Lin^{2,b}, Hongyang Chao^{3,c}, Weifeng Li^{4, d*}, Qingmin Liao^{5,e}

¹Department of Electronic Engineering/Graduate School at Shenzhen, Tsinghua University, China

²School of Software, Sun Yat-sen University, China

³School of Software, Sun Yat-sen University, China

⁴ Department of Electronic Engineering/Graduate School at Shenzhen, Tsinghua University, China

⁵ Department of Electronic Engineering/Graduate School at Shenzhen, Tsinghua University, China ^axu-b13@mails.tsinghua.edu.cn, ^be931786222@qq.com, ^c isschhy@mail.sysu.edu.cn, ^{d*}li.weifeng@sz.tsinghua.edu.cn, ^eliaogm@tsinghua.edu.cn

Keywords: Linear Regression Classifier, Local Binary Patterns, Tuned Linear Regression Classifier.

Abstract. In this paper we propose a novel classification method based on Linear Regression Classification (LRC) for offline signature verification. The class-specific models can be simply established by using the registered samples (examples), and a test signature can be linearly represented by these registered samples. Then the tuned-LRC is constructed to capture the nonlinear information when the fundamental linear assumption is invalid in LRC. In contrast to the conventional classifiers used in signature verification, our proposed methods are very simple and no training stage is needed, and the dictionary can be easily expanded by additional samples. The experiments conducted on GPDS960Graysignature database demonstrate the effectiveness of the proposed methods.

Introduction

Handwritten signatures are widely accepted as a convenient means of document authentication, and play a very special role in the wide series of biometrics authentication. Writer authentication systems can be categorized into on-line and off-line systems.

Similar to other biometric recognition systems, an off-line handwritten signature verification system generally involves two processes: effective and robust features and discriminant classifiers. Regarding feature extraction, how to extract and apply pseudo-dynamic information in a reasonable and accurate way is very important for off-line signature verification. There are plenty of approaches to reconstruct dynamic information from static handwriting records [1]. It is comprehensively studied on extracting the robust features from a static signature image. For instance, local binary pattern (LBP), which has already been applied in face and handprint biometrics, was adopted in [2,3]. Statistical texture analysis, e.g. the Gray Level Co-occurrence Matrix (GLCM), has been proposed and successfully applied to signature verification tasks [4].

In the verification phase, most classification approaches can be generally categorized into three groups: template matching, statistical and structural approach. Regarding template matching techniques, a test sample is matched against the templates of authentic/forgery signatures where the most common approaches adopt dynamic time warping (DTW) [5] for signature matching. Inside statistical approaches, distance-based classifiers can be deployed [6]. Neural networks [7] have also been widely used, due to their capabilities in learning and generalizing. More recently, special attention has been devoted to the use of hidden Markov models (HMMs) [8]. Support vector machines (SVMs) is another promising structural approach to signature verification. It trains separating hyperplane that maximally separates the classes in the high dimension and tends to be among the best methods with the limited data available [2].

However, most of these classifiers are complicated and need adequate data to train the model.

In this paper, Linear Regression Classifier is deployed for off-line writer authentication. LRC techniques have been applied to different fields, such as face recognition [9], etc. For signature verification, it is assumed that samples from a specific object class lie on a linear subspace. Then, a class-specific model by simply using the registered samples (examples) is constructed. A test signature can be linearly represented by these registered samples. The proposed LRC based method is simple and fast since no training stage is needed, and the dictionary can be easily expanded with additional samples.

Linear Regression Classifier Based Method

v = Da.

Linear regression classifier (LRC) has been introduced in face recognition [9], which is based on the well-acknowledged concept that face images from a specific class lie on a linear subspace. Nevertheless, little relevant study was found in signature verification research. For a signature verification problem, let x be the feature vectors (e.g., LBP) extracted from a static signature image.

For a particular client, we have the given samples (or examples) $x_1, x_2, ..., x_M$ to construct a dictionary $D = [x_1, x_2, ..., x_M]$, where M is the total number of the samples. Once a test signature image in terms of LBP feature vector y is given, a linear representation of y can be written as

where a is the coefficient vector whose entries are associated with the *i*-th training or registered examples.

(1)

From the viewpoint of Linear Regression Classification (LRC), the solution of a can be found out by the below equation.

$$a = \arg \min \left\| y - Da \right\|_2$$
(2)

where $\|\cdot\|_2$ is l_2 -norm and $\|y - Da\|_2$ represents the signal reconstruct residual. Eq. 2 is well conditioned and a can be estimated using least-squares estimation as below.

$$\mathbf{a} = (D^T D)^{-1} D^T \mathbf{y} \tag{3}$$

Once *a* is determined, whether *y* is a genuine signature can be judged by comparing the residual $||y - Da||_2$ with a threshold T_1 , i.e.,

$$\left\| y - D\boldsymbol{a} \right\|_2 \frac{\boldsymbol{p}}{\boldsymbol{f}} T_1, \tag{4}$$

If the residual is less than T_1 , the test feature vector y is then judged as an authentic signature, and vice versa.

Actually, LRC works fine when genuine samples lie on a linear subspace spanned by registered samples. However, the structure of subspace is usually nonlinear, which will result in large errors in reconstruction stage. Therefore, it is very important to model the nonlinear space structure of registered samples. In order to capture nonlinear structure information, the tuned-LRC based classifier is designed, as shown below. For a particular client, given his/her signature samples (or examples) $x_1, x_2, ..., x_M$, we compute the mean feature vector as the centroid

$$c = \frac{1}{M} \sum_{m=1}^{M} x_m \tag{5}$$

let $y = x_{M+1}$, then the new centroid c' considering this test sample y.

$$c' = \frac{1}{M+1} \sum_{m=1}^{M} x_m,$$
(6)

Based on the origin and the new centroid, the term d representing change of clustering degree with regard to subspace spanned by training samples and this given test sample.

$$d = \frac{M \sum_{m=1}^{M+1} \|x_m - c'\|_2}{(M+1) \sum_{m=1}^{M} \|x_m - c\|_2},$$
(7)

where, $x_{M+1} = y$. Still, based on LRC, the linear coefficients could be attained by solving Eq. 2. But then whether y is a genuine signature will be judged by comparing $\|y - Da\|_2 * d$ with a threshold T_2 , i.e.,

$$\left\| y - Da \right\|_{2}^{*} d \frac{\mathbf{p}}{\mathbf{f}} T_{2}$$
(8)

Hence, the term d tunes LRC method when registered samples could not properly linearly represent test sample.

Experimental setting

Regarding to GPDS960Gray database, the first 300 users are selected for our experimental evaluation. As to each user, there are 24 genuine signatures and about 30 simulated forgeries. In our experiment, both fictitious signatures and simulated forgery types were used to test. 10 genuine signature samples are selected randomly from the signatures of the other users to construct the training set while the remaining genuine signatures are used for testing. As to random experiment, 30 fictitious signatures are randomly selected from all the remaining users, and as to simulated experiment, the 30 simulated forgeries of each signer are selected as test samples. The training set is randomly selected, and the equal error rate (EER) are provided, which is the mean of ten repeated experiments.

Regarding to features, Local Binary Pattern (LBP) features were evaluated on GPDS960Graysignature database, which use same background removal and block division approaches with [2]. In the experiment, a signature image is firstly pre-processed by signature detection and background removal. Then the local binary pattern (LBP) is extracted. Spatial histogram is used to model the distribution of the local pattern to avoid losing the location of the different structures inside the image. There are several papers dividing signature images into many blocks to achieve a good performance, to keep consistent with current study and also to demonstrate influence on LRC with different blocks. In our experiment, there are several kinds of block divisions, that is the image is divided into several equal vertical blocks and horizontal blocks which overlapped by 60%. For example, $3 \times 4 = 12$ blocks means three equal horizontal blocks and four equal vertical blocks, hence there are twelve blocks in total. As to LBP, the histograms of LBP inside each block are calculated with 255 bins.

Experimental results

To keep consistent with current study and also to demonstrate influence on LRC with different blocks, $1 \times 1 = 1$ block, $1 \times 2 = 2$ blocks, $2 \times 3 = 6$ blocks, in the experimental settings, notably, $3 \times 4 = 12$ blocks was adopted by several papers.



Fig. 1. Performance analysis with different blocks in Random and Simulated experiment in terms of equal error rate (EER) in percentages (%)

Fig. 1 illustrates the performance of LRC decreases with increase of blocks when there is 5 training samples, which means this block division method is not a good choice for LRC. That is because signatures in the database have different sizes and the rigid block division method gives rise to different block textural information in two signatures written by the same signer. Actually, with regard to LRC, low dimensional features could achieve a well promising performance.

Regarding LRC, we attain the reconstruction residual to verify whether a test sample is genuine or forgery. However, in many cases the fundamental linear assumption is invalid. Hence a new term is introduced to capture this nonlinear information, which represent clustering degree. After attaining this term, the multiple with previous reconstruction residual constructs our proposed tuned-LRC. Table 1 demonstrates the verification performance in terms of equal error rate (EER) using LRC and tuned-LRC.

	Classifier	TestType	1*1	1*2	2*3	3*4
five training samples	LRC	Random	10.22	11.7	13.86	15.27
		Simulated	13.56	14.68	15.18	16.49
	tuned-LRC	Random	7.81	9.45	11.87	12.76
		Simulated	13.02	13.58	14.31	15.29
ten training samples	LRC	Random	8.37	9.56	11.21	12.65
		Simulated	9.92	11.15	12.93	14.55
	tuned-LRC	Random	6.38	7.62	8.76	10.53
		Simulated	9.21	10.34	12.15	13.89

Table 1: Performance comparison among LRC, tuned-LRC in terms of EER (%)

Firstly, comparing with 5 training samples, the verification performance is improved on the whole in Random and Simulated experiment with different classifiers by using ten training samples, that is using more training (registered) samples helps us separate genuine signature and fictitious forgeries. Secondly, as to LRC and tuned-LRC, it can be observed that the performance have been improved both in Random experiment and Simulated experiment by deploying clustering degree information. From Table I, the performance of Random experiment as to random forgeries is improved such as EER (%) has been reduced about 2 for 1×1 , 1×2 , etc. Similarly, consistent trend has been demonstrated in Simulated experiment. Therefore, tuned-LRC could modify LRC properly, and achieve a well promising performance facing the fictitious forgeries and simulated forgeries.

Conclusions

A novel classification method was proposed for off-line signature authentication based on Linear Regression Classification. Our signature verification experiments on GPDS960Graysignature database demonstrate it can achieve a comparable performance. To further attain a more effective classifier, the tuned-LRC based classifier was designed. The experiment results illustrate the effectiveness using the proposed method considering nonlinear information. In contrast to the conventional classifiers used in signature authentication, the effectiveness of tuned-LRC comes from several aspects including its simplicity and timesaving since no training stage is needed, and the dictionary can be easily expanded by additional samples.

References

[1] V. Nguyen, M. Blumenstein, and G. Leedham, Globalfeatures for the off-line signature verification problem, in Document Analysis and Recognition, 2009. ICDAR'09. 10th International Conference on. IEEE, 2009, 1300–1304.

[2] M. A. Ferrer, J. Vargas, A. Morales, and A. Ordonez, Information Forensics and Security, IEEE Transactionson, 7, 3 (2012) 966–977

[3] J. Hu and Y. Chen, Offline signature verification using real adaboost classifier combination of pseudo-dynamic features, Document Analysis and Recognition. ICDAR 2013, IEEE, 2013.

[4] J. Vargas, M. Ferrer, C. Travieso, and J.B. Alonso, Pattern Recognition, 44, 2 (2011) 375–385.

[5] A. Shanker and A. Rajagopalan, Pattern Recognition Letters, 28 (2007) 1407–1414.

[6] S.N. Srihari, A. Xu, and M.K. Kalera, Learning strategies and classification methods for off-line signature verification, Frontiers in Handwriting Recognition, 2004. IWFHR-9 2004. Ninth International Workshopon, 161–166, IEEE, 2004.

[7] H. Baltzakis and N. Papamarkos, Engineering applications of Artificial intelligence, 14, 1 (2001) 95–103.

[8] L. Yang, B. Widjaja, and R. Prasad, Pattern recognition, 28, 2 (1995) 161-170.

[9] I. Naseem, R. Togneri, and M. Bennamoun, IEEE Transactions on, 32, 11 (2010) 2106–2112.