On-line Signature Verification Using Local Shape Analysis

Mingfu Zou, Jianjun Tong, Changping Liu, Zhengliang Lou
714 Room, Automation Building, Institute of Automation,
Chinese Academy of Sciences, P.O.Box 2728, Beijing, P.R.China 100080
zoumingfu@tsinghua.org.cn

Abstract

This paper presents a novel approach to the on-line signature verification using local shape Analysis. First, segment the input signature into several segments using HMM (Hidden Markov Model). Then, combine two adjacent segments to form a long segment and get its spectral and tremor information using FFT (Fast Fourier Transformation). At last, accept it or reject it based on the similarity between the spectral and its prototype. In addition, we proposed a novel initialization algorithm to avoid the local optimal of the HMM's re-estimation and a novel algorithm to avoid losing the important information at cusps in preprocessing. Combining the local shape analysis with the local time-based comparison, we get promising experimental results.

Keywords: signature Verification; Hidden Markov Model; Segmentation; Biometric authentication; Shape analysis;

1. Introduction

Signature verification is a biometric attribute. Well-known biometric methods include iris-, retina-, face- and fingerprint-based identification and verification [1]. They are being used more and more in our ordinary life. Although attributes like iris, retina and fingerprints do not change over time, they require special and relatively expensive hardware to capture the image. An important advantage of the signature verification compared with other biometric attributes is its long tradition in many common commercial fields [2], such as e-business, which includes on-line banking transactions, electronic payments, access control and so on. So signature verification is a very popular research area right now. Generally, it is accepted that an individual's signature is unique, although nobody verifies it. In fact signature verification is a difficult pattern problem because the intra-class variations could be large.

An on-line signature is best represented by multiple models, which could be local, global, shape-based, or time-based [3]. Compared with global models, the other models have many advantages [4]. This article introduces a technique that combines the local time-based model and the segmental shape-based model together.

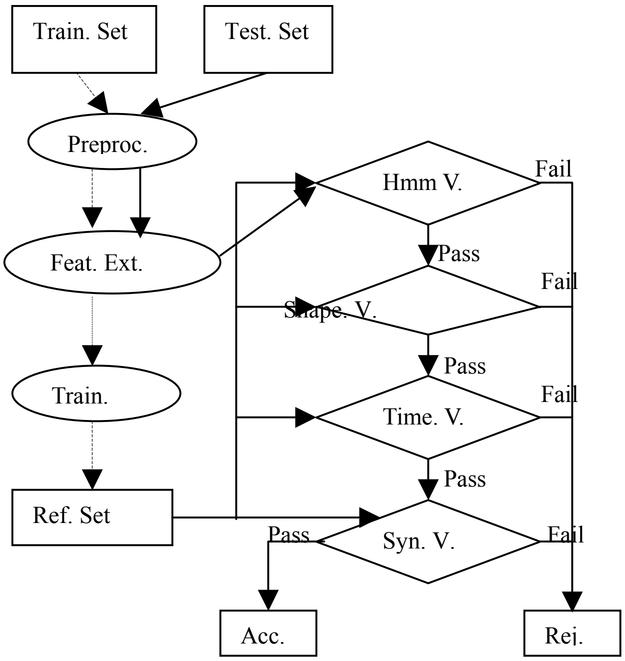
Segmentation is one of the most difficult problems and it is also so important that it should be considered separately. There are many techniques about segmentation reported in the literature [4,5,6]. Immunity to intra-signer variations offers the greatest challenge to most techniques applied. In this paper, a novel algorithm is proposed to yield consistent and reliable results for different signature instances of a signer.

Signatures include many smooth curve elements other than line segments and sharp angle elements. Important information about individuals' biometric features is included in signatures [7]. They change from one to another instance, but the forged signatures deviate much bigger than the genuine. In order to analysis these smooth curve elements, [4] segments signatures at relatively non-complex shape. In this paper, a novel algorithm is also proposed to analysis them. First, divide the input signature to several segments using HMM. Then, combine two conjoint segments together to form a long segment and get its spectral and tremor information using FFT. At last, accept it or reject it based on the distance between the spectral and its prototype.

In addition, a novel initialization algorithm is used to avoid local optimal of the re-estimation algorithm and at meantime a novel algorithm is also proposed to retain the important information at the cusps.

The rest of the paper is organized as follows. Section 2 presents the proposed signature verification. Section 3 shows the experimental result and conclusion follows in section 4.

2. System design



Train. = Training   Test. = Testing   V. = Verification
Preproc. = preprocessing   Feat. = Features
Ext. = Extracting   Ref. = Reference
Seg. = Segmentation   Syn. = Synthetic
Acc. = Accept   Rej. = Reject

Fig. 1. Modules of the signature verification system. The dashed line arrows show the flow of data during training, while the solid line arrows show the data flow during verification.

Fig.1 describes our system.

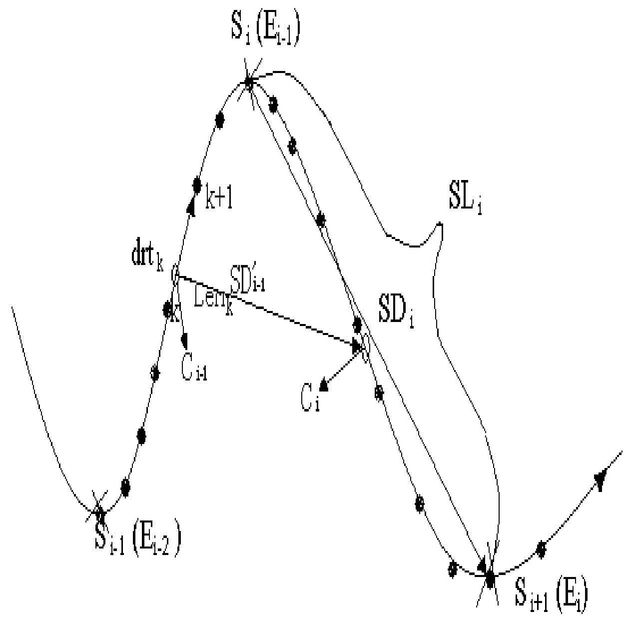
2.1. Data acquisition and preprocessing

The raw available from tablet consists of three dimensional series data:

$$(x(t), y(t), p(t) \in R^2 \times \{0,1,\Lambda\ \Lambda\ ,255\}, t = 1,2,......,T \quad (1)$$

where $(x(t), y(t)) \in R^2$ is the pen position at time t, and $p(t) \in \{0,1,\Lambda\ \Lambda\ ,255\}$ represents the pen pressure.

There are many methods used for preprocessing. In [5], they used Fourier normalization as preprocessing. In [4], they uniformly re-sampled the signature to N user-defined equidistant point along the signature curve. In [8], they not only did the same as in [4], but also retained the critical points, such as start and end points of a stroke and points of trajectory change, because they carry important information. Hence, every instance of a signer has different preprocessed points. In [7], they resample the signature based on the current direction. While in [9], they opposed any preprocessing for they deem that the original points retained the difference in writing speed and writing rhythm in various partitions of the signature.



·=sampling point $Len_k = ((x_{k+1} - x_k)^2 + (y_{k+1} - y_k)^2)^{0.5}$
X= segmenting site $drt_k = [(x_{k+1} - x_k) + j(y_{k+1} - y_k)] / Len_k$
C = segmental center

Fig. 2. Curve showing the geometric parameters used in the following algorithms. Drt_k is the direction vector of the k^{th} point. S_i (E_i) is the start (end) point of the i^{th} segment. SL_i and SD_i are the length and direction of the i^{th} state, respectively. SD'_{i-1} is the direction from the i-1[st] segmental center to i [th] segmental center.

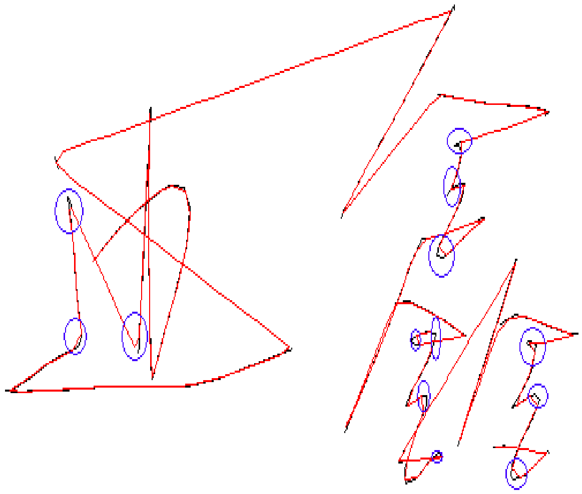
In our experiment, we use different preprocessing ways for different models. For local time-based model, we do it as in [9]. But for shape-based model, we resample a signature based on the variance of direction and the distance between two points. The proposed algorithm in this paper can be described as shown in Algorithm 1.

ALGORITHM 1: THE PREPROCESSING ALGORITHM
Step 1: Compute the pseudo-distance
1.1 Initialize
PreDrt= drt_1, PreDst = Len_1
$PsdLen_i$=0 $1 \le i \le T$
1.2 Iteration: for i=2,…,T-1{
CurDrt= drt_i, CurDst = Len_i

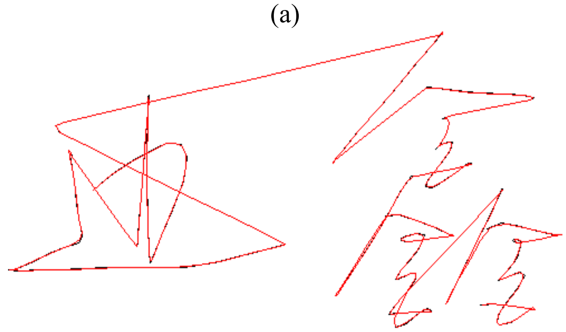
$$PsdLen_i = |CurDrt - PreDrt| * \sqrt{\text{PreDst} + \text{CurDst}}/4 \quad (2)$$

$PsdLen_{i-1}$+=$PsdLen_i$
PreDrt=CurDrt, PreDst=CurDst }
Step 2: Uniformly resample the signature based on the

Proceedings of the Seventh International Conference on Document Analysis and Recognition (ICDAR 2003)
0-7695-1960-1/03 $17.00 © 2003 IEEE



(a)



(b)

Fig.3 Preprocessing of on-line signature: (a) Re-sampling the signature uniformly with equidistant spacing. These blue circles label the lost cusps. (b) Re-sampling with our proposed algorithm. There has no lost cusp. The black tracking represents the original handwriting while the red tracking denotes the re-sampled handwriting.

Fig. 3 shows the result using different preprocessing methods. From the two pictures, it is natural to draw a conclusion that our method retains the important information at the cusps while the common equidistant re-sampling method will lost it.

2.2. Feature extraction

In our experiments, we extracted four kinds of features: the global, original local time-based, preprocessed spatial and segmental features. The details about global features are omitted. Local time-based features are: (i) the direction drt, (ii) the relative speed v (absolute speed normalized by average signing speed), (iii) and delta pressure Δp. They are all extracted from the original sampling points. The third kind of feature is used for HMM. Because every state of the HMM corresponds to a segment in the signature, we use the preprocessed local direction to represent it. Section 2.4 describes the last kind of features in detail.

By experiments, we found our extracted features are reliable and useful.

2.3. HMM learning and verification

In our experiments, each signature is modeled by a single left-to-right HMM with loop and forward transition whose probabilities are re-estimated during training. Although signature is a time-evolving non-stationary signal, for every segment in a signature, we can assume it is stationary signal. Therefore, the number of state equals the number of segment of the signature which has the highest similarity with the others in the training data.

It is well known that the re-estimation algorithm of the HMM should reach a local maximum for the likelihood function [10]. In order to make the local maximum become the global maximum, we must choose the right initial estimate of the HMM parameters. In [5], the initial model parameters are obtained through equal-length segmentation of all the training samples. While in [11], the model is initialized by the segments bounded by the condition $v_t=0$. The proposed algorithm in this paper can be described as shown in Algorithm 2.

Training of the HMM parameters is done by using Maximum Likelihood criterion and applying the Viterbi approximation [12]. Based on the trained model, λ, we can get the log-likelihood for every training samples:

$$LikeliHood[i] = \log(-p(Xi \mid \lambda)) \qquad (3)$$

where Xi is the ith training signature's feature string. What's more, we can segment the input signature by backtracking. Fig.4 shows the segment result using HMM. From these pictures, it is obvious that our method can get reliable segment.

ALGORITHM 2: THE INITIALIZATION ALGORITHM

Step 1: Initially Segment using an external segmentation method:

$SegSite_{ij}$ = jth segment site of the ith sample

Seg_{ij} = jth segment information of the ith sample

$1 \le i \le N$, $1 \le j \le NumSegi$

Step 2: Get the similarity between training samples' segment using dynamical programming:

$$Simil_{ij}=Simil_{ji}=DP(Seg_i,Seg_j) \quad 1 \le i \le N \qquad (4)$$

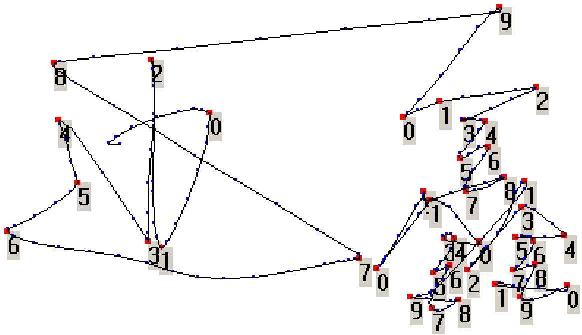
Step 3:Get the best segmenting instance

$$\text{The best segment} = \underset{1\le i\le N}{Max} \sum_{1\le j\le N} Simil_{ij} \qquad (5)$$

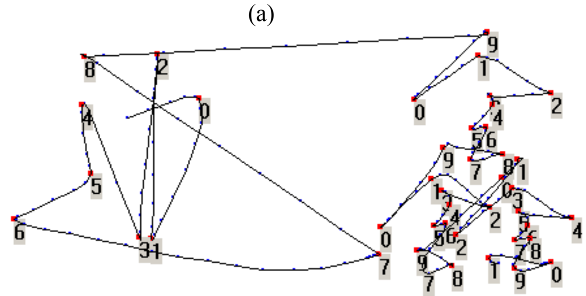
Step 4: Initialize the model based on the best instance

Each of the segment in the best corresponds a state in the HMM. Based on the matched information in Step 2, initialize parameter using their means along the matched path.

0-7695-1960-1/03 $17.00



(a)



(b)

Fig.4 Reliable and consist segmenting using HMM. The red blocks label the segment site and the digits beside the red blocks label the units of the current segment's serial number.

Given several signatures signed by a same person, we can use the statistical method to estimate the mean and the variance of the log-likelihood:

$$MeanH = \sum_i LikeliHood[i] / N$$

$$VarH = \sum_i (LikeliHood[i] - MeanH)^2 / N \qquad (6)$$

During verification, the Viterbi algorithm is used to obtain the likelihood that the signature can be modeled by the HMM of the particular subject and segmentation information which is used for further fine verification. Assuming that the distribution of the genuine log-likelihood is Gaussian distributions, we proposed the following measure of similarity between the signature and the model:

$$SimilH = (LogLikelihood - MeanH) / VarH \qquad (7)$$

If $SimilH$ is less than a threshold, we can reject it directly. Otherwise, we should further verify it.

2.4. Verification using shape and tremor analysis

For every segment, there are many features that can be extracted. There are 32 features extracted in [11] while only 11 features are extracted in [6]. In these papers, the detailed steps for i^{th} segment's feature extraction are listed as follows:

Step 1): Compute the feature between i^{th} and $i+1^{st}$ segment and combine them together to form a long segment. Between the two segments, we extract the following four features: SD'_i and $ID_i = SD_{i+1} - SD_i$ (They are all showed in Fig. 2.). After getting the preprocessed signature's segment information using HMM, we get the original signature's information by finding its corresponding site.

Step 2): Normalize and re-sample the combined segment. First, we move the center of the combined segment to origin. Then, normalize it in order to make the distance between width and height of the segment as big as a predicted value (we set it to 1000 in our experiment). At last, uniformly re-sample the segment to N+1 (in our experiments, N is 32) user-defined equidistant points along the curve.

Step 3): Compute the combine segment's spectral information. Let $z(n) = x(n) + jy(n)$, $0 \le n \le N$ be the boundary sequence. As in [13], we can get the FD (Fourier Descriptor):

$$Z(k) = \sum_{n=0}^{n=N-1} \left(b(n) \big/ |b(n)| \right) * \exp\left(-j * 2 * \pi * n * k \big/ N\right) \qquad (8)$$

where $b(n) = z(n+1) - z(n)$. In our experiments, not only do we use the first ten harmonic components of Z(k) as the feature vector as in [14], but also use the last ten components because of they containing much more tremor information than others.

At last, we get the segmental feature vector:

$$v = (SD_{i_x}, SD_{iy}, ID_{ix}, ID_{iy}, Z(1), \ldots Z(10), Z(22), \ldots z(31)) \qquad (9).$$

Using the statistical method, we can get the mean value $\bar{v}$ and standard deviation σ of these features.

During verification, we use the follow formula to compute the shape and tremor similarity:

$$SimilS = \left(\sum_{Seg=1}^{SegNum} \sum_{i=0}^{23} |v[i] - \bar{v}[i]| / \sigma[i] \right) / (24 * SegNum) \qquad (10)$$

We can use the same strategy as in section 2.3 to get the similarity $SimilS$ and to decide whether to reject a signature or not.

2.5. Local time-based verification

As in section 2.3, we can get the local time-based similarity, $SimilL$, and make decision to accept or reject a signature.

2.6. Combining all the verifications

After getting the score (difference) of every method, we combine all the results as following formula:

$$Simil = WL * SimilL + WS * SimilS + WH * SimilH \qquad (11)$$

where W (*Weight*), we set it be equal to the inverse standard deviation of similarity during training, is the weight for each verification method and $WL + WS + WH = 1$.

If $Simil$ is less than a thresh, we reject it as a forger, otherwise, accept it as a genuine signature.

3. Experimental Result

This section reports our preliminary experiment using the algorithm described above. Three datasets, called DB1, DB2 and DB3, are used for evaluation. But for the type of forgery signature and the number of participator, they are all the same. DB1, there are only two different signers, includes 40 genuine signatures and 80 random forgeries. DB2, there are 40 different signers, contains 640 genuine signatures and 800 imitated forgeries.

We should give more details about DB3, although there are only two signers. This dataset includes 1000 genuine 10000 over-the-shoulder imitated forgeries. In practice, we created it by testing the signature verification system in [15]. Hence, it is in live experiments. In testing, while the signer is seated and writes signatures, the forger is looking over the shoulder of the signer to capture the dynamic information of the written signature. Then, the forger imitates the signature and tries to imitate the movements of the original signer to obtain the over-the-shoulder imitated forgeries. This procedure is repeated with the signer and forger switching roles until the forger is familiar with the original shape and movements. What's more, at any time, the forger can capture the dynamic information by asking the signer to re-write its signatures. After two weeks hard work, it was completed.

In our experiments, we use the first 5 genuine signatures as the training samples, the other genuine and all the forged signatures as the test set. FAR (False Accept Ratio) and FRR (False Reject Ratio) are described in table 1:

Table 1: Experimental result

	FRR	FAR
DB1	6. 67%	0. 00%
DB2	9. 94%	0. 50%
DB3	11. 30%	2. 00%

4. Conclusion

This paper proposed a novel on-line signature verification algorithm based on local shape analysis. A new preprocessing algorithm and a new initial estimate for HMM are both adopted, too. Preliminary experimental results look promising.

References:

[1] Luan L. Lee, T. Berger, E. Aviczer, “Reliable On-Line Human Signature Verification Systems”, IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL 18, NO. 6, JUNE 1996

[2] R.w Dai, H.W. Hao, X.H. Xiao, “System and Intergration of Chinese Character Recognition”, pp. 280-320

[3] Ma Mingming, W. S. Wijesoma and Eric Sung, “An automatic on-line signature verification system based on tree models”, in Proc of Canadian Conference on Elect and Comp. Eng., pp.890-894,2000.

[4] Yue,K.W.;Wijesoma,W.S. Wijesoma, “Improved segmentation and segment association for on-line signature verification”, Systems ,Man ,and Cybernetics , 2000 IEEE International Conference on, Volume : 4, 2000

[5] Kashi, R.S. ; Hu, J., Nelson, W.L. Turin, “On-line handwritten signature verification using hidden Markov model features”, Document Analysis and Recognition, 1997., Proceedings of the Fourth International conference on, Volume: 2 ,1997

[6] Taik H. Rhee, Sung J. Cho, Jin H. Kim, “On-line Signature Verification Using Model-Guided Segmentation and Discriminative Feature Selection for Skilled Forgeries”, http://ai.kaist.ac.kr/publication/sign01.pdf

[8] A. K. Jain, F. D. Griess, S. D. Connell, “On-line signature verification”, Patter Recognition 35(2002) 2963-2972

[7] T. Ohishi, Y. Komiya, T. Matsumoto, “On-line Signature Verification using Pen-Position, Pen-Pressure and Pen-Inclination trajectories”, ICPR'00-volume 4, September 03-08,2000

[9] Gerhard Rigoll, Andreas Kosmala, and Daniel Willett. “A Systematic Comparison of Advanced Modeling Techniques for Very Large Vocabulary On-Line Cursive Handwriting Recognition”, In Seong-Whan Lee, editor, *Advances in Handwriting Recognition*, chapter 2, pages 69-78. World Scientific, 1999

[10] X.d. Huang, A. Acero, H.W. Hon, “Spoken Language Processing: A Guide to Theory, Algorithm and System Development”, Prentice Hall PTR, 2001, pp.367-404

[11] J.G.A. Dolfing, E.H.L. Aarts and J.J.G. M., ”On-line signature verification with hidden markov models”,In Proceedings of the International Conference on Pattern Recognition, pages 1309, August 1998.

[12] L. Rabiner and B. Juang, “Fundmentals of speech recognition”. Prentice hall, first edition, 1993.

[13] E. Persoon and K.S. Fu, “Shape Discrimination Using Fourier Descriptors”, IEEE Trans. On Computers, Vol. C-21, No.3, Mar 1972.

[14] Nai-Jen Cheng ; Chi-Jain Wen ; Hon-Fai Yau ; David Hwang Liu; Kuei Liu ; Kun-Chi Cheng ,Bor-Shenn Jeng On-line Chinese signature verification with mixture of experts, Security Technology , 1998. Proceeding . ,32nd Annual 1998 International Carnahan Conference on ,1998.