NONLINEAR SEPARATION OF SIGNATURE TRAJECTORIES FOR ON-LINE PERSONAL AUTHENTICATION

Mitsuru Kondo, Daigo Muramatsu, Masahiro Sasaki, Takashi Matsumoto

Department of Electrical, Electronics and Computer Engineering, Graduate School of Science and Engineering, Waseda University

ABSTRACT

Authentication of individuals is rapidly becoming an important issue. This paper proposes a new nonlinear algorithm for pen-input on-line signature verification incorporating pen-position, pen-pressure and pen-inclinations trajectories.

A preliminary experiment is performed on a database consisting of 1849 genuine signatures and 3174 *skilled*† forgery signatures from fourteen individuals. Since no fine tuning was done, this preliminary result looks very promising.

1. INTRODUCTION

Personal identity verification has a great variety of applications including EC, access to computer terminals, buildings, credit card verification, to name a few. Algorithms for personal identity verification can be roughly classified into four categories depending on static/dynamic and biometric/physical or knowledge-based as shown in Fig1.1. (This figure has been partly inspired by a brochure from Cadix Corp, Tokyo.) Fingerprints, iris, retina, DNA, face, blood vessels, for instance, are static and biometric. Algorithms which are biometric and dynamic include lip movements, body movements and on-line signature. Schemes which use passwords are static and knowledge-based, whereas methods using magnetic cards and IC cards are physical.

In [1]-[5], we proposed algorithm PPI (pen-position/pen-pressure/pen-inclination) for on-line pen input signature verification. The algorithm considers writer's signature as a trajectory of pen-position, pen-pressure and pen-inclination which evolves over time, so that it is dynamic and biometric. Since the algorithm uses pen-trajectory information, it naturally needs to incorporate stroke number (number of pen-ups/pen-downs) variations as well as shape variations.

In our previous work [1]-[5], genuine signatures were separated from forgery signatures in a linear manner. This paper attempts to perform the nonlinear separation.

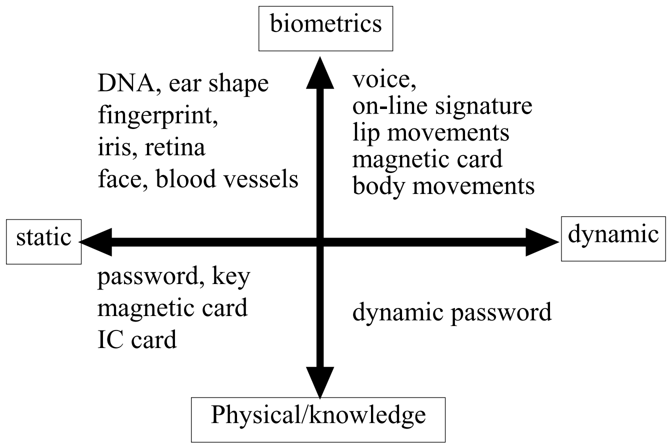

Fig1.1 Authentication Methods

2. THE ALGORITHM

2.1 Overall algorithm

Fig2.1 describes an overall algorithm. A database of signatures is divided into two groups : signatures for learning and signatures for testing.

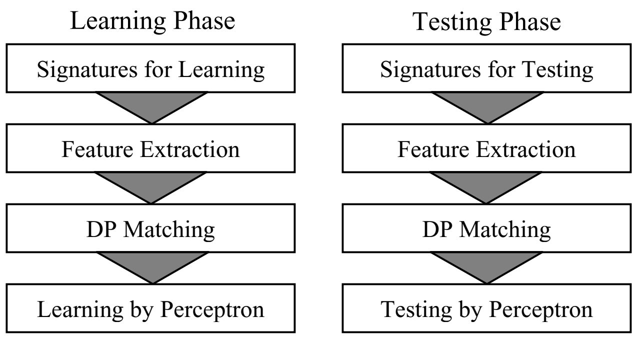

Fig2.1 Overall algorithm

†There are three types of forgery signature – Random forgery, Simple forgery and *Skilled* forgery. A forgery is called Random when forger has no access to genuine signature. A forgery is called Simple when forger knows only the name of the person who write genuine signature. A forgery is called *Skilled* when forger can view and train genuine signature.

2.2 Feature Extraction

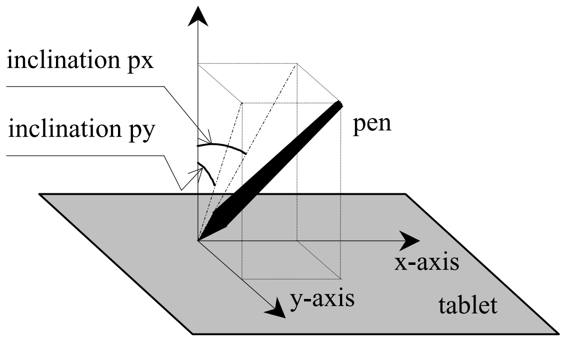

Fig2.2 Raw data from tablet

The raw data available from our tablet (WACOM Art Pad 2 pro Serial) consists of five dimensional time series data :

$$(x(t_i), y(t_i), p(t_i), px(t_i), py(t_i)) \in R^2 \times \{0,1,...,255\} \times R^2 \quad (2.1)$$
$$i = 1,2,...,I$$

where $(x(t_i), y(t_i)) \in R^2$ is the pen position at time t_i , $p(t_i) \in \{0,1,...,255\}$ represents the pen pressure, $px(t_i)$ and $py(t_i)$ are pen inclinations with respect to the x - and y -axis as shown in Fig 2.2.

Define

$$X_g = \frac{\frac{\sum_{i=1}^{I} x(i)}{I} - x_{\min}}{x_{\max} - x_{\min}} \times \alpha \qquad (2.2)$$

$$Y_g = \frac{\frac{\sum_{i=1}^{I} y(i)}{I} - y_{\min}}{y_{\max} - y_{\min}} \times \alpha \qquad (2.3)$$

where $x_{\min}$ and $x_{\max}$ stand for the minimum and the maximum value of $x(t_i)$, $y_{\min}$ and $y_{\max}$ stand for the minimum and the maximum value of $y(t_i)$, respectively, shown in Fig 2.3. α is a scale parameter to be chosen. The pair (X_g, Y_g) can be thought of the *centroid* of signature data (2.1).

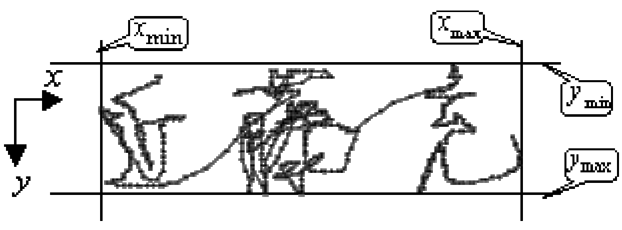

Fig2.3 $x_{\max}$, $x_{\min}$, $y_{\max}$ and $y_{\min}$ of signature

Let

$$V(i) = (dx(i), dy(i))$$
$$:= \left(\frac{x(i) - x_{\min}}{x_{\max} - x_{\min}} \times \alpha - X_g, \frac{y(i) - y_{\min}}{y_{\max} - y_{\min}} \times \alpha - Y_g\right) \quad (2.4)$$
$$i = 1,2,...,I$$

be the relative pen position with respect to the centroid. Then the length $\Delta f(i)$ and the angle $\theta(i)$ of each pen position are given by

$$\Delta f(i) = \sqrt{dx(i)^2 + dy(i)^2} \qquad (2.5)$$
$$i = 1,2,...,I$$

$$\theta(i) = \begin{cases} \tan^{-1} \frac{dy(i)}{dx(i)} & (dx(i) > 0) \\ sign(dy(i)) \times \frac{\pi}{2} & (dx(i) = 0) \\ \tan^{-1} \frac{dy(i)}{dx(i)} + \pi & (dx(i) < 0, dy(i) \geq 0) \\ \tan^{-1} \frac{dy(i)}{dx(i)} - \pi & (dx(i) < 0, dy(i) < 0) \end{cases} \qquad (2.6)$$
$$i = 1,2,...,I$$

Our feature consists of the following five dimensional data

$$(\theta_i, \Delta f_i, p_i, px(t_i), py(t_i)) \in R^2 \times \{0,1,..., N\} \times R^2 \qquad (2.7)$$
$$i = 1,2,...,I$$

where I is the number of the data.

A typical original signature trajectory given by Fig2.4 is converted into the relative trajectory given by Fig2.5.

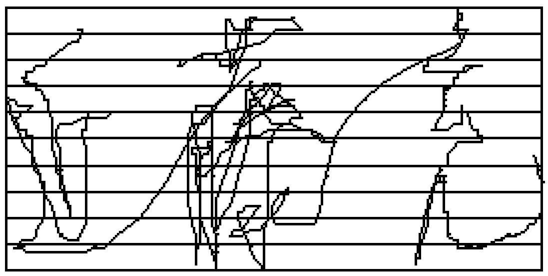

Fig2.4 Original Signature Trajectories

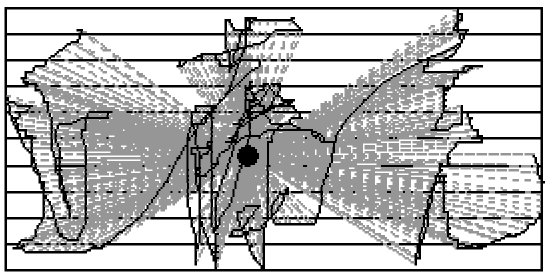

Fig2.5 Relative Trajectories

2.2 Angle-Arc Length Distance Measure

Let

$$(\eta_k, \Delta g_k, q_k, qx(t_k), qy(t_k)) \in R^2 \times \{0,1,\ldots,N'\} \times R^2 \quad k = 1,2,\ldots,K \qquad (2.8)$$

be the feature trajectory of a template signature and consider

$$|\theta_i - \eta_k| S(p_i, q_k) S(\Delta f_i, \Delta g_k) \qquad (2.9)$$

where S is defined by

$$S(u,v) = \begin{cases} 1 & (u = v) \\ |u - v| & (u \neq v) \end{cases} \qquad (2.10)$$

Since 1 is the minimum value of p_i and q_k , function S puts penalty on discrepancies between p_i and q_k . Similarly, $S(\Delta f_i, \Delta g_k)$ puts penalties on the length of each local trajectory length.

The following is our angle arc length distance measure.

$$D1 := \min_{\substack{i_s \le i_{s+1} \le i_s+1 \\ k_s \le k_{s+1} \le k_s+1}} \sum_{s=1}^{S} |\theta_{i_s} - \eta_{k_s}| S(p_{i_s}, q_{k_s}) S(\Delta f_{i_s}, \Delta g_{k_s}) \qquad (2.11)$$

where $i_1 = k_1 = 1,\ i_s = I,\ k_s = K$ are fixed.

Because of the sequential nature of the distance function, Dynamic Programming is a feasible means of the computation.

$$D1(0,0) = 0$$

$$D1(i_s, k_s) = \min \begin{cases} D1(i_s - 1, k_s - 1) + |\theta_{i_s} - \eta_{k_s}| \\ \quad \times S(p_{i_s}, q_{k_s}) S(\Delta f_{i_s}, \Delta g_{k_s}) \\ D1(i_s - 1, k_s) + |\theta_{i_s} - \eta_{k_s}| \\ \quad \times S(p_{i_s}, q_{k_s}) S(\Delta f_{i_s}, 0) \\ D1(i_s, k_s - 1) + |\theta_{i_s} - \eta_{k_s}| \\ \quad \times S(p_{i_s}, q_{k_s}) S(0, \Delta g_{k_s}) \end{cases} \qquad (2.12)$$

2.3 Pen Inclination Distances

Define pen-inclination distances

$$D2 := \min_{\substack{i_{s'} \le i_{s'+1} \le i_{s'}+1 \\ k_{s'} \le k_{s'+1} \le k_{s'}+1}} \sum_{s'=1}^{S'} |px_{i_{s'}} - qx_{k_{s'}}| \qquad (2.13)$$

$$D3 := \min_{\substack{i_{s''} \le i_{s''+1} \le i_{s''}+1 \\ k_{s''} \le k_{s''+1} \le k_{s''}+1}} \sum_{s''=1}^{S''} |py_{i_{s''}} - qy_{k_{s''}}| \qquad (2.14)$$

which are computable via DP also.

2.4 Template Generation

To choose three template signatures, we compute the sum of the distance measure $D = D1 + D2 + D3$ between each of the signatures in a group of genuine signatures for learning and sort them according to their distances, then choose three signatures with the smallest distances. These will be used as templates.

2.5 Nonlinear Separation

Given the set of feature vector $(D1, D2, D3)$, our previous algorithm proposed in [1]-[5], performs linear separation of genuine signatures from forgery signatures. We naturally expect that appropriate nonlinear separation would improve the performance even though our linear scheme already perform reasonably well.

One of the means to perform nonlinear saparation is to prepare a parameterized family $f(D1, D2, D3; w)$ of functions and adjust the parameter vector w in such a way that

$$E := \sum_{i=1}^{M} \| y_i - f(x_i; w) \|^2 \to \min \qquad (2.15)$$

where

$$x_i := (D1_i, D2_i, D3_i)$$
$$y_i := \begin{cases} 1 & \text{(when input signature is genuine)} \\ 0 & \text{(when input signature is forgery)} \end{cases} \qquad (2.16)$$
$$i = 1,2,3,\ldots,M$$

M is the number of training data sets. The family $f(x; w)$ can be, for instance, the perceptrons;

$$f(x; w) := \sigma\left(\sum_{i=1}^{h} \sigma(b_i^T x + c_i)\right) \qquad (2.17)$$

where $\sigma(u) := \dfrac{1}{1 + e^{-u}}$ (2.18)

$$w = (a_1, \ldots, a_h, b_{11}, b_{12}, b_{13}, \ldots, b_{h1}, b_{h2}, b_{h3}, c_1, \ldots, c_h) \qquad (2.19)$$

With this family of functions, one obtains a probability interpretation. Let w^* be an optimal parameter vector with respect to (2.15) and let $x_{test} := (D1_{test}, D2_{test}, D3_{test})$ be a test feature vector. Then the following interpretation is possibile;

$$P(x_{test} \textit{ is genuine}) = f(x_{test}; w^*) \qquad (2.20)$$

thus, a test signature is predicted to be genuine if (2.21) is satisfied, while a test signature is predicted to be forgery if (2.22) is satisfied.

$$0.5 \le f(x_{test}; w^*) \le 1.0 \qquad (2.21)$$
$$0 \le f(x_{test}; w^*) < 0.5 \qquad (2.22)$$

3. EXPERIMENT

This section reports our preliminary experiment using the algorithm described above. Fourteen individuals participated in the experiment. The data was taken for the period of three months. There are 1849 genuine signatures and 3174 *skilled* forgery signatures. Signatures are divided into two groups as shown in Table 3.1.

Table3.1 Database

Signatures for Learning		Signatures for Testing	
genuine	forgery	genuine	forgery
418	1573	1431	1601

Table3.2 shows verification error rates as a function of h (the number of hidden layer). FAR represents False Acceptance Rate and FRR represents False Rejection Rate.

Table3.2 Verification Error Rate

h	FAR (%)	FRR (%)
5	1.42	1.31
6	1.08	1.37
7	1.56	1.56
8	1.42	1.63
9	1.08	1.38
10	1.42	1.00
Average	1.33	1.37

Since our previous scheme [1]-[5] contains a free parameter c to be chosen, direct comparison is impossible, however, the minimum FAR 1.08%, and the minimum FRR 1.00%, of the proposed algorithm are very encouraging.

4. CONCLUSION

A new nonlinear algorithm was proposed for pen-input on-line signature verification which incorporates trajectory of pen-position, pen-pressure and pen-inclinations.

Once the problem is formulated in a probabilistic setting, we can use much more advanced learning schemes such as Hierarchical Bayesian algorithms [6], which we will pursue in a future project.

Our experiment tells us that signatures change with passage of time so that a better algorithm should be able to update w^* with respect to time. This can be done, for instance, by the Online Bayes algorithms [7].

5. REFERENCES

[1] T. Ohishi, Y. Komiya and T. Matsumoto, "An On-Line Pen Input Signature Verification Algorithm", Proc. IEEE ISPACS 2000, Vol.2, pp589-592, 2000.

[2] T. Ohishi, Y. Komiya and T. Matsumoto, "On-line Signature Verification using Pen Position, Pen Pressure and Pen Inclination Trajectories", Proc. IEEE ICPR 2000, Vol. 4, pp547-550, 2000.

[3] D. Sakamoto, T. Ohishi, Y. Komiya, H. Morita and T. Matsumoto, "On-line Signature Verification Algorithm Incorporating Pen Position, Pen Pressure and Pen Inclination Trajectories", Proc. IEEE ICASSP 2001, Vol. 2, pp993-996, 2001.

[4] Y. Komiya, T. Ohishi, T. Matsumoto: "A Pen Input On-Line Signature Verifier Integrating Position, Pressure and Inclination Trajectories", IEICE Trans. Information Systems, vol. E84 - D, No.7, July, 2001.

[5] T. Ohishi, Y. Komiya, H. Morita, T. Matsumoto, "Pen-input On-line Signature Verification with Position, Pressure, Inclination Trajectories", Pros. IEEE IPDPS 2001, pp170, April 2001.

[6] T. Matsumoto, Y. Nakajima, M. Saito, J. Sugi, H. Hamagishi, "Reconstructions and Predictions of Nonlinear Dynamical Systems: A Hierarchical Bayesian Approach", IEEE Trans. On Signal Processing, vol. 49, No. 9, pp.2138-2155, Sep. 2001.

[7] T. Kurihara, Y. Nakada, K. Yosui, T. Matsumoto, "Bayesian On-line Learning: A Sequential Monte Carlo with Importance Resampling", Pros. IEEE International Workshop on Neural Networks for Signal Processing (NNSP 2001), p.163-p.172, Falmouth, Sep.10-12, 2001.