ELSEVIER

Online signature verification using a new extreme points warping technique

Hao Feng *, Chan Choong Wah

School of Electrical and Electronic Engineering, Nanyang Technological University, 56, West Coast Crescent, 02-12, West Bay Condominium 128038, Singapore

Abstract

There are two common methodologies to verify signatures: the functional approach and the parametric approach. In this paper, we propose a new warping technique for the functional approach in signature verification. The commonly used warping technique is dynamic time warping (DTW). It was originally used in speech recognition and has been applied in the field of signature verification with some success since two decades ago. The new warping technique we propose is named as extreme points warping (EPW). It proves to be more adaptive in the field of signature verification than DTW, given the presence of the forgeries. Instead of warping the whole signal as DTW does, EPW warps a set of selected important points. With the use of EPW, the equal error rate is improved by a factor of 1.3 and the computation time is reduced by a factor of 11.
© 2003 Published by Elsevier B.V.

Keywords: Signature verification; Dynamic time warping; Functional approach

1. Introduction

Plamondon and Lorette (1989) categorized the various signature verification methodologies into two types: functional approach and parametric approach. In the functional approach, complete signals ($x(t)$, $y(t)$, $v(t)$, etc.) directly or indirectly constitute the feature set. The two signals, one from a genuine signature and the other from a forgery, are then compared point-to-point. However in the parametric approach, only the parameters abstracted from the complete signals are compared. Though the parametric approach enjoys the advantages of algorithmic simplicity and computation speed, the task of selecting the right set of parameters is not trivial. The comparison based on the complete signals generally yields better results (Plamondon and Lorette, 1989). The two approaches may be applied in different applications, where there are distinct requirements in error rate performance, speed etc. In this paper, our research is focused on the functional approach.

In the functional approach, a straightforward way to compare two signal functions is to use a

---

* Corresponding author. Fax: +65-9070-3130.
 E-mail addresses: haofeng@pmail.ntu.edu.sg (H. Feng), ecwchan@ntu.edu.sg (C.C. Wah).

linear correlation (Plamondon and Lorette, 1989), but a direct computation of the correlation coefficient is not valid due to the following two problems:

1. Difference of overall signal duration.
2. Existence of non-linear distortions within signals.

For a signal function, e.g. $x(t)$, $y(t)$, $v(t)$, it is unlikely that the signal duration is the same for different samples even from the same signer. In addition, for different signings, distortions occur non-linearly within the signals. To correct the distortion, a non-linear warping process needs to be performed before comparison. An established warping technique used in speech recognition is dynamic time warping, or DTW (Sankoff and Kruskal, 1983). For the past two decades, the use of DTW has also become a major technique in signature verification (Hangai et al., 2000; Nalwa, 1997). Though DTW has been applied to the field with some success, it has some drawbacks, as we will explain in details in Section 2.

2. Drawbacks of DTW

The DTW technique is based on the dynamic programming (DP) matching algorithm to find the best matching path, in terms of the least global cost, between an input signal and a template (Sankoff and Kruskal, 1983). The DTW takes a signature sample as the input and aligns it non-linearly with respect to the stored reference signature. The process changes the input signal waveform in two aspects:

1. The end of the input waveform will be aligned with that of the reference.
2. Peaks and valleys will be shifted to align with those of the reference.

An illustration of the waveforms before and after DTW is shown in Fig. 1. In Fig. 1, the top two graphs (a) and (b) are drawn from the reference signature. The middle two graphs (c) and (d) are from the sample signature before DTW, while the bottom two graphs (e) and (f) are from the warped sample signature. Both x and y are independently
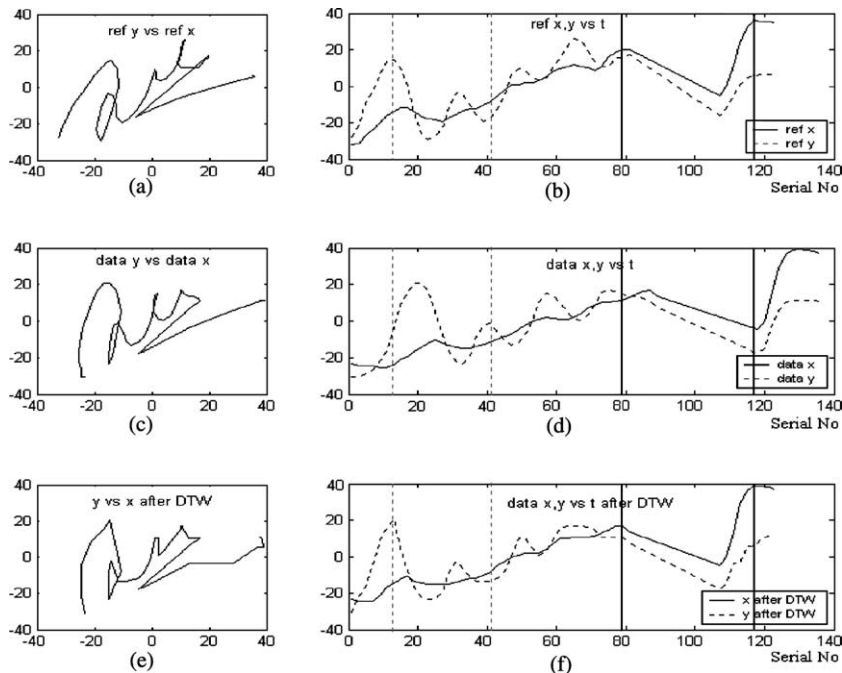

Fig. 1. Waveforms before and after DTW.

warped through DTW. From graphs (b), (d) and (f), one may notice that peaks and valleys of the sample waveforms are shifted to align with those of the reference ones. Some of such shifts of peaks and valleys have been highlighted in graphs (b), (d) and (f) of Fig. 1.

In general, DTW has two main drawbacks when applied in signature verification: (i) heavy computational load, (ii) warping of forgeries. The first drawback is a known problem in speech recognition. This is because DTW performs non-linear warping on the whole signal. The execution time is proportional to the square of the signal size (Sankoff and Kruskal, 1983). To reduce the computation time, Hangai et al. (2000) defined boundary conditions in the DTW matching matrix. However the resultant computation time is still relatively long. It takes on average around 0.4 s, as we will explain in Section 3.1. This problem of the heavy computation load may turn out to be critical if an online signature verification system deals with multi-user requests at the same time.

The second drawback, however, is not well documented in the past literature. When used in the speech recognition, DTW searches a best way to trim the input signal to be more recognizable. However in signature verification, with the presence of the forgeries, forged signals also undergo DTW to be trimmed, so as to be more 'authentic'. Hence some adaptations of the algorithm in the field of signature verification need to be made. This problem can be implicitly addressed by the use of an additional motion measure defined by Sato and Kogure (1982). The motion measure takes the warping path into account. The warping path is less likely to be straight for a forged signal if it undergoes a lot of trimmings during the DTW process. However, the inclusion of the motion measure adds to the complexity of the data classification and the decision-making. Hence it is not used in many recent researches (Nalwa, 1997; Anil et al., 2002).

In this paper, we will introduce a simple solution, without using the motion measure, to adequately address the above two drawbacks. Considering the fact that the DTW process warps every point on the signal, we propose a new warping technique to warp only selective important points of the signal.

3. A new warping technique

The proposed warping technique is called the extreme points warping (EPW). As the name suggests, the technique warps only the extreme points (EPs) of the signal. The EPW process comprises three steps: (i) EPs marking, (ii) EPs matching, and (iii) segments warping.

3.1. Extreme points marking

The EPs are defined as the signal peaks and valleys. We first define a rise-distance, denoted by 'r', as the amplitude from a valley to the following peak. Similarly we define a drop-distance, denoted by 'd', as the amplitude from a peak to the following valley. For any peak (or valley), a rise-distance can be computed at one side of the curve, while a drop-distance can be computed at the other side of the curve. The peak or valley is marked as an EP only if the following condition is met:

$$r \geqslant h_0, \quad d \geqslant h_0 \tag{1}$$

where h_0 is defined as a threshold. Small ripples are not considered as EPs. This is because small ripples are unreliable most of the time. In the project, the threshold h_0 is chosen as one pixel. Hence any ripples with rise or drop distance less than h_0 will not the counted as EPs. Our simple EPs marking algorithm can identify the important peaks and valleys along the signal, while excluding the small ripples, as shown in Fig. 2.
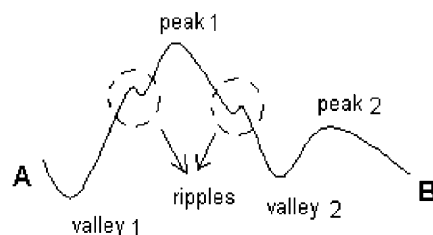

Fig. 2. EPs and ripples.

After identifying the important peaks and valleys as EPs, we will match the EPs of the sample signal and the reference signal correspondingly.

3.2. Extreme points matching

Due to the signature variations, the two sets of EPs are not one-to-one matched. Missing or extra EPs may occur in either signal. Fig. 3(a) and (b) show two torque signals plotted from a reference signature and a sample signature, respectively. The definition of the torque (Nalwa, 1997) will be explained in Section 4. The EPs are marked with '∗' along the two signals.

Through studying the variation phenomena from the collected database, which comprises 25 users and 1000 signatures, we can summarize three types of variations. They are:

1. non-synchronicity for the start point—the first EPs of the two signals may not synchronously start from a peak (or a valley);
2. existence of ripples—a ripple may be found at the start, between the consecutive peak/valley pair and at the end of either signal; however, the occurrence of two or more ripples is rare between a genuine signal and a reference signal;
3. non-synchronicity for the end point—the last EPs of the two signals may not synchronously end up with a peak (or a valley).

We will define a matching algorithm to identify the matching pairs of the corresponding EPs despite the variations mentioned above. Our EPs match-
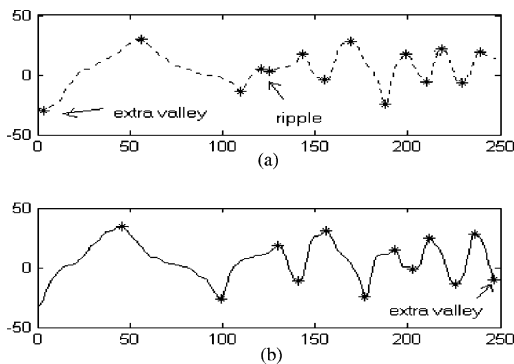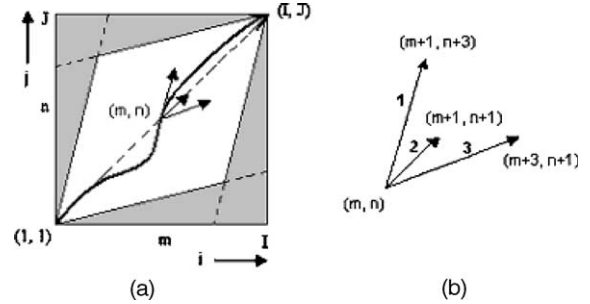

Fig. 4. EPs matching algorithm: (a) the global warping path and (b) the local warping path.

ing algorithm is based on the DP matching technique (Sankoff and Kruskal, 1983). In the classical DP matching process (Hangai et al., 2000), one point on one signal can be matched to any point on the other signal. However in our case, the EPs are alternating peaks and valleys. Hence the corresponding matching pairs of EPs have to be peak-peak or valley-valley matching. We need to introduce some new rules in the DP algorithm to suit the application.

In the EPs matching process, an EP–EP matrix is first established as in Fig. 4(a). In the matrix, the EPs on the reference signal form the horizontal axis and the EPs on the sample signal form the vertical axis. Note in the matrix, the two sets of the EPs need to synchronously start with a peak (or a valley). The global costs at the elements within the unshaded region in Fig. 4(a) are to be computed. The warping path is defined by following the least-global-cost path from $(1,1)$ to (I,J).

Fig. 4(b) shows three local warping paths at the element (m,n). That is to say, if we assume that (m,n) is a matching pair (i.e. the mth EP on the reference signal is matched to the nth EP on the sample signal), the next matching pair can be one of the three: $(m+1,n+1)$, $(m+1,n+3)$ and $(m+3,n+1)$. The $(m+1,n+3)$ means that the two EPs after the nth EP on the sample signal are regarded as a pair of ripple and hence are skipped in the matching. Similarly the matching at $(m+3,n+1)$ means that the two EPs after the mth EP on the reference signal are regarded as a pair of ripple. Hence they are skipped in the matching process. Whenever a pair of EPs is skipped, a skipping cost $S(k,k+1)$ is incurred. It is defined as

Fig. 3. EPs from two signals.

the city block distance between the kth and the $(k+1)$th EPs on the signal.

The EPs on the reference signal can be expressed as two-dimensional data (x_i, y_i), where x_i is the horizontal position of the EP and y_i is the vertical amplitude of the EP. Similarly the EPs on the sample signal are expressed as (x_j, y_j). We define the local distance at the element (i, j) in the matrix as

$$d(i, j) = |x_i - x_j| + |y_i - y_j| \qquad (2)$$

Note in Eq. (2), the city block distance is adopted instead of the Euclidean distance. This is to avoid the situation when a big difference in position or amplitude may over-influence the matching decision. We define $D(i, j)$ as the global distance at the element (i, j) in the matrix. As an initial condition, we have

$$D(1,1) = d(1,1), \quad D(1,3) = d(1,3),$$
$$D(3,1) = d(3,1) \qquad (3)$$

We then compute the global costs at the rest of the elements, progressively using Eq. (4).

$$D(i,j)$$
$$= \min \begin{bmatrix} D(i-1, j-3) + d(i,j) + \rho_s \times S(j-2, j-1) \\ D(i-1, j-1) + 1/2 d(i,j) \\ D(i-3, j-1) + d(i,j) + \rho_s \times S(i-2, i-1) \end{bmatrix}$$
$$(4)$$

where ρ_s is defined as the skipping factor. As different from the normal DTW process (Hangai et al., 2000), we introduce a skipping cost into Eq. (4). The skipping cost is usually very small when skipping a ripple. However it will be much larger if a pair of important peak and valley is misinterpreted as a ripple and skipped. The skipping factor ρ_s is to adjust the influence of the skipping cost in the decision of matching. Through fine-tuning, it is found that $\rho_s = 2$ is an appropriate value, which will be explained in Section 4.

As an example, we will apply the algorithm to match the two sets of EPs (see Fig. 3). Firstly an EP–EP matrix is established in Fig. 5, where the EPs on the reference signal form the horizontal axis, and the EPs on the sample signal form the vertical axis.

As we have explained, the matrix assumes synchronicity of the two sets of EPs. Hence the first
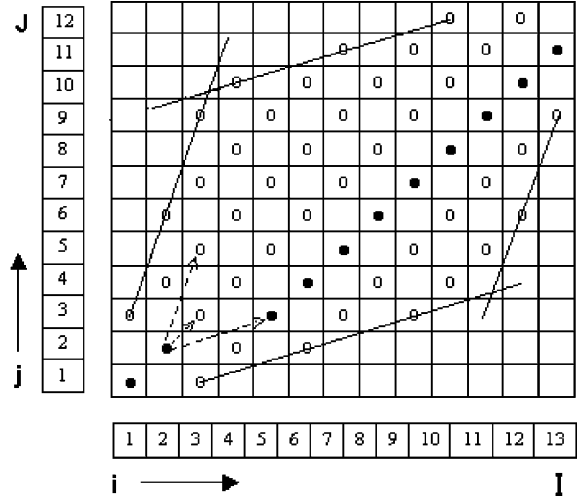

Fig. 5. An example of EPs matching.

EP (an extra valley) on the reference signal is removed, so that the two sets of EPs start synchronously with a peak point (see Fig. 3). One may also remove the first EP (a peak) on the sample signal to make two sets of EPs synchronously start with a valley. Though it is not proper from visual inspection (see Fig. 3), it will still from a valid EP–EP matrix. But the much higher global cost incurred at the elements in the second matrix will indicate that it is not the correct matrix.

In Fig. 5, the circled cells (include the dotted cells) within the defined boundaries indicate all the possible peak–peak or valley–valley matching pairs. Apart from that, peak–valley or valley–peak matching pairs are regarded as mismatching and have been avoided in the definition of the matrix. The global costs are to be computed for every circled cell (include the dotted cells). The dotted cells indicate the correct matching pairs, which follow the least global cost path, as shown in Fig. 5. Fig. 6 shows the result after the EPs matching process, where the matching pairs are ordered in sequence.

3.3. Segments warping

After determining the correct matching pairs of the two sets of the EPs, we will linearly warp the segments within the consecutive EPs. Fig. 7 shows
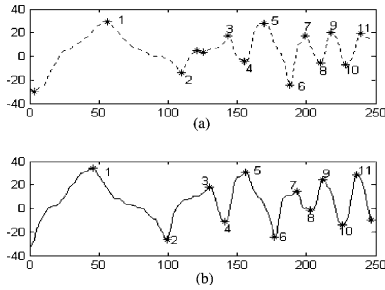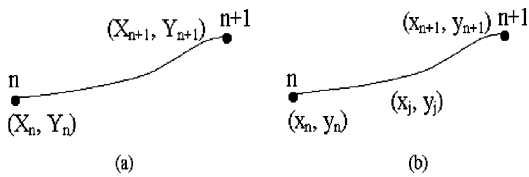
Fig. 6. Result after the EPs matching.

Fig. 7. Linearly warp the segments: (a) the segment of a reference and (b) the segment of a sample.

the two corresponding segments. The point (x_j, y_j) is an arbitrary point on the segment of a sample signal.

In the segment warping process, the sample segment will be linearly stretched to align with the reference segment. The stretching only changes the position of a point (i.e. x), without changing the magnitude (i.e. y). After segment warping, we have
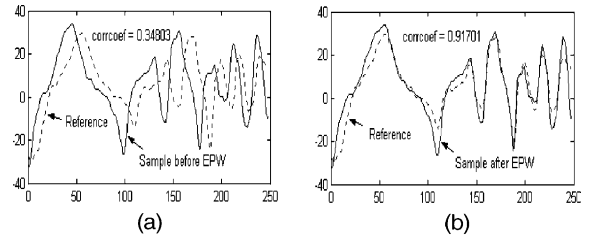

Fig. 8. Result from the segment warping: (a) before EPW and (b) after EPW.

$$x'_n = X_n, \quad x'_{n+1} = X_{n+1} \tag{5}$$

$$x'_j = X_n + (x_j - x_n) \times \frac{X_{n+1} - X_n}{x_{n+1} - x_n} \tag{6}$$

Fig. 8 shows that after segments warping, the correlation coefficient between the reference signal (see Fig. 3(a)) and the sample signal (see Fig. 3(b)) is increased from 34.8% to 91.7%. It is noted that by warping a set of selective EPs, we have achieved the goal of warping the whole signal.

As different from DTW, which warps every point on the signal hence destroys the local curvatures, EPW warps a few EPs and the local curvatures between the EPs are preserved. Fig. 9 shows a comparison of the warping results using EPW for a genuine signal and a forged signal.

As shown in Fig. 9, after EPW, the correlation coefficient for the genuine signal is increased from
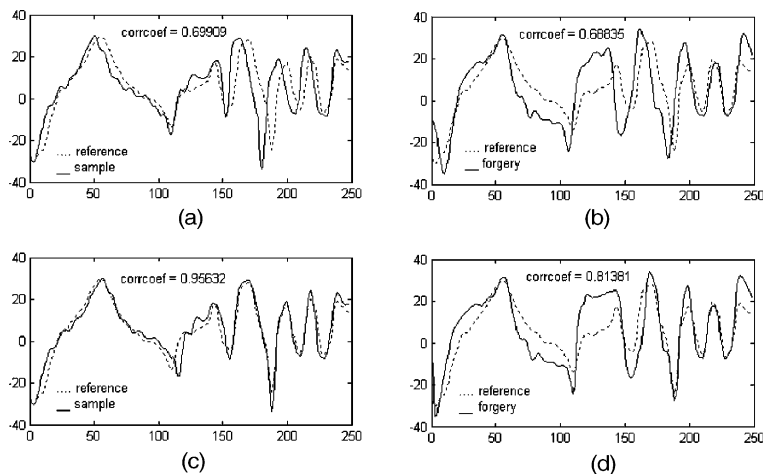

Fig. 9. EPW for the genuine and the forged signals: (a) genuine torque before EPW, (b) forged torque before EPW, (c) genuine torque after EPW and (d) forged torque after EPW.

70% to 95.6%, while for forged signal from 70% to only 81.4%. For the forged signal, even after its EPs are correctly aligned with the corresponding reference EPs, it has a much lower correlation coefficient because of the difference in local curvatures.

4. Verification experiments

To evaluate the proposed new technique, we perform a comparative analysis between EPW and DTW in two aspects: error rate and speed. The comparison is based on the same database and under the same test conditions.

4.1. Database

A signature database comprising 25 users was built. Each user donated 30 genuine signature samples and 10 forgeries in two phases with one-month interval. Overall 1000 signatures were collected and stored in the database.

The forgeries were collected by encouraging participants to mimic each other's signatures as closely as possible. The forger was allowed to view the static images of all the authentic samples and practiced for several minutes before providing the forgeries. Some genuine and forgery samples are shown in Fig. 10.

Fig. 10. Signature samples: (a) genuine 1, (b) genuine 2, (c) forgery 1 and (d) forgery 2.

4.2. Experimental results

To describe the signature shape, the x, y trajectories along the curve are often used. Besides x, y signals, Nalwa (1997) proposed the use of the torque, $T(l)$, and the center of mass, $\bar{x}(l)$, $\bar{y}(l)$ as robust characteristic functions to describe the shape. The three signals are defined below:

$$\bar{x}(l) = \sum_{\lambda=-L}^{L} g(\lambda)x(l+\lambda) \tag{7}$$

$$\bar{y}(l) = \sum_{\lambda=-L}^{L} g(\lambda)y(l+\lambda) \tag{8}$$

$$T(l) = \int_{-L}^{+L} g(\lambda)(y(l+\lambda)\,\mathrm{d}x(l+\lambda) - x(l+\lambda)\,\mathrm{d}y(l+\lambda)) \tag{9}$$

where $\bar{x}(l)$ and $\bar{y}(l)$ are the x, y coordinates along the signature curve, while $g(\lambda)$ is defined as a Gaussian computation window by Nalwa (1997) and 'L' is the window length. To sum up, these five signals x, y, $\bar{x}(l)$, $\bar{y}(l)$, and $T(l)$ will be used for performance evaluation.

Before evaluating the system performance, we will first need to set the value of the skipping factor ρ_s, as shown in Eq. (4). The skipping factor ρ_s is to facilitate the correct matching between EPs. An appropriate value of ρ_s should result in the minimum mismatches between the genuine signal EPs and the reference signal EPs. Table 1 shows the mean correlation coefficients between the EPW-warped signals with the reference signals among the 25 users.

While the skipping cost does little help on the signals like x, y, $\bar{x}(l)$, $\bar{y}(l)$, it improves the EPs matching for the torque signal, which is more complex and fluctuant than x, y etc. (Nalwa, 1997). In the project, a value of 2 is chosen for ρ_s.

Table 1
The effect of the skipping factor

ρ_s	x	y	$\bar{x}(l)$	$\bar{y}(l)$	$T(l)$
0	0.988	0.932	0.993	0.952	0.720
1	0.989	0.961	0.994	0.969	0.903
2	0.989	0.961	0.994	0.968	0.905
3	0.989	0.962	0.994	0.968	0.899

To measure the similarity of the two signals, Nalwa (1997) and Ma and Wijesoma (2000) computed the correlation coefficient between the warped signal and the reference signal. Hangai et al. (2000) and Anil et al. (2002) computed the Euclidean distance between the two signals. In our performance evaluation, we use both to measure the performance. The results in terms of the equal error rate (EER) obtained by adopting DTW and EPW based on the five signals and two measures are presented in Table 2.

It should be noted that the values of EER in Table 2 are obtained based on the signature shape only, without dynamic features involved, since we aim to compare DTW and EPW only. For an effective signature verification system, which adopts both the shape and dynamic features, the resultant EER would be significantly lower.

From Table 2, it is noted that under the same test conditions, the EER of EPW shows improvement over DTW. When the Euclidean distance is used, the average EER of five signals is 33.0% for DTW and 25.4% for EPW, *an improvement of the factor of 1.30*. When the correlation coefficient is used, the average EER is 35.0% for DTW and 27.7% for EPW, *an improvement of the factor of 1.26*. Overall, the EPW shows a moderate improvement of the factor of 1.3 in terms of EER as compared with, using the conventional DTW.

Besides error rate, speed is another attribute used to compare the performance between the DTW and the proposed EPW. The simulation is done using Matlab 6.1 on a Pentium IV 1.9 GHz PC with 256 MB RAM, running Windows 98. Table 3 summarizes the averaged computation time of the five signals for each user.

Table 2
EER for EPW and DTW (%)

Signal	x	y	$\bar{x}(l)$	$\bar{y}(l)$	$T(l)$	Average
Euclidean distance						
DTW	30.5	34.4	29.5	35.1	35.2	33.0
EPW	23.6	26.3	25.5	26.4	25.4	25.4
Correlation coefficient						
DTW	33.5	36.0	33.2	34.8	37.4	35.0
EPW	27.3	27.8	30.6	28.4	2]4.2	27.7

Table 3
Computation times using DTW and EPW for all users (ms)

Users	DTW	EPW
1	415.9	15.8
2	414.5	17.9
3	414.5	17.2
4	416.0	44.0
5	416.8	50.8
6	415.2	28.7
7	415.2	71.1
8	416.3	41.2
9	416.3	30.4
10	416.0	60.8
11	415.1	51.5
12	416.6	23.3
13	414.2	13.4
14	416.5	39.6
15	416	18.2
16	417	49.8
17	417	49.8
18	415.1	15.5
19	416.4	50.9
10	414.7	24.7
21	414.5	22.4
22	415.5	21.0
23	415.8	48.2
24	415.2	23.6
25	416.7	95.1
Average	415.7	37.0

In Table 3, the computation time for EPW is varying among users. This is because the number of EPs identified on a signature signal depends on the signature complexity while the points involved in DTW matching process are roughly the same. The average computation time among the 25 users using DTW is 415.7 ms, while using EPW is only 37.0 ms, *an improvement of the factor of 11*.

From 0.4 to 0.037 s, one may not be able to perceive the difference in the real-time applications. However the improvement would be most evident if it runs on a slower PC and deals with multiple users' requests simultaneously.

5. Conclusion

In this paper, we proposed a new warping technique call EPW to replace the commonly used DTW. Instead of warping the whole signal as DTW does, EPW warps a set of selective points, i.e. the EPs on the signal. Through matching the EPs and

warping the segments linearly, we achieve the goal of warping the whole signal. Since EPW warps only EPs, the local curvatures between the EPs are preserved, which prevents forged signals taking advantages from the warping process. With the adoption of EPW, the EER is improved by a factor of 1.3 over using DTW and the computation time is reduced by a factor of 11. Hence the new technique, EPW, is quite promising to replace DTW to warp signals in the functional approach, as part of a more effective signature verification system.

References

Anil, K.J., Friederike, D.G., Scott, D.C., 2002. Online signature verification. Pattern Recognit. 35 (12), 2963–2972.

Hangai, S., Yamanaka, S., Hammamoto, T., 2000. Writer verification using altitude and direction of pen movement. In: Proceedings of 15th International Conference on Pattern Recognition, vol. 3. pp. 479–482.

Ma, M.M., Wijesoma, W.S., 2000. Automatic online signature verification based on multiple models. In: Proceedings of the IEEE/IAFE/INFORMS Conference on Computational Intelligence for Financial Engineering. pp. 30–33.

Nalwa, V.S., 1997. Automatic online signature verification. Proc. IEEE 85 (2), 215–239.

Plamondon, R., Lorette, G., 1989. Automatic signature verification and writer identification—the state of the art. Pattern Recognit. 22 (2), 107–131.

Sankoff, D., Kruskal, J.B., 1983. Time Warps, String Edits and Macromolecules: The Theory and Practice of Sequence Comparison. Addison-Wesley Publishing. pp. 125–160.

Sato, Y., Kogure, K., 1982. Online signature verification based on shape motion and writing pressure. In: Proceedings of the 6th ICPR. pp. 823–826.