ematical formula recognition," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-6, no. 1, Jan. 1984.

[5] J. Kim and C. C. Tappert, "Handwriting recognition accuracy versus tablet resolution and sampling rate," in *Proc. IEEE Seventh Int. Conf. Pattern Recognition*, Montreal, P.Q., Canada, July 30-Aug. 2, 1984, pp. 917-918.

[6] E. M. Riseman and R. W. Ehrich, "Contextual word recognition using binary diagrams," *IEEE Trans. Comput.*, vol. C-20, no. 4, pp. 397-403, Apr. 1971.

[7] C. R. Blair, "A program for correcting spelling errors," *Inform. Contr.*, vol. 3, pp. 60-67, 1960.

A Comparative Analysis of Regional Correlation, Dynamic Time Warping, and Skeletal Tree Matching for Signature Verification

MARC PARIZEAU AND RÉJEAN PLAMONDON

Abstract—**This correspondence reports on a comparative study of three different signal matching algorithms in the context of signature verification: regional correlation, dynamic time warping, and skeletal tree matching. The algorithm performances are compared in a single experimental protocol over the same database. Algorithm performance is analyzed in terms of verification error rates, execution time, and number and sensitivity of algorithm parameters. Three different script types (normal signatures, handwritten passwords, and initials) and three different signal representation spaces (position, velocity, and acceleration) are considered. Verification errors show that no algorithm consistently out-performs the others in all circumstances. Where significant differences are observed, regional correlation comes first in four out of the five cases. Skeletal tree matching is a close second to regional correlation in one case and comes first in another (dynamic time warping being a close second in this latter case). The complexity of the algorithms varies greatly. Regional correlation is the fastest, followed by dynamic time warping, while skeletal tree matching is very time-consuming. Finally, it is observed that regional correlation is more parameter-sensitive than dynamic time warping.**

Index Terms—**Algorithm comparison, dynamic time warping, regional correlation, signal matching, signature verification, skeletal tree matching.**

I. Introduction

In on-line signature verification, many kinds of information can be extracted while the signature is being executed (as opposed to the off-line problem where the only information available is the signature image itself) [1]. But, in general, there are two groups of methods depending on the type of features used for the classification process, functions or parameters [2]. In the first group, signals measured with an instrumented pen or other apparatus are considered as mathematical time functions whose values directly or indirectly constitute the feature set. In the second group, parameters computed from these signals are used as features. With this last approach, both global and local information can be considered with the greatest efficiency in terms of algorithmic simplicity, computational speed, and memory requirements. However, the problem of selecting the right features is not trivial and methods based on complete signals have so far yielded better results [2].

The main objective of this research was to study the relative performances of three signal comparison algorithms in the context of signature verification: regional correlation [3]-[5], dynamic time warping [6]-[8], and skeletal tree matching [9], [10]. Three comparison criteria are considered: classification error rates, execution time, and number of parameters. Obviously, a good signal matching algorithm needs to yield low classification error rates, but at what cost in terms of execution time? Considering the fact that multiple signal comparisons are usually necessary in a complete signature verification system and that the user does not want to wait more than a few seconds for an answer, the algorithm should not only perform well, but quickly. The number of parameters is also an important aspect of algorithm performance. Different signals from different signers may need different parameter values. Too many parameters needing fine-tuning can significantly degrade the optimal classification performance of an algorithm when there is not enough data to precisely estimate these optimal parameter values.

Regional correlation has been used for signature verification on acceleration and pressure signals obtained from a special accelerometer pen [3]-[5]. Dynamic time warping has also been used for signature verification on position and pressure signals obtained from a digitizer [7], [8]. Tree matching has never been used for signature verification. These previous results are, however, difficult to compare since they originate from different experiments with different databases and experimental protocols. Differences in the quality and types of forgeries are enough to render any comparison meaningless, as are the differences in the sizes of the training and test subsets, in the number of trials permitted and in the type of classifier used, etc. [2]. With a unique database and experimental protocol, we have been able to study and compare the relative merits and pitfalls of these algorithms.

In the next section, the experimental protocol used to compare the algorithms with different script and signal types is presented. Then, in Section III, the particular algorithm implementations are described. Also, each algorithm is characterized in terms of its computational complexity and invariance to time-scale transforms on the input signals. Section IV gives the experimental results and discusses algorithm performances in terms of the three criteria previously mentioned.

II. Experimental Protocol

The experimental protocol is based on signals extracted by a digitizer which measures the position of the pentip along two orthogonal axes of a writing surface. These signals ($x(t)$ and $y(t)$) are sampled at a fixed frequency and stored in a signature database. Information about the state of the pen switch, although available in the database, is not used in this study because it was judged to be unreliable. Most commercial digitizers have been designed for CAD applications, not for handwriting. Pen switch travel is too long and most signers cannot maintain sufficient pressure to activate the switch while the pen is in contact with the paper. New digitizers specially designed for handwriting applications are now available (see for example [11]).

Although portions of signals between a pen-lift and a pen-down may be considered less stable, not removing them will not affect our results because of the relative nature of this study. Indeed, each algorithm will have to deal with the same signals. Furthermore, one can argue that keeping these signal portions is advantageous since, not being visible, they are harder to imitate.

Manuscript received March 31, 1989; revised November 1, 1989. Recommended for acceptance by A. K. Jain. This work was supported by the National Sciences and Engineering Research Council of Canada under Grant A0915 and the Ministère de l'Education du Quebec under Grant CRP2097. M. Parizeau has received scholarships from NSERC, Canada, and FCAR, Quebec.

The authors are with the Department de Genie Electrique, Laboratoire Scribens, Ecole Polytechnique de Montreal, Montreal, P.Q. H3C 3A7, Canada.

IEEE Log Number 9034371.

0162-8828/90/0700-0710$01.00 © 1990 IEEE

A. Database

The digitizer used to build the signature data base was a Summagraphics (model: bit pad one) which sampled coordinates at the rate of 65 Hz and with a resolution of 200 lines/inch. Signature samples were signed along a line parallel to the horizontal axis of the digitizer (x direction). Volunteers were chosen from a population of students and professors at the Ecole Polytechnique, half of them being male and 18% left-handed.

A total of 50 signatures, 50 handwritten passwords, and 50 initials were collected from each of 39 volunteers. Data were collected in five sessions, usually one session per day for a week (10 samples of each script type per session). After database validation, 48 samples of each script type, on average, remained for each volunteer. There were no forgeries.

B. Random Forgeries

Random forgeries were used to simulate real forgeries. A random forgery is defined as an authentic signature from a signer other than the one serving as the reference. Random forgeries were judged sufficient in this study for the following reasons. First, because of the comparative nature of the experiment, the quality of forgeries may influence the absolute results, but should not affect the relative results. Moreover, even a skilled forger cannot easily imitate the dynamics of a signature, especially in the acceleration domain where a random forger might even do a better job. Second, good skilled forgeries are very difficult to obtain, especially in large numbers. However, to make random forgeries more realistic, they were scaled to fit into the same space (box) as the corresponding reference. The rationale behind this normalization is that the most easily forged parameter in a signature is probably its global size.

C. Script Types

Three different types of script were used as signatures to compare the algorithms: the normal signature, a handwritten password and the initials of the subjects. These script types enabled us to verify the effect of the length and stability of the signals on the algorithms. The signature is usually the longest and the most stable. It can be considered as a well-practiced handwritten password. Passwords are shorter. Each volunteer chose his/her own 5- or 6-letter word and was instructed to practice writing it for as long as necessary. Initials are even shorter but generally well practiced.

D. Signal Types

Three different types of signature signals were used to compare the algorithms: position, velocity and acceleration of the pentip. As for the script types, these different signal types enabled us to verify that the experimental results were not an artifact of a particular representation space. The velocity and acceleration signals were computed from the position signal after first- and second-time derivatives. Only normal signatures were considered for these signal types because a previous study has shown that they out-perform passwords and initials [12]. The differentiation algorithm was an eleven-coefficient FIR filter $h(n)$ with a Lanczos window $w(n)$:

$$h(n) = \frac{1}{\pi}\left[\frac{\sin(n\omega_c)}{n^2} - \frac{\omega_c \cos(n\omega_c)}{n}\right], n = 1, 2, \cdots, N \quad (1)$$

$$w(n) = \frac{[\sin(\pi n/N)]}{[\pi n/N]} \quad (2)$$

where $\omega_c = 2\pi f_c$ is the cut-off frequency and N is the number of coefficients. The cut-off frequency was fixed at 16 Hz.

E. Classification Experiment

The first criterion used to evaluate algorithm performances is the verification error. The process by which an error rate is estimated for a particular algorithm and its parameter set is defined by what we shall call here an experiment.

An experiment consists of $2T$ signal comparisons: T comparisons using authentic signals and T comparisons using random forgeries. Thus an experiment is defined by $2T + 1$ signals selected from the signature database: 1 reference signal selected at random within the database, T different authentic test signals selected at random from scripts of the same signer as the reference (but not the reference itself) and T different test signals selected at random from subjects other than the reference (random forgeries). To represent a wide range of variations, the T random forgeries are chosen from T different subjects.

When the $2T$ comparison results are obtained, the verification error is evaluated by searching for the threshold which minimizes the total error. The total error consists of the false rejection rate (type I error) and the false acceptance rate (type II error).

III. Signal Matching Algorithms

The signal matching algorithms were taken from the literature but were adapted for the particular constraints of our experimental protocol [13].

A. Regional Correlation

The idea behind this algorithm is to cut the signals into regions and to correlate corresponding regions over different time lags to find the best possible match for each pair of regions. According to Worthington *et al.* [5], the handwritten signature is more stable during the pen-down parts of the signature and so, in this context, acceleration signals have been segmented using the pen-down pen-up transitions, rejecting the pen-up parts of the signature. However, this method often results in two incompatible lists of regions for the reference and test signals because pen-lifts are not always detected accurately and also because signers are not always consistent.

Our version of the regional correlation algorithm does not use pen-lift information because, based on the hypothesis that the signature is a learned process, that is, it is the result of a ballistic motion with essentially no visual feedback, the pen-up parts of the signature should be almost as stable as the pen-down parts and certainly harder to imitate. The segmentation process used is based only on the observation that handwriting signals tend to fall out of phase beyond a certain time interval. Hence, both the reference and test signals are cut into an equal number of regions. For a given signal, all regions are of the same length and regions are correlated in pairs over all allowed time lags. Regions are initially middle-aligned.

The time complexity of the regional correlation algorithm is $O(n\delta)$, where n is the length of the shortest signal (in sample points) and δ is the length of the regions. If the lengths of the regions are kept constant, then the complexity is $O(n)$.

The regional correlation algorithm is invariant to linear transformation on the input signals because of its similarity measure, linear correlation. Furthermore, this invariance is local to each region. This property means that neither scale nor offset will affect its results.

In terms of time invariance, regional correlation is somewhat limited to the synchronization of the different regions. This synchronization will accommodate minor hesitation in the execution of the signature, but will not compensate for linear or nonlinear time warping within the regions.

B. Dynamic Time Warping

Our particular implementation is inspired by the work of Sakoe and Chiba [6] in the field of speech recognition. Let W be a warping function which maps, in sequence, sample points from a reference signal $a(t)$ to samples points of a test signal $b(t)$:

$$W = \{w(1), w(2), \cdots, w(k), \cdots, w(K)\}$$
$$w(k) = (i(k), j(k)) \quad (3)$$

where i and j represent the time axes of the reference and test signals respectively ($i = 1, \cdots, I$ and $j = 1, \cdots, J$). Let $d(w(k))$ be the distance between two sample points:

$$d(w(k)) = d(i(k), j(k)) = \left|\left[a(i(k)) - \mu_a\right] - \left[b(j(k)) - \mu_b\right]\right| \quad (4)$$

where μ_a and μ_b are the means of signals $a(t)$ and $b(t)$. Then the total distance D_{ab} between signals $a(t)$ and $b(t)$ is defined by the following dynamic programming equations:

$$D_{ab} = g(I, J) \quad (5)$$

if $S = 0$ then

$$g(i,j) = \min \begin{bmatrix} g(i, j-1) \\ g(i-1, j-1) \\ g(i-1, j) \end{bmatrix} + d(i,j) \quad (6A)$$

if $S = 1/2$ then

$$g(i,j) = \min \begin{bmatrix} g(i-2, j-3) + d(i, j-2) + d(i, j-1) \\ g(i-1, j-2) + d(i, j-1) \\ g(i-1, j-1) \\ g(i-2, j-1) + d(i-1, j) \\ g(i-3, j-2) + d(i-2, j) + d(i-1, j) \end{bmatrix} + d(i,j) \quad (6B)$$

if $S = 1$ then

$$g(i,j) = \min \begin{bmatrix} g(i-1, j-2) + d(i, j-1) \\ g(i-1, j-1) \\ g(i-2, j-1) + d(i-1, j) \end{bmatrix} + d(i,j) \quad (6C)$$

$$\vdots$$

$$\text{with } j - \Gamma \le i - \frac{(I-J)}{2} \le j + \Gamma \quad (7)$$

$$\text{and } g(1, 1) = d(1, 1). \quad (8)$$

These equations respect the usual monotonicity, continuity, and boundary conditions [6]. Also, Γ and S are, respectively, the window and slope constraints [6].

The time complexity of dynamic time warping is $O(nm)$, where n is the length (in sample points) of the reference signal and m is the length of the test signal. This is true for a given slope condition. Assuming that both signals are approximately the same length, the complexity is $O(n^2)$. The window condition reduces the computational effort.

The distance measure between two sample points used by the dynamic time warping algorithm is very sensitive to scale. However, an offset on the complete signal will not affect the result because the distance measure is centered around the means. With respect to its invariance to timing fluctuations, the algorithm can accept both linear and nonlinear transformations. This property is controllable by the constraints on the warping function.

C. Skeletal Tree Matching

Tree matching is a method which estimates the distance between two signals by the distance between their corresponding trees. The tree representation of a waveform is a description of the succession of peaks and valleys in the waveform and of their self-embedded structure. In 1985, Cheng and Lu [9] introduced two new types of trees, the skeletal tree and the complete tree. In this study, only skeletal trees are considered. The distance between two trees is estimated in terms of node operations which can be linked directly to transformation of the peaks and valleys.

Four types of operations on tree nodes are defined for transforming one tree into another. The minimum number of these operations is used as a dissimilarity measure. The operations are: father–son splitting and merging, and brother–brother splitting and merging. For skeletal trees, a father–son split or merge corresponds to the growing or shortening of a peak by one quantization interval. A brother–brother split or merge corresponds to the deepening or shallowing of a valley by one quantization interval. The algorithm to find the minimal number of operations is given in [10].

The time complexity of the tree matching algorithm is $O(nM^2)$, where n is the number of nodes in the reference tree and M is the size of a matching table used to accumulate the best results during the processing of the test tree. In practice, M is usually chosen proportional to n, and thus the complexity is $O(n^3)$. The size of the skeletal tree is proportional to the number of quantization levels multiplied by the number of peaks in the signal.

The tree matching algorithm when used with skeletal trees does not vary at all with the timing fluctuations of signers. Indeed, the skeletal tree representation contains only the sequence of peaks, their self-embedding structure and their amplitude. The duration of the peaks is not taken into account.

Regarding the effect of scale and offset on the input signals, skeletal tree matching is not invariant to either, but can adapt easily and without great penalty to a local or global offset. The way the algorithm measures the distance between two signals is like counting the number of elementary deformations of peaks and valleys necessary to transform one signal into the other, and this makes it a very attractive paradigm.

IV. Experimental Results

A. First Criterion: Verification Errors

Algorithm performance is evaluated by verification error, execution time and number of parameters. In Section II, an experiment was defined as 1 reference signal compared to $2T$ test signals. Thus, for each experiment a classification error rate can be observed. By replicating this experimental design K times for one type of signal, the effect of the reference signal can be eliminated by blocking this variable and conducting an analysis of variance. Based on total processing time considerations and on the number of signers in the database, $T = K = 30$ was chosen. It was further decided that each replication of the experimental design would represent a different signer (reference signals were chosen from different signers) so that a wider range of variations between individuals could be considered.

To minimize the effect of algorithm implementation on its execution speed, all of the programming was carried out by the same person. Tree matching was given particular attention because of its time complexity.

Each algorithm has a parameter set: the region length and the maximum allowed time lag for regional correlation, the window length and the slope constraint for dynamic time warping and the quantization step for skeletal tree matching. For the first simulation, the parameters were chosen so as not to impose severe limitations on the timing adaptation of the algorithms. The length of the regions in regional correlation was fixed at 0.7 second as in [5]. The maximum time lag permitted was $\pm 25\%$ of the longest region. The window length for dynamic time warping was also fixed at $\pm 25\%$ of the longest signal. No slope constraint was imposed on the warping function ($S = 0$). The quantization interval for the skeletal tree was set at 1 mm for the position signals. This resolution is considered to be coarse, but acceptable considering the time complexity of the tree matching algorithm and the size of the skeletal tree (proportional to the number of quantization steps).

Fig. 1 shows the total number of errors (type I + type II) for each of the 30 experiments in the case of the position signals of normal signatures. Table I gives the complete average results (in %) for each script and signal type. The reader should remember that these results have only relative significance. No effort has been made to minimize the error rates by choosing the best references or by any other method. The object of this first simulation was to find out whether or not one algorithm was significantly better than the others.

As can be seen from Fig. 1, although the mean results in Table I are somewhat different, individual results of the various experiments show considerable fluctuation. To study these differences more precisely, a variance analysis was conducted with the following model [14]:

$$y_{ij} = \eta + \tau_i + \beta_j + \epsilon_{ij} \tag{9}$$

where y_{ij} is the observed result for the ith algorithm and jth experiment, η is the general average, τ_i is the algorithm effect, β_j is the effect of the experiment, and ϵ_{ij} is the error of the model. This model permits the study of the algorithms while removing the effect of the references. It was validated by testing the underlying assumptions: linearity and additivity. Residuals were found to have zero mean but with a variance proportional to the estimated values. To stabilize the variance, the following transformation was used [14]:

$$x = \sin^{-1}\left[(y/2T)^{1/2}\right] \tag{10}$$

where x is a new metric for the observation y ($2T$ is the number of signal comparisons for one experiment). This transformation has whitened the residuals sufficiently to allow the variance analysis to be conducted.

Let H_0 be the null hypothesis, i.e., the average performances of the three algorithms are equal, let S_a^2 be the variance between algorithms (variance of τ_i) and let S_r^2 be the variance of the residuals (variance of ϵ_{ij}). Then, based on the additive model of (9) (after transformation) and Fisher's randomization theory, H_0 can be tested from the ratio S_a^2/S_r^2 which has an F distribution with 2 degrees of freedom associated with the algorithms and 58 with the residuals.

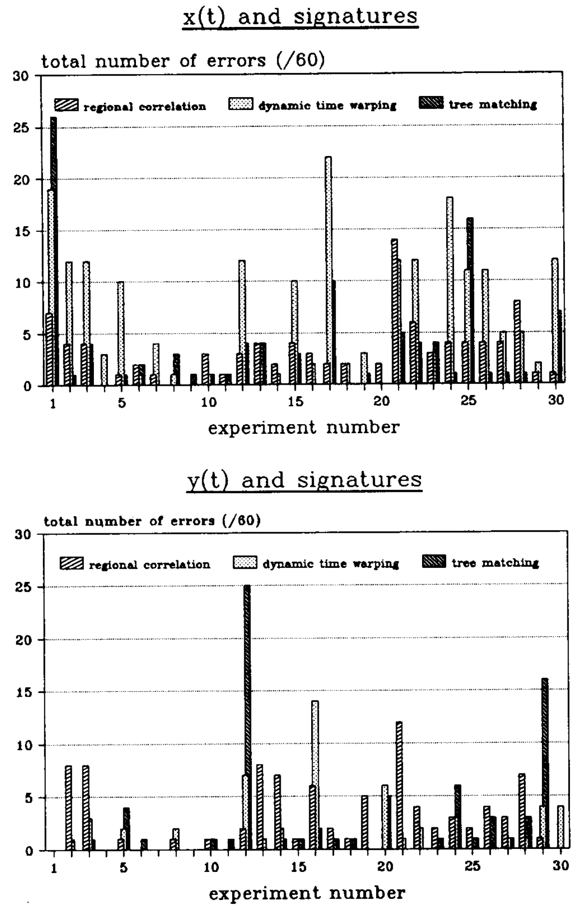

Fig. 1. Experimental results for signatures and position signals.

TABLE I

AVERAGE ERROR RATES FOR THE THREE ALGORITHMS (IN %)

	signatures						passwords		initials	
	x	y	v_x	v_y	a_x	a_y	x	y	x	y
Regional correlation	5.2	4.9	7.0	8.0	10.0	10.5	9.3	7.2	9.0	7.6
Dynamic Time Warping	11.6	3.0	6.2	4.0	16.7	12.2	18.6	5.6	14.3	10.9
Tree Matching	5.7	4.2	6.7	3.3	13.2	11.3	16.6	7.4	16.3	13.1

Table II gives the results of the variance analysis for the various signal and script types. The row $P(F)$ gives the probability for the ratio S_a^2/S_r^2 under the null hypothesis. Fig. 2 illustrates the average performances in Table I with an histogram. The cases where significant differences ($P(F) < 2\%$) between the algorithms have been observed are indicated by an arrow.

For $y(t)$, $v_x(t)$, and $a_y(t)$ of a signature, the null hypothesis cannot be rejected. If real and consistent differences do exist, they

TABLE II
RESULTS OF THE VARIANCE ANALYSIS

	signatures						passwords		initials	
	x	y	v_x	v_y	a_x	a_y	x	y	x	y
Sa^2	495	119	126	464	339	13.3	610	118	522	255
Sr^2	51	67	71.9	55.1	78.8	73.2	40	40	45	74
Sa^2/Sr^2	9.7	1.8	1.8	8.4	4.3	0.2	15.3	3.0	11.6	3.4
P(F)	<.1%	17%	17%	<.1%	1.8%	82%	<.1%	5.8%	<.1%	4.0%

TABLE III
AVERAGE COMPUTATION TIMES FOR ONE SIGNAL COMPARISON (IN SECONDS)

	Regional Correlation	Dynamic Time Warping	Tree Matching
signatures	0.31 (1)	1.4 (4.6)	7.5 (24)
passwords	0.15 (1)	0.38 (2.5)	0.64 (4.3)
initials	0.10 (1)	0.18 (1.7)	0.24 (2.4)

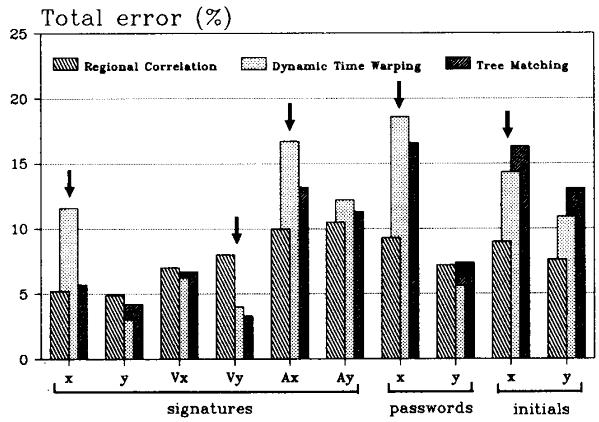

Fig. 2. Average results for the three algorithms.

have not been observed. The cases of $y(t)$ signals of passwords and initials are borderline. Small differences probably do exist but they have not been clearly observed.

Looking at the histogram in Fig. 2, it is possible to make the following observations in terms of the error rate criterion:

1) For $x(t)$ signals, whatever the script type, regional correlation performs either as well as or better than the others. Dynamic time warping performs badly for all script types. This suggests that it cannot handle the moving average which is usually encountered for $x(t)$ signals. Tree matching did almost as well as regional correlation for signatures but its relative performance deteriorated with passwords and initials. This suggests that the algorithm needs a minimum of information (peaks and valleys) to be discriminating.

2) For $y(t)$ signals, whatever the script type, no significant differences were observed between the three algorithms.

3) For cases where significant differences were observed, regional correlation came first in four out of five cases ($x(t)$ for signatures, passwords and initials, and $a_x(t)$ for signatures) but last (by far) in the fifth ($v_y(t)$ of signatures). Dynamic time warping never came first but was close to tree matching for $v_y(t)$ signals of signatures. Tree matching came first for $v_y(t)$, and close to regional correlation for $x(t)$, signals of signatures.

4) For the position representation space, whatever the script type, the algorithms lead to lower error rates with signals measured along the vertical axis.

5) All algorithms are less discriminant with acceleration signals of normal signatures.

B. Second Criterion: Computation Time

Table III gives the average computation time for one signal comparison by each of the three algorithms (the numbers in parentheses are relative to regional correlation). The algorithms were coded in Pascal and the simulations were run on an IBM 4381 computer.

This table shows that important differences exist in the time needed to compare two signals with the three algorithms. For the signatures used in the experiments, it took, on average, 24 times longer to compare two signals with skeletal tree matching than with regional correlation, and for dynamic time warping, it took 4.6 times longer. The effect of the time complexity of the algorithms is clearly visible when looking at the results for the passwords and initials. Indeed, signatures are generally longer than passwords which are themselves longer than initials. For initials, the ratio for tree matching drops to 2.4, and to 1.7 for dynamic time warping.

Regional correlation (which has the smallest time complexity) is thus the fastest algorithm. Dynamic time warping is somewhat slower, however several hardware architectures for VLSI implementation of the algorithms have been developed that offer solutions to this problem. Tree matching is very slow both in absolute and in relative terms. Tree matching requires an application where processing time is not an issue.

C. Third Criterion: Number and Sensitivity of Parameters

The effect of the permitted time lags over the regional correlation algorithm is not critical. Fig. 3 compares the results of simulations for maximum time lags of $\pm 20\%$ and $\pm 30\%$ with the previous results ($\pm 25\%$) for position signals of normal signatures. Under $\pm 20\%$, the permitted time lags are too restrictive and normal time shifts in the signals can no longer be compensated.

Region length is a more important parameter, as shown in Fig. 4. When the regions are too short and need to be shifted to match correctly, the overlap between the regions becomes small and, hence, a relatively greater portion of the signal is lost. Simulations have shown that for the $x(t)$ signals, the region length chosen in this study was appropriate, but for $y(t)$ signals the correct region length would have been around 1.5 seconds. Region length could be a possible explanation for the bad performance observed in the

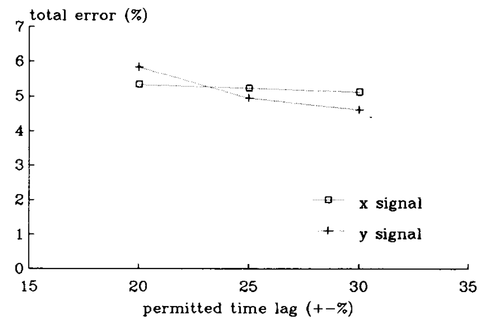

Fig. 3. Total error versus permitted time lags for regional correlation with region length of 0.7 seconds.

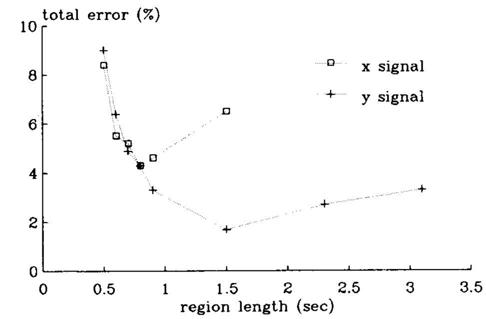

Fig. 4. Total error versus region length for regional correlation with permitted time lags of $\pm 25\%$.

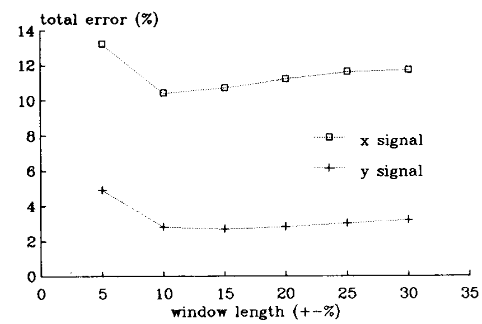

Fig. 5. Total error versus window length for dynamic time warping with no slope constraint.

case of $v_y(t)$ signature signals. The sensitivity of the algorithm over this parameter is an undesirable characteristic since it suggests the need for user-dependent fine-tuning.

For dynamic time warping, the effect of the window length, as in the time lags of regional correlation, is not important. Results of simulations for window lengths of $\pm 5\%$ to $\pm 30\%$ on position signals of signatures are illustrated in Fig. 5. They show that a value around $\pm 15\%$ is optimal.

Up to now, the residual distance obtained by the dynamic time warping algorithm has been divided by the length of the longest of the reference or test signals to normalize the result. Indeed, as opposed to regional correlation which has a result in the range [0, 1], the result of dynamic time warping is usually proportional to the length of the signals because every sample point is considered in the warping function. The logic behind the normalization by the longest signal was that the minimum number of points in the warping function corresponded to the number of points in the longest signal because of the boundary condition. As shown in Fig. 6, dividing by the length of the reference is much better than not dividing by anything or dividing by the longest. This result can be explained by the fact that with the longest signal approach, forged signals longer than the reference tend to reduce the measured distance and thus reduce its discriminating power. Using the length of the reference is a safer approach because authentic signals are more stable, although those longer than the reference will be penalized compared with those that are shorter.

The slope constraint on the warping function is time-consuming. The time penalty is proportional to the number of terms in (6) which changes with the slope constraint. Fig. 7 gives simulation results for $S = 1/2$ and $S = 1$. They show that a small constraint on the

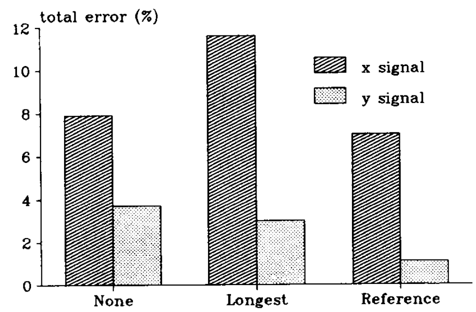

Fig. 6. Total error versus normalization method for dynamic time warping with window length of $\pm 15\%$ and no slope constraint.

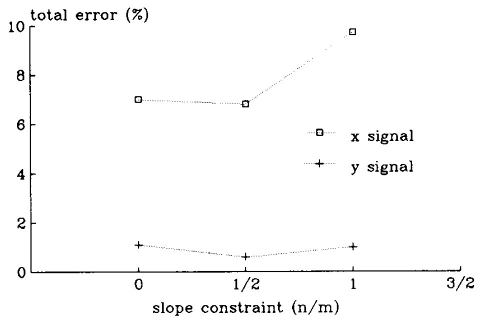

Fig. 7. Total error versus slope constraint for dynamic time warping with window length of $\pm 15\%$ and normalized by the reference.

slope of the warping function ($S = 1/2$) may be beneficial on a relative basis, but this is not an overwhelming factor.

The resolution of the skeletal tree determines the minimum size of peaks considered in the matching process. Although a smaller step should help discriminate between authentic and forged signals, the cubic complexity of the algorithm in conjunction with the linear relation between resolution step and tree size rapidly creates an overwhelming problem. For example, doubling the resolution would require approximately 60 hours of CPU time on the 4381 to execute a complete run of the experimental design on signatures. Also, small peaks may not be very stable which suggests that this resolution should not surpass the size of unstable peaks. For these reasons, skeletal tree matching was not fine-tuned.

V. Conclusion

This correspondence has compared three signal matching algorithms in the context of signature verification. These algorithms all employ very different methods to measure either the similarity or distance between two signals. Algorithm performances were evaluated in terms of classification error rates, execution time, and parameter sensitivity. In terms of classification error rates, experimental results have shown that no algorithm out-performs the others in all circumstances. Significant differences between the algorithms were observed in five cases: $x(t)$, $v_y(t)$, and $a_x(t)$ signals of signatures, and $x(t)$ signals of both handwritten passwords and initials. Of the five cases, regional correlation was first four times and skeletal tree matching once. Both dynamic time warping and skeletal tree matching were a close second once.

In terms of execution time, regional correlation is on average 4.6 times faster than dynamic time warping which, itself, is 5 times faster than skeletal tree matching.

In terms of parameter sensitivity, regional correlation has one critical parameter: region length. For the particular case of $y(t)$ signature signals, its fine-tuning reduced the error rate by half. The allowed time lags can be chosen around $\pm 25\%$ of the longest region. Dynamic time warping has no critical parameter. Window length should be around $\pm 15\%$ and slope constraint should be $1/2$. Results have shown that the residual distance should be normalized by the length of the reference signal. The parameter of skeletal tree matching, the resolution step, is critical to execution time.

Finally, it may be concluded that for an on-line signature verification system based on signals extracted by a digitizer, either position or velocity signals of normal signatures should be considered. If only position signals are used, Regional Correlation could be used on $x(t)$ and $y(t)$, if short processing time is required. Otherwise, skeletal tree matching remains a good candidate. Dynamic tree warping should only be used on $y(t)$ signals. In the velocity domain, the dynamic time warping algorithm is preferred because its parameters are less critical than those of regional correlation. Skeletal tree matching is still a good candiate if processing time is not an issue.

Acknowledgment

We would like to thank D. Lavoie and J. F. Nadeau, who collected the information for the database, and the 39 volunteers who contributed to its construction.

References

[1] Nat. Bureau Standards, "Guidelines on user authentication techniques for computer network control," Federal Information Processing Standards Publ., 1980, 83 pp.

[2] R. Plamondon and G. Lorette, "Automatic signature verification and writer identification: The state of the art," *Pattern Recogn.*, vol. 22, no. 2, 1989.

[3] N. M. Herbst and H. Morrisey, "Signature verification method and apparatus," U.S. Patent 3 983 535, 1976.

[4] N. M. Herbst and C. N. Liu, "Signature verification method and apparatus utilizing both acceleration and pressure characteristics," U.S. Patent 4 128 829, 1978.

[5] T. K. Worthington, T. J. Chainer, J. D. Williford, and S. C. Gundersen, "IBM dynamic signature verification," in *Computer Security*, J. B. Grimson and H. J. Kugler, Eds. Amsterdam, The Netherlands: North-Holland, 1985, pp. 129–154.

[6] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-26, no. 1, pp. 43–49, Feb. 1978.

[7] Y. Sato and K. Kogure, "Online signature verification based on shape, motion, and writing pressure," in *Proc. 6th Int. Conf. Pattern Recognition*, Munich, Oct. 1982, pp. 823-26.

[8] M. Yasuhara and M. Oka, "Signature verification experiment based on non-linear time alignment: A feasibility study," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-17, pp. 212-216, 1977.

[9] Y. C. Cheng and S. Y. Lu, "Waveform correlation by tree matching," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-7, no. 3, pp. 299-305, May 1985.

[10] S. Y. Lu, "A tree matching algorithm based on node splitting and merging," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-6, no. 2, pp. 249-256, Mar. 1984.

[11] B. Blesser and J. R. Ward, "The design of an electronic tablet technology with acceptable performances for handwriting capture and analysis," in *Proc. Third Int. Symp. Handwriting and Computer Applications*, Montreal, July 1987.

[12] M. Parizeau and R. Plamondon, "What types of scripts can be used for signature verification," in *Computer Recognition and Human Production of Handwriting*, R. Plamondon, C. Y. Suen, M. L. Simner, Eds. Singapore: World Scientific, 1989.

[13] M. Parizeau, "Etude comparative d'algorithmes pour mesurer la similitude de signaux dans un contexte de verification de l'identité par la signature manuscrite," mémoire de maîtrise (M.Sc.A.), Ecole Polytechnique de Montréal, Jan. 1987.

[14] G.E.P. Box, W. G. Hunter, and J. S. Hunter, *Statistics for Experimenters*. New York: Wiley, 1978.

[15] R. Plamondon and M. Parizeau, "Signature verification from position, velocity and acceleration signals: A comparative study," in *Proc. 9th Int. Conf. on Pattern Recognition*, vol. I, Rome, Italy, 1988, pp. 260-265.