Online Signature Verification Using Locally Weighted Dynamic Time Warping via Multiple Fusion Strategies

MANABU OKAWA, (Member, IEEE)
Forensic Science Laboratory, Metropolitan Police Department, Tokyo 100-8929, Japan
Corresponding author: Manabu Okawa (e-mail: m_okawa.lab1@m.ieice.org).
This work was supported in part by JSPS KAKENHI Grant Number JP21H04330.

ABSTRACT The process of employing a dynamic signature verification system to verify the writer’s identity is known as online signature verification. It can be used as a security system to verify entrance applications and password substitutes, and as a forensic tool to support expert’s investigation, for example. This study proposes a novel online signature verification system based on a single-template strategy to improve performance in real-world scenarios. It uses discriminative mean signature template sets as well as fusion strategies of multiple local weighting and warping schemes for dynamic time warping (DTW). First, there is the creation of a set of user-specific mean signature templates for each feature using a recent time-series averaging method, i.e., Euclidean barycenter-based DTW barycenter averaging. Then, we acquire a local weighting estimate considering local stability sequences based on multiple and direct matching points between the mean signature templates and references for dependent and independent DTW. Moreover, we derive fusion strategies to calculate locally weighted DTW sets and concatenate them as a feature vector for each warping, followed by constructing a support vector machine (SVM) classifier, respectively. Finally, in the verification phase, we employ the single-template technique to obtain a discriminative fused score using SVMs between the mean template sets and a query sample. The suggested method’s efficiency is demonstrated by extensive experimental results acquired utilizing three public online signature datasets: SVC2004 Task1/Task2, and MCYT-100.

INDEX TERMS Biometrics, forensics, signature verification, time series analysis, fusion strategy, dynamic time warping (DTW), support vector machine.

I. INTRODUCTION

Handwriting is a common tool of communication in our daily lives, and signatures are socially and legally accepted as a form of individual authentication based on the behavioral characteristics and unique features of each person. Recently, online signature verification systems have been used in biometrics [1]–[3] and forensics [4]–[7] because of the widespread adoption of consumer electronics applications and products (e.g., tablets, phablets, and cell/mobile phones).

Parametric and functional approaches are used to extract a set of features from dynamic signatures. The former method depicts signatures as a set of parameters or vectors (e.g., total signature duration, number of pen ups/downs, and average/maximum speed), whereas the latter method represents signatures as time functions (e.g., pen position trajectory, pressure, and velocity). The functional technique has consistently outperformed the parametric approach [1]–[3]. As a result, the functional approach is the emphasis of this research.

The useful method has been adopted for template matching using a distance measure such as dynamic time warping (DTW). Template matching can be classified into multiple- and single-template strategies. The former compares respective distances between a query pattern and each reference using several statistical measures (e.g., max, median, min, and mean), whereas the latter calculates a single template chosen or generated from a reference set. The single-template strategy has numerous advantages, such as speed, security, and tolerance based totally on one-to-one matching. However, its accuracy is lower than that of the multiple-template approach [8].

A novel single-template strategy using a time-series averaging method, namely, Euclidean barycenter-based DTW
barycenter averaging (EB-DBA) [8], and locally/globally weighted DTW (LG-DTW), has been recently proposed [9], [10]. The study [9] adopted multiple matching points (MMPs) [11] for the local weighting estimate and the variable importance obtained using gradient boosting (GB) [12] for the global weighting estimate, whereas the other study [10] used direct matching points (DMPs) [13] instead of the MMPs for the local weighting estimate in LG-DTW. Based on the recent demand for high-speed systems in the big data era, these single-template strategies can decrease calculation complexity while achieving efficient performance, which results in an accuracy improvement.

However, there are some challenges regarding verification performance in real use as follows:

1. Using MMPs and DMPs independently reduces their complementary effects and may result in the loss of detailed local stability information that existed between the mean signature templates and reference sets.

2. The individual use of MMPs and DMPs makes the system difficult to adapt to changes in writing conditions (e.g., device and signal types), template aging, and skilled forgery attacks, all of which occur frequently in real-world scenarios.

Therefore, we have obtained the following solutions to mitigate these challenges:

1. We introduced a modified local weighting scheme for DTW using both MMPs and DMPs (namely, LM-DTW and LD-DTW, respectively) for dependent and independent warping to incorporate more detailed and flexible local stability information and to minimize intra-class discrepancies effectively.

2. To enhance the inter-user variability, we applied the multiple fusion strategies: the representation-level fusion to concatenate the LM-DTW and LD-DTW as a single vector (namely, F-DTW) for each warping, followed by the score-level fusion to combine each score from multiple support vector machine (SVM) classifiers [14], which were constructed for each warping.

Notably, this study is an extension of our previous research [15]. Then, we revised the previous method and conducted additional experiments as outlined below:

1. We updated the global weighting scheme with GB into the score-level fusion with SVM to enhance the discriminative power of F-DTW and to further improve performance.

2. In addition to the previous experiments conducted using an SVC2004 Task1 dataset [16], we conducted comprehensive experiments using two public datasets, SVC2004 Task2 [16] and MCYT-100 [17], to confirm the generalization performance of the proposed method.

The remainder of the paper is organized as follows: In Section II, we review recent online signature verification methods and distance measures relevant to this study. Throughout Section III, we present the proposed online signature verification method. In Section IV, we explain the experimental methods and results. In the penultimate part (Section V), we discuss the findings and their applicability in real-world scenarios. Finally, in Section VI, we present the conclusion.

II. RELATED WORK

A. ONLINE SIGNATURE VERIFICATION

In the past decade, numerous online signature verification systems have been proposed [1]–[3]. We can classify the systems into two main matching methods: model-based and distance-based approaches. Model-based approaches describe data distribution by employing generative models (e.g., Gaussian models [18] and hidden Markov models (HMMs) [19]) and discriminative models (e.g., SVMs [20], convolutional neural networks (CNNs) [21], and recurrent neural networks (RNNs) [22]). Distance-based approaches match query signatures with reference sets by employing distance measures such as DTW [23]. Distance-based approaches are superior in forensic situations with limited data availability for enrollment because a model-based approach would face overfitting problems.

Among various distance-based systems, template matching is commonly used for online signature verification [18]. Template matching includes single- and multiple-template strategies. The single-template strategy has more advantages, such as speed, security, and tolerance, which are in high demand in the current digital era. However, it does not perform as well as the multiple-template strategy [18].

A recent study [8] proposed an effective single-template strategy that uses mean signature templates created using a novel time-series averaging method called EB-DBA to tackle the difficulties. The template creation method expanded the possibility of template matching in online signature verification, as in [9]–[11], [13], [24].

In template matching techniques, distance measures play an important role in calculating the dissimilarity between the templates and the query signatures with different lengths.

B. DISTANCE MEASURES

The distance measures can be categorized into lockstep and elastic distance measures [25]. The lockstep measures are computed by strictly aligning the time-series indices with one-to-one mappings, e.g., Euclidean distances. However, these measures are sensitive to noise, outliers, and basic shape variations with irregular lengths. To overcome these drawbacks, elastic measures such as DTW [23] have been proposed for optimally aligning the indices of a time series with a one-to-many mapping based on dynamic programming.

1) DTW

For \( K \)-dimensional multivariate time series, we can calculate DTW using dependent and independent warping [26]. A
DTW with independent warping (DTW) is individually calculated for each time sequence, assuming that each DTW is a distance measure with a one-dimensional trajectory in one-dimensional Euclidean space. DTW with dependent warping (DTW) is directly derived as a single DTW corresponding to the set of time sequence, assuming the considered K-dimensional time series as a one-dimensional trajectory in K-dimensional Euclidean space. Recent online signature verification studies [8] proposed a novel distance measure with a one-dimensional trajectory in one-dimensional Euclidean space, which adds a multiplicative weight penalty on the distances between the points in the warping path. The cost matrix is updated using this approach to incorporate a modified logistic weight function that assigns an additional weight to the DTW cost function between the reference and test points. A sliding window DTW (SW-DTW) has also been proposed as a relevant approach [28]. SW-DTW adopted the modified DTW cost function using a window function to consider the context by incorporating a weighted average of the neighboring distances.

Some studies have incorporated the weighting scheme in the DTW matching to estimate local stability domains in online signature verification. The study proposed a stability-modulated DTW (SM-DTW) to incorporate the most similar parts between a pair of signatures into the distance measure computed by DTW. Other studies [30], [31] analyzed the most stable domains for each signer using the distance measure computed by DTW. However, DTW is sensitive to noise and outliers in time sequences because it needs to pair all elements of a sequence. To compensate for such drawbacks, some weighting methods for DTW have been proposed.

2) Weighting Schemes for DTW

The weighting schemes for DTW can be classified into local and global weighting schemes. The details are summarized below.

a: Local weighting scheme

A weighting function that adds weight to the DTW cost function between matching points is included in the local weighting scheme.

A previous study [27] proposed a weighted DTW (WDTW), which adds a multiplicative weight penalty based on the distances between the points in the warping path. The cost matrix is updated using this approach to incorporate a modified logistic weight function that assigns an additional weight to the DTW cost function between the reference and test points. A sliding window DTW (SW-DTW) has also been proposed as a relevant approach [28]. SW-DTW adopted the modified DTW cost function using a window function to consider the context by incorporating a weighted average of the neighboring distances.

These studies bring to light the potential for local weighting methods to be applied to DTW. However, most of the previous approaches needed a multiple-template strategy and/or adequate parameter optimization, which results in high computational complexity. To compensate for these difficulties, we have proposed the locally weighted DTW (LWT) to incorporate the local stability of the mean signature template set through MMPS or DMPs and apply the weights to the DTW cost functions.

b: Global weighting scheme

The global weighting scheme incorporates a feature-weighting/selection and applies it to DTW or its variants. Canonical time warping (CTW) [32] combines DTW with canonical correlation analysis (CCA) to calculate spatial projections and determine the linear combinations of variables between two different multivariate sequences. CTW can incorporate feature-weighting/selection, and a dimensionality reduction mechanism for aligning signals of different dimensions. The other research [24] proposed a novel
single-template strategy that uses a global weighting scheme to combine the multiple DTWs while weighting them with the variable importance through GB.

These studies highlighted the limited performance of the single-template strategy in online signature verification. However, the separate/independent use of local or global weighting results in the limited discriminative power of DTW. Therefore, we proposed a novel single-template strategy to overcome these challenges.

III. PROPOSED METHOD

A. OUTLINE

Figure 1 shows the outline of the proposed online signature verification method.

After online signature input, we applied preprocessing to improve quality and extract the common function-based features. We implemented a single-template strategy in the enrollment phase, including mean signature template generation based on the EB-DBA and fusion strategies of multiple local weighting and warping schemes for DTW using the reference set. In the verification phase, we computed the fused score between a test sample and the mean signature templates of a purported user. Finally, the system provides a genuine or forged result for the test sample if the fused score is less or more than a designated threshold for each user.

B. PREPROCESSING

To address the natural fluctuations in signatures regarding writing conditions, we employed a common normalization approach for horizontal and vertical pen coordinates \( \{x(t), y(t)\} \) in [8]–[11], [13], [18], [24] as follows:

\[
\hat{x}(t) = \frac{x(t) - x_g}{x_{\max} - x_{\min}}, \quad \hat{y}(t) = \frac{y(t) - y_g}{y_{\max} - y_{\min}}
\]

(6)

where \( (x_g, y_g) \) is the centroid of a signature; \( \{x_{\min}, y_{\min}\} \) and \( \{x_{\max}, y_{\max}\} \) are the minimum and maximum values of \( \{x(t), y(t)\} \), respectively, for \( t = 1, 2, \ldots, T \).

Hereafter, the pen coordinates \( \{x(t), y(t)\} \) imply the preprocessed functions, \( \{\hat{x}(t), \hat{y}(t)\} \).

C. FEATURE EXTRACTION

This study adopts seven common function-based features [8]–[11], [13], [18], [19], [24], followed by z-score normalization.

- There are three unique features: horizontal and vertical pen coordinates \( x(t), y(t) \), and pen pressure \( p(t) \).
- Four additional features are derived from the original \( x(t), y(t) \) as follows: path-tangent angle \( \theta(t) \), path velocity magnitude \( \nu(t) \), log curvature radius \( \rho(t) \), and total acceleration magnitude \( \alpha(t) \):

\[
\theta(t) = \arctan(\hat{y}(t)/\hat{x}(t)),
\]

(7)

\[
\nu(t) = \sqrt{\dot{x}(t)^2 + \dot{y}(t)^2},
\]

(8)

\[
\rho(t) = \log(\nu(t)/\dot{\theta}(t)),
\]

(9)

\[
\alpha(t) = \sqrt{\dot{\nu}(t)^2 + (\nu(t) \cdot \dot{\theta}(t))^2},
\]

(10)

where the derivatives of discrete-time signals are computed by a second-order regression that removes small noisy variations using the following formula [33]:

\[
f(t) = \frac{\sum_{\epsilon=1}^{2}(f(t + \epsilon) - f(t - \epsilon))}{2 \sum_{\epsilon=1}^{2} \epsilon^2}
\]

(11)

It should be noted that some digital devices (e.g., cell/smart phones) provide original signals without pen pressure. In that case, we select six of the above seven function-based features that do not include \( p(t) \).

D. SINGLE-TEMPLATE STRATEGY

The proposed single-template strategy comprises three steps: 1) the mean signature templates, 2) local weighting estimates, and 3) fusion schemes (Fig. 2). The details of each step are described in the following subsections, referring to the definitions in Section II-B1.

1) Mean Signature Templates

The single-template strategy uses user-specific mean signature templates (i.e., multiple prototypes corresponding to each user’s feature) through EB-DBA [8] to consider intra-user variations among all reference samples.

EB-DBA is an effective time series averaging method that iteratively refines Euclidean barycenter (EB) sequence to minimize its DTW to average target sequences based on expectation–maximization scheme. Concretely, we created a EB sequence of \( N \) references in which the elements are resampled to reach their average length \( I \) equally. We then proceeded with the DBA [34] from the original \( N \) references employing the EB sequence for the initial sequence.

From \( K \) function-based features (i.e., \( K = 6 \) for the SVC2004 Task1; \( K = 7 \) for the SVC2004 Task2 and the MCYT-100 in this study), we obtained \( K \) mean signature templates of lengths \( I \) through EB-DBA for each user.

2) Local Weighting Estimates

We estimate local stability regions in signatures to detect local fluctuations and incorporate intra-user variations into the distance measure. This study adopted MMPS and DMPs to estimate the complementary local stability of the mean signature templates (Fig. 3).

a: MMPS

MMPS [11] detect multiple matching points in DTW trajectories where the mean signature template set and the references are significantly distorted. Consequently, the MMP sequence indicates the local instability of the mean
FIGURE 1. Outline of the proposed online signature verification method.

FIGURE 2. Process of the proposed single-template strategy: (1) mean signature template set creation per feature (pen coordinates “X” and “Y,” pen pressure “P,” path-tangent angle “Ang,” path velocity magnitude “Vel,” log curvature radius “Logcr,” and total acceleration magnitude “Tam”) through EB-DBA (solid black lines) using the five original reference sequences (dashed lines in different colors); (2) local weighting estimates with MMPs and DMPs for independent and dependent paths; (3) fusion schemes in representation- and score-level fusions through the SVM models. Finally, we obtain the Score using the five original reference sequences (dashed lines in different colors).

signature template sequence, and the inverse of the averaged MMP sequence \( \{mmp_i\}_{i=1}^I \) can be considered as the local stability.

For the sake of simplicity, let us assume that there is an \( I \)-length univariate time sequence corresponding to mean signature template \( A \) and the original set of \( N \) references \( B = \{B^n\}_{n=1}^N \) with a \( J_n \)-length univariate time sequence.

Then, the estimation process of the MMP-based local stability can be outlined as follows.

1. We first calculate the standard DTW for each warping between \( A \) and \( B \), and obtain a set of \( N \) optimal warping paths according to the formula below:

\[
W(A, B) = \{W^n(A, B^n)\}_{n=1}^N.
\]

2. Next, we compute \( N \) MMP sequences from \( W(A, B) \), and obtain the averaged MMP sequence as below:

\[
\{mmp_i\}_{i=1}^I = \left\{ \frac{1}{N} \sum_{n=1}^N c^n_i \right\}_{i=1}^I
\]

(12)

where \( c^n_i \) is the cardinality of a set, represented as \( \text{card}\{\cdot\} \), belonging to the \( i \)-th point of \( A \) defined as follows:

\[
c^n_i = \text{card}\{(i^n, j^n) \in W^n(A, B^n) \mid i^n = i\},
\]

By following the steps outlined above, we finally obtained \( I \)-length local weight sequences \( LM_I = \{LM^I_k\}_{k=1}^K \) for independent warping and \( LD_I \) for dependent warping, defined as follows:

\[
LM^I_k = \{lm^I_k(1), lm^I_k(2), \ldots, lm^I_k(i), \ldots, lm^I_k(I)\}
\]

\[
LD^I_k = \{ld^I_k(1), ld^I_k(2), \ldots, ld^I_k(i), \ldots, ld^I_k(I)\}
\]

where \( lm(i) = 1/mmp_i \), \( 0 < lm(i) \leq 1 \); \( lm(i) \) is 1 when all matching point pairs are DMPs and approaches 0 with an increase in the number of MMPs.

b: DMPs

DMPs [13] detect the DTW trajectories’ averaged matching points where one-to-one matching relations exist between the mean signature template set and all references.

The estimation process of the DMP-based local stability is described as follows.

1. In the same way as the MMPs, we first calculate a set of \( N \) optimal warping paths

\[
W(A, B) = \{W^n(A, B^n)\}_{n=1}^N,
\]

where a \( Z_n \)-length warping path is described by

\[
W^n(A, B^n) = \{(p^n_z, q^n_z)\}_{z=1}^{Z_n} \text{ with } 1 \leq p^n_z \leq I, 1 \leq q^n_z \leq J_n, \text{ and } \max(I, J_n) \leq Z_n \leq I + J_n - 1.
\]
(2) Next, we compute the $N$ DMP sequences from the set of warping paths. When the multiplicity of the warping relation for each component is defined as the number of consecutive occurrences of the component index in $W^n(A, B^n)$, the multiplicities corresponding to the respective matching components of $p^n_i$ and $q^n_j$ are obtained as:

$$m^n_i = \text{card}\{(p^n_i, q^n_j) \in W^n(A, B^n) \mid p^n_i = i\}$$

$$m^n_j = \text{card}\{(p^n_i, q^n_j) \in W^n(A, B^n) \mid q^n_j = j\}.$$ 

The $i$th point of $A$, where the multiplicity simultaneously satisfies both $m^n_i = 1$ and $m^n_j = 1$, can be defined as a DMP.

By following the steps outlined above, we finally obtained $I$-length local weight sequences $LD_1 = \{LD^k_1\}_{k=1}^K$ for independent warping and $LD_D$ for dependent warping, defined as follows:

$$LD^k_1 = \{ld^k_1(1), ld^k_1(2), \ldots, ld^k_1(i), \ldots, ld^k_1(I)\},$$

$$LD_D = \{ld_D(1), ld_D(2), \ldots, ld_D(i), \ldots, ld_D(I)\},$$

where

$$\{ld(i)\}_{i=1}^I = \left\{ \frac{1}{N} \sum_{n=1}^N c^n_i \right\}_{i=1}^I.$$ (13)

with

$$c^n_i = \begin{cases} 1 & \text{if an } i\text{th point is a DMP} \\ 0 & \text{otherwise} \end{cases}$$

and $0 \leq ld(i) \leq 1$; $ld(i)$ is 0 and 1 when all matching point pairs are MMPs and DMPs, respectively.

3) Fusion Schemes

After calculating $LM_1$ and $LM_D$ with MMPs and $LD_1$ and $LD_D$ with DMPs, we applied fusion schemes to obtain a discriminative score while maximizing the inter-user variations. This study used two fusion strategies [5]: representation- and score-level fusions.

a: Representation-level Fusion

For representation-level fusion, we obtained a locally weighted DTW with MMPs and DMPs (i.e., LM-DTW and LD-DTW, respectively), followed by concatenating them into a single vector, F-DTW.

To obtain LM-DTW, cost function $d(\cdot, \cdot)$ between two points of the considered time series, as defined in Eqs. (1) and (5), can be rewritten by weighting it by the corresponding local weight sequences, $LM_1$ and $LM_D$, respectively, as follows:

$$d(a_k(i), b_k(j)) = ld^k_D(i) \times (a_k(i) - b_k(j))^2,$$ (14)

$$d(a(i), b(j)) = ld_D(i) \times \sum_{k=1}^{K} (a_k(i) - b_k(j))^2.$$ (15)

Likewise LM-DTW, to obtain LD-DTW, the cost function $d(\cdot, \cdot)$ in Eqs. (1) and (5) can be rewritten using $LD_1$ and $LD_D$ as follows:

$$d(a_k(i), b_k(j)) = ld^k_D(i) \times (a_k(i) - b_k(j))^2,$$ (16)

$$d(a(i), b(j)) = ld_D(i) \times \sum_{k=1}^{K} (a_k(i) - b_k(j))^2.$$ (17)

As a result, we obtained $\{LM-DTW^k_D\}_{K=1}^K$ and $\{LD-DTW^k_D\}_{K=1}^K$ for independent warping; LM-DTW and LD-DTW for dependent warping.

Finally, we obtained F-DTW with independent and dependent warping (namely, $F-DTW_1$ and $F-DTW_D$, respectively), described as follows:

$$F-DTW_1 = \{F-DTW^k_1\}_{i=1}^{2K},$$

$$= \{LM-DTW^k_1, \ldots, LM-DTW^k_1, LD-DTW_1^k, \ldots, LD-DTW_1^k\},$$

$$F-DTW_D = \{F-DTW^k_D\}_{i=1}^{2K},$$

$$= \{LM-DTW_D, LD-DTW_D\}.$$}

b: Score-level Fusion

For score-level fusion, we constructed two SVM classifiers using $F-DTW_1$ and $F-DTW_D$, respectively, and obtained the final score by fusing the scores derived from the classifiers.

SVM [14] is a well-known machine learning classifier that is widely used in writer and signature verification systems [4], [5], [20]. Geometrically, an SVM builds a maximum-margin hyperplane based on the principle of structural risk minimization from statistical learning theory.

When constructing an SVM model, we used positive instances (the intra-user variations between the target user’s mean signature template set and reference set) and negative instances (the inter-user variations between the target user’s mean signature template set and the other users’ mean signature template sets) for each user. For example, we obtained 5 positive and 39 or 99 negative instances by using five genuine signatures as the reference set in the SVC2004 Task1/Task2 datasets and MCYT-100, respectively (Section IV). Then, we employed a linear SVM with the $L_2$-norm penalty and the squared hinge loss using a cost-sensitive learning method to handle the imbalanced class distributions. A grid search is applied to tune the SVM parameter (i.e., the penalty constant $C$).

In the enrollment phase, we constructed two SVM classifiers using $F-DTW_1$ and $F-DTW_D$ (namely, SVM$_1$ and SVM$_D$, respectively). When inputting a query sample to SVM$_1$ and SVM$_D$, we obtain the confidence scores, Score$_1$ and Score$_D$, respectively, each of which is proportional to the signed distance of that sample to the hyperplane.

Finally, we obtained a final score, Score$_{ID}$, by combining the two scores as follows:

$$\text{Score}_{ID} = \text{Score}_1 + \text{Score}_D.$$ (18)
FIGURE 3. Example of the estimation process using toy samples for local stability: (1) the warping relations between the mean signature template “MT” and five references, “R1” to “R5,” with different lengths (“0” to “8” in each sequence revealing the indices), (2) the MMPs, (3) the DMPs, and (4) the local stability sequences, which are applied to DTW calculations and obtain LM-DTW/LD-DTW.

E. OUTPUT

After evaluating the scores between the mean signature template sets of the purported user and the test samples in the verification phase, the system outputs an accept or reject result depending on whether the extent of dissimilarities is below or above the user-specific threshold. In this study, we defined the threshold by analyzing the equal error rate (EER) (Section IV-A2).

IV. EXPERIMENTS

A. METHODS

In real scenarios, skilled forgery detection is a challenging task, especially for forensic document examiners (FDEs) [6], [7].

To overcome such challenges, we conducted experiments using the public online signature datasets: SVC2004 Task1/Task2 [16] and MCYT-100 [17]. These datasets contain the various stylized signatures with highly skilled forgeries collected from other contributors who had sufficient training time to produce valid forgeries. These situations are consistent with the addressed challenge; therefore, we adopted the three datasets in these experiments.

1) Signature Datasets

a: SVC2004 Task1 and Task2

The SVC2004 Task1 and Task2 datasets contain 1,600 signatures, including Western and Asian signatures from 40 users (for a total of 3,200 signatures from 80 users). For each user, both datasets contain 20 genuine and 20 skillfully forged signatures. After sufficient practice, the original writers were advised to provide simple, invented signatures as genuine to avoid privacy issues. The SVC2004 Task1 includes horizontal and vertical pen coordinates with time stamps and pen up/down status, all captured using a digitizing tablet at a sampling rate of 100 Hz. However, the SVC2004 Task2 includes pen pressure, azimuth, and inclination signals. Among the seven function-based features (Section III-C), only the six features without the pen pressure feature are derived from the SVC2004 Task1.

b: MCYT-100

The MCYT-100 dataset consists of 5,000 Western signatures gathered from 100 users. The data include horizontal and vertical pen coordinates, pressure, azimuth, and inclination with time stamps, all of which were captured by a digitizing tablet at a sampling rate of 100 Hz. Each user is represented by 25 samples of both genuine and skillfully forged signatures.

2) Evaluation

Finally, we evaluated the signature verification performance by analyzing the EER with a user-dependent threshold, in which the false rejection and false acceptance rates are equal.

According to the experiments conducted in previous studies and real scenarios, we randomly selected $N = 5$ genuine signatures as the reference set in each experiment. For the test samples in the verification phase, the remaining 15 genuine signatures and 20 skillfully forged signatures were used on the SVC2004 Task1 and Task2; the remaining 20 genuine signatures and 25 skillfully forged signatures were used on the MCYT-100.

To prevent selection bias, we repeated all experiments five times on these three datasets. Finally, we obtained the average EERs.

B. RESULTS

1) Overall Performance

To confirm the effectiveness of the proposed method in template matching, we compared various combinations of
templates and distance measures using three datasets under the same experimental conditions (Section IV-A).

The sets of templates and distance measures used are described below:

- **Template strategies:**
  1. “MT(Mean)”:
     a multiple-template strategy with a mean measure, which was found to perform best among statistical measures [8], after the distances between a test sample and all the references were calculated.
  2. “ST(Rep)”:
     a single-template strategy in which a representative template set was chosen directly from the reference set using the minimum average distance measures from other samples.
  3. “ST(MST)”:
     a single-template strategy with a mean signature template set created through EB-DBA.

- **Distance measures:**
  1. “DTW”:
     the traditional DTW [23] with no weighting for the cost function.
  2. “G-DTW”:
     the previous DTW [24], applying global weighting to combine the multiple DTWs through GB.
  3. “LM-DTW”:
     the recent DTW [11], with the applied MMP-based local stability sequence as the weights for the cost function.
  4. “LD-DTW”:
     the recent DTW [13], with the applied DMP-based local stability sequence as the weights for the cost function.
  5. “LG-DTW”:
     the relevant measure [15], applying local and global weighting to the DTW.

Concretely, after obtaining the F-DTW, we calculated the global weighting factors estimate through GB: \( \{ \alpha_u \}_{u=1}^{2K} \) for the \( \{ \text{F-DTW}^u \}_{u=1}^{2K} \) and \( \{ \beta_v \}_{v=1}^{2} \) for the \( \{ \text{F-DTW}^D \}_{v=1}^{2} \), each of which satisfies \( \sum_{u=1}^{2K} \alpha_u = 1 \) and \( \sum_{v=1}^{2} \beta_v = 1 \), followed by computing the LG-DTW:

\[
\text{LG-DTW} = \sum_{u=1}^{2K} (\alpha_u \times \text{F-DTW}^u) + \sum_{v=1}^{2} (\beta_v \times \text{F-DTW}^D) \tag{19}
\]

- **“Scorep”**: the proposed method, applying the single-template technique to obtain a discriminative fused score through SVMs constructed using F-DTW.

Figure 4 shows the overall performance of the proposed method in terms of EER. As shown in Fig. 4, we deduced the following results:

- Among DTW measures, performance using mean signature templates (“ST(MST)”) is considerably better than the conventional “ST(Rep)” and competitive with the multiple-template strategy (“MT(Mean)”).
- In “ST(MST),” the performance is further improved by applying weighting schemes for DTW.

Notably, the independent use of LM-DTW and LD-DTW reveals data dependency in performance among the datasets; thus, it is rational to use both LM-DTW and LD-DTW in the proposed method. In fact, LG-DTW and Scorep, both of which use F-DTW, outperform the independent use of LM-DTW and LD-DTW.

- Among methods using F-DTW, performance using the score-level fusion with SVM (Scorep) is better than the global weighting scheme with GB (LG-DTW).
- Overall, the proposed single-template strategy (Scorep with “ST(MST)”) achieves the lowest EERs across all datasets.

To confirm statistical significance between the proposed Scorep and the other seven methods, we applied the statistical hypothesis tests. After confirming the global hypothesis tests to be significant on all datasets (i.e., the Friedman test [35] with a significant level of less than 0.001), we applied the Matched-Pairs test [36] along with the Holm method [37]. The Matched-Pairs test determines whether the difference in errors between two methods tested for equivalent subjects on the same dataset is statistically significant. The Holm method is used to adjust the pre-defined significance level for the multiple comparisons. As a result, the Scorep outperformed four methods (i.e., all three DTWs and G-DTW) on all datasets at a significance level of less than 0.05 (i.e., 1.58e-08 ≤ \( p \)-value ≤ 1.37e-03). Hence, we can conclude that there is a significant difference between the results of the proposed method and these four methods.

On the other hand, we cannot confirm the statistical significance between Scorep and LM-DTW or LD-DTW, which depends on the datasets. This result demonstrates that the individual use of LM-DTW and LD-DTW is susceptible to the writing conditions; therefore, it is reasonable to use both LM-DTW and LD-DTW in the proposed method. Additionally, we cannot confirm the statistical significance between Scorep and LG-DTW, both of which use F-DTW, on all datasets. This result indicates that the proposed F-DTW provides sufficient discriminative power in any fusion schemes; thus, it is meaningful to use the score-level fusion with linear SVM in Scorep, where it needs only a few parameters and low computational complexity compared to the global weighting scheme with GB in LG-DTW.

These results confirm that the proposed single-template strategy provides an effective template-matching approach for online signature verification.

2) Comparative Analysis of Weighted DTW

To further confirm the effectiveness of the proposed method, we compared the performance with the previous weighted
DTW methods [27], [28], [32] by applying the single-template strategy under the same experimental conditions. The baselines of the weighted DTW are shown below:

- “WDTW”:
  applying the modified logistic weight function as the weight for the cost function [27].

- “SW-DTW”:
  modifying the cost function by incorporating a weighted average of the neighboring distances using a window function with a width $\delta \in \mathbb{N}$ while weighting with a constant $\alpha \in [0, 1]$ between the cost in amplitude and first order derivative [28].

- “CTW”:
  combining DTW and CCA to allow feature-weighting/selection and dimensionality reduction mechanisms [32].

Notably, WDTW and SW-DTW were calculated using dependent warping based on the findings [8], [26] and their parameters were selected according to previous studies.

Figure 5 compares the EERs of the proposed method with the previous weighted DTW using the single-template strategy with the mean signature templates. As a baseline, we displayed the results of the conventional DTW in Fig. 5, which we confirmed the statistically significant differences with ScoreID (Section IV-B1). As shown in the figure, the proposed method (ScoreID) provides data independence and the lowest EERs compared with conventional methods for all datasets.

To confirm the statistical significance of the performance between the proposed method and the recent WDTW, SW-DTW, and CTW, we applied the statistical hypothesis tests following the previous experiments (Section IV-B1). As a result, the proposed method outperformed all the recent methods on all datasets at a significance level of less than 0.05, where most of the experiments were further...
outperformed at a significance level of less than 0.001 (i.e., $2.60 \times 10^{-12} \leq p$-value $\leq 1.99 \times 0.02$). Hence, we can conclude that there is a significant difference between the results of the proposed method and each of the recent WDTW, SW-DTW, and CTW.

These results confirm that the proposed method provides an effective measure, especially for the single-template strategy in online signature verification.

3) Comparative Analysis of State-of-the-Art Systems
To assess the effectiveness of the proposed single-template strategy in online signature verification, we compared the results of the proposed method's EER with those of state-of-the-art systems.

Tables 1–3 present the results obtained using the SVC2004 Task1/Task2 and MCYT-100, respectively, where only genuine signatures for the enrollment phase and genuine and skillfully forged signatures for the verification phase. In these tables, we displayed EERs of the proposed method as a representative of the single-template strategy following the previous experimental results (Section IV-B1). It should be noted that the comparative analysis of the SVC2004 Task1/Task2 datasets were set up as experiments using not only $N = 5$ but also $N = 10$ as the reference signatures for fair comparisons with the previous studies.

These tables show that the proposed single-template strategy outperforms other recent literature systems in all datasets. The results confirm the effectiveness of the proposed method for online signature verification, even while investigating skilled forgery scenarios.

V. DISCUSSION
In this study, we proposed a novel single-template strategy that provides lower calculation complexity and higher verification performance simultaneously compared with the multiple-template strategy. The main contributions of this study are summarized as follows:

- We adopted the mean signature template creation method with EB-DBA to incorporate the intra-user variations within the reference samples.
- It provides locally weighted DTWs using both MMPs and DMPs (i.e., LM-DTW and LD-DTW, respectively) derived from the intra-user variations between the mean signature template and reference samples for independent and dependent warping to incorporate detailed and flexible local stability information and to minimize intra-class discrepancies effectively.
- It employs multiple fusion strategies to improve inter-user variability: the representation-level fusion, which concatenates LM-DTW and LD-DTW into a single vector, F-DTW, for each warping; and score-level fusion, which combines each score, through SVMs constructed using F-DTW for each warping.

Unlike recent black-box modeling strategies such as deep learning algorithms that require high computational complexity and many training samples, the proposed approach is superior, especially in forensic situations with limited available data [4]–[7].

The proposed method also relies on explainable stepwise methods that support FDEs to explore their differences and similarities and explain the rationale behind their assessment for legal professionals in the decision-making process. For example, it can provide detailed matching between the mean signature template and a query signature (Figs. 1 and 2).

Furthermore, the proposed method can provide visual investigation tools to assist FDEs for forensic analysis in an explainable way. For instance, we can visualize F-DTW$_D$ matching between a mean signature template obtained through EB-DBA and query signatures in the spatiotemporal domain (Fig. 6). In this figure, the farther the distance with higher local weighting between the matching points, the higher the chances of the query signature being a forgery.

Therefore, it is particularly useful for applications, such as forensics and security, in which fairness, accountability, and transparency are critically important.

VI. CONCLUSION
To increase the performance of online signature verification, we devised a unique single-template technique for DTW based on mean signature template sets and fusion strategies of several local weighting and warping schemes.

In the enrollment phase, we used EB-DBA to obtain user-specific mean signature template sets, taking into account intra-user heterogeneity across reference samples. Then, for independent and dependent DTW, we calculated local weighting estimate by analyzing MMPs and DMPs between the mean signature template sets and reference samples, respectively, to incorporate detailed and flexible local stability information and to effectively minimize intra-class discrepancies. To improve inter-user variability, we used the representation-level fusion to concatenate LM-DTW and LD-DTW calculated with MMPs and DMPs into a single vector, F-DTW, for each warping, followed by the score-level fusion to combine each score through SVMs constructed using F-DTW for each warping.

The results of the experiment on the public online signature datasets SVC2004 Task1/Task2, and MCYT-100 proved...
and compensate for the biometrics–forensics divide.

The usefulness of the suggested method for online signature verification.

The proposed explainable stepwise strategy can bridge and compensate for the biometrics–forensics divide.

**REFERENCES**

[1] R. Plamondon, G. Pirlo, E. Anquetil, C. Rémi, H.-L. Teulings, and M. Nakagawa, “Personal digital bodyguards for e-security, e-learning and e-health: A prospective survey,” Pattern Recognit., vol. 81, pp. 633–659, 2018.

[2] M. Díaz, M. A. Ferrer, D. Impedovo, M. I. Malik, G. Pirlo, and R. Plamondon, “A perspective analysis of handwritten signature technology,” ACM Computing Surveys (CSUR), vol. 51, no. 6, pp. 117:1–117:39, 2019.

[3] D. Impedovo and G. Pirlo, “Automatic signature verification in the mobile cloud scenario: survey and way ahead,” IEEE Trans. Emerg. Topics Comput., vol. 9, no. 1, pp. 1–15, 554–568 2021.

[4] M. Okawa and K. Yoshida, “Online writer verification based on forensic expertise: Analyzing multiple patterns by combining the shape and advanced pen pressure information," Jpn. J. Forensic Sci. Tech., vol. 22, no. 2, pp. 61–75, 2017.

[5] M. Okawa, “Synergy of foreground-background images for feature extraction: Offline signature verification using Fisher vector with fused KAZE features,” Pattern Recognit., vol. 79, pp. 480–489, 2018.

[6] C. M. Deviterne-Lapeyre, “Interpol review of questioned documents 2016–2019,” Forensic Sci. Int.: Synergy, vol. 2, pp. 429–441, 2020.

[7] D. Mazzolini, P. Mignone, P. Pavan, and G. Vessio, “An easy-to-explain decision support framework for forensic analysis of dynamic signatures,” Forensic Sci Int.: Digital Investigation, vol. 38, p. 501216, 2021.

[8] M. Okawa, “Template matching using time-series averaging and DTW with dependent warping for online signature verification,” IEEE Access, vol. 7, pp. 81 010–81 019, 2019.

[9] M. Okawa, “Online signature verification using single-template matching through locally and globally weighted dynamic time warping,” IEEE Trans. Inf. & Syst., vol. E103-D, no. 12, pp. 2701–2708, 2020.

[10] M. Okawa, “Modified dynamic time warping with dependent warping for online signature verification,” in Proc. IEEE 3rd Global Conf. Life Sciences and Technologies (LifeTech 2021), pp. 124–125.

[11] M. Okawa, “Online signature verification using a single-template strategy with mean templates and local stability-weighted dynamic time warping,” in Proc. 11th IEEE Int. Workshop Comput. Intell. Appl. (IWCAIA 2019), pp. 83–88.

[12] J. H. Friedman, “Greedy function approximation: A gradient boosting machine,” Annals of statistics, vol. 29, no. 5, pp. 1189–1232, 2001.

[13] M. Okawa, “Time-series averaging and local stability-weighted dynamic time warping for online signature verification,” Pattern Recognit., vol. 112, p. 107699, 2021.

[14] V. N. Vapnik, Statistical Learning Theory. Wiley, 1998.

[15] M. Okawa, “Multiple local weighting scheme for dynamic time warping in online signature verification,” in Proc. IEEE 4th Global Conf. Life Sciences and Technologies (LifeTech 2022), pp. 280–281.

[16] D.-Y. Young, H. Chang, Y. Xiong, S. George, R. Kashi, T. Matsumoto, and G. Rigoll, “SVC2004: First international signature verification competition,” in Biometric Authentication (Proc. ICBA 2004), ser. LNCS 3072, 2004, pp. 16–22.
M. Okawa: Online Signature Verification Using Locally Weighted Dynamic Time Warping via Multiple Fusion Strategies

[17] J. Ortega-Garcia, J. Fierrez-Aguilar, D. Simon, J. Gonzalez, M. Faundez-Zanuy, V. Espinosa, A. Salue, I. Hernaez, J.-J. Igarza, C. Vivaracho et al., “MCYT baseline corpus: A bimodal biometric database,” *IEEE Proceedings-Vision, Image and Signal Processing*, vol. 150, no. 6, pp. 395–401, 2003.

[18] L. Tang, W. Kang, and Y. Fang, “Information divergence-based matching strategy for online signature verification,” *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 4, pp. 861–873, 2018.

[19] J. Fierrez, J. Ortega-Garcia, D. Ramos, and J. Gonzalez-Rodriguez, “HMM-based on-line signature verification: Feature extraction and signature modeling,” *Pattern Recognit. Lett.*, vol. 28, no. 16, pp. 2325–2334, 2007.

[20] C. Gruber, T. Gruber, S. Krinninger, and B. Sick, “Online signature verification with support vector machines based on LCSS kernel functions,” *IEEE Trans. Syst. Man Cybern. B Cybern.*, vol. 40, no. 4, pp. 1088–1100, 2010.

[21] S. Lai, L. Jin, Y. Zhi, Z. Li, and L. Lin, “SynSig2Vec: Forgery-free learning of dynamic signature representations by sigma lognormal-based synthesis,” *IEEE Trans. Pattern Anal. Mach. Intell.*, 2021. (Early Access).

[22] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, and J. Ortega-Garcia, “Exploring recurrent neural networks for on-line handwritten signature biometrics,” *IEEE Access*, vol. 6, pp. 5128–5138, 2018.

[23] H. Sakoe and S. Chiba, “Dynamic programming algorithm optimization for spoken word recognition,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 26, no. 1, pp. 43–49, 1978.

[24] M. Okawa, “Online signature verification using single-template matching with time-series averaging and gradient boosting,” *Pattern Recognit.*, vol. 102, p. 107277, 2020.

[25] A. Mueen and E. Keogh, “Extracting optimal performance from dynamic time warping,” in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, 2016, pp. 2129–2130.

[26] M. Shokoohi-Yeka, B. Hu, H. Jin, J. Wang, and E. Keogh, “Generalizing DTW to the multi-dimensional case requires an adaptive approach,” *Data Min. Knowl. Disc.*, vol. 31, no. 1, pp. 1–31, 2017.

[27] Y.-S. Jeong, M. K. Jeong, and O. A. Omitaomu, “Weighted dynamic time warping for time series classification,” *Pattern Recognit.*, vol. 44, no. 9, pp. 2231–2240, 2011.

[28] F. Dolfado, M. Barandas, R. Matias, R. Martins, M. Carvalho, and H. Gamboa, “Time alignment measurement for time series,” *Pattern Recognit.*, vol. 81, pp. 268–279, 2018.

[29] A. Parizade, M. Diaz, M. A. Ferrer, and A. Marcelli, “SM-DTW: stability modulated dynamic time warping for signature verification,” *Pattern Recognit. Lett.*, vol. 121, pp. 113–122, 2019.

[30] D. Impedovo, G. Pirlo, M. Diaz, and M. A. Ferrer, “Weighted direct matching points for user stability model in multiple domains: A proposal for on-line signature verification,” in *Proc. 15th Int. Conf. Document Analysis and Recognition (ICDAR 2019)*, 2019, pp. 1320–1325.

[31] D. Impedovo, G. Pirlo, and L. Sarcinella, “Signatures’ stability evaluation in a multi-device scenario,” in *Proc. 17th Int. Conf. Frontiers in Handwriting Recognition (ICFHR 2020)*, 2020, pp. 367–372.

[32] F. Zhou and F. De la Torre, “Generalized canonical time warping,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 279–294, 2016.

[33] S. Young, G. Evermann, M. Gales, T. Hain, D. Kershaw, X. Liu, G. Moore, J. Odell, D. Ollason, D. Povey, A. Ragni, V. Valitchev, P. Woodland, and C. Ziehe, *The HTK book (for HTK Version 3.5, documentation alpha version)*. Camb. Univ. Eng. Dep., 2015.

[34] F. Petitjean, A. Ketterlin, and P. Gançarski, “A global averaging method for dynamic time warping, with applications to clustering,” *Pattern Recognit.*, vol. 44, no. 3, pp. 678–693, 2011.

[35] M. Friedman, “A comparison of alternative tests of significance for the problem of m rankings,” *The Annals of Mathematical Statistics*, vol. 11, no. 1, pp. 86–92, 1940.

[36] L. Gillick and S. J. Cox, “Some statistical issues in the comparison of speech recognition algorithms,” in *Proc. 1989 IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP-89)*, 1989, pp. 532–535.

[37] S. Holm, “A simple sequentially rejective multiple test procedure,” *Scand. J. Statist.*, pp. 65–70, 1979.

[38] K. Wang, Y. Wang, and Z. Zhang, “On-line signature verification using wavelet packet,” in *Proc. 2011 Int. Joint Conf. Biometrics (IJCB)*, 2011, pp. 1–6.

[39] K. Barkoula, G. Economou, and S. Fotopoulos, “Online signature verification based on signatures turning angle representation using longest common subsequence matching,” *International Journal on Document Analysis and Recognition (IJDAR)*, vol. 16, no. 3, pp. 261–272, 2013.

[40] A. Alhary, H.-j. Chiang, and S.-i. Kamata, “On-line signature matching based on hilbert scanning patterns,” in *Proc. Int. Conf. Biometrics*, 2009, pp. 1190–1199.

[41] A. Sharma and S. Sundaram, “An enhanced contextual DTW based system for online signature verification using vector quantization,” *Pattern Recognit. Lett.*, vol. 84, pp. 22–28, 2016.

[42] X. Xia, X. Song, F. Luan, J. Zheng, Z. Chen, and X. Ma, “Discriminative feature selection for on-line signature verification,” *Pattern Recognit.*, vol. 74, pp. 422–433, 2018.

[43] A. Sharma and S. Sundaram, “On the exploration of information from the DTW cost matrix for online signature verification,” *IEEE Trans. Cybern.*, vol. 48, no. 2, pp. 611–624, 2018.

[44] Y. Jia, L. Huang, and H. Chen, “A two-stage method for online signature verification using shape contexts and function features,” *Sensors*, vol. 19, no. 8, p. 1809, 2019.

[45] Y. Liu, Z. Yang, and L. Yang, “Online signature verification based on DCT and sparse representation,” *IEEE Trans. Cybern.*, vol. 45, no. 11, pp. 2498–2511, 2015.

[46] N. Liu and Y. Wang, “Template selection for online signature verification,” in *Proc. 19th Int. Conf. Pattern Recognit. (ICPR 2008)*, 2008, pp. 1–4.

[47] K. Cpałka, M. Zalasiewski, and L. Rutkowski, “A new algorithm for identity verification based on the analysis of a handwritten dynamic signature,” *Appl. Soft Comput.*, vol. 43, pp. 47–56, 2016.

[48] N. Sae-Bae and N. Memon, “Online signature verification on mobile devices,” *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 6, pp. 933–947, 2014.

[49] Z. Xia, T. Shi, N. N. Xiong, X. Sun, and B. Jeon, “A privacy-preserving handwritten signature verification method using combinational features and secure KNN,” *IEEE Access*, vol. 6, pp. 46695–46705, 2018.

[50] A. Fischer and R. Plamondon, “Signature verification based on the kinematic theory of rapid human movements,” *IEEE Trans. Human-Mach. Syst.*, vol. 47, no. 2, pp. 169–180, 2017.

[51] D. Guru, K. Manjunatha, S. Manjunath, and M. Somashekara, “Interval valued symbolic representation of writer dependent features for online signature verification,” *Expert Syst. Appl.*, vol. 80, pp. 232–243, 2017.

[52] S. Lai and L. Jin, “Recurrent adaptation networks for online signature verification,” *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 6, pp. 1624–1637, 2019.

[53] S. Lai, L. Jin, L. Lin, Y. Zhu, and H. Mao, “SynSig2Vec: Learning representations from synthetic dynamic signatures for real-world verification,” in *Proc. 34th AAAI Conf. Artificial Intelligence (AAAI-20)*, 2020.

MANABU OKAWA received the B.S. degree in science from Tokyo Metropolitan University, Tokyo, Japan in 2001, the M.S. degree in engineering from Shinshu University, Nagano, Japan in 2008, and the Ph.D. degree in engineering from the University of Tsukuba, Tsukuba, Japan in 2014. At present, he is with the Forensic Document Section of Forensic Science Laboratory of the Metropolitan Police Department, Tokyo, Japan. From 2016 to 2019, he also worked as a part-time lecturer at the University of Tsukuba. His research interests include biometrics, forensics, and data science.