

# A Comparative Study on the Consistency of Features in On-line Signature Verification

Hansheng Lei<sup>1</sup>, Venu Govindaraju

*CUBS, Center for Unified Biometrics and Sensors,  
State University of New York at Buffalo, Amherst, NY 14260, USA*

---

## Abstract

A large number of features have been proposed by researchers for on-line signature verification. However, little work has been done in measuring the consistency and discriminative power of these features. This paper presents a comparative study of features commonly used in on-line signature verification. A consistency model is developed by generalizing the existing feature-based measure to distance-based measure. Experimental results show that the simple features like  $X$ -,  $Y$ - coordinates, the *speed* of writing and the *angle* with the  $X$ -axis are among the most consistent.

*Key words:* On-line Signature verification, Feature selection, Consistency

---

## 1 Introduction

Significant research has been conducted in feature extraction and selection for the application of on-line signature verification [8,9,12–19]. Digitizing devices capture dynamic information of the pen trajectory ( $X$ -,  $Y$ -coordinates, pressure, altitude, etc.). The speed and acceleration of writing can be derived from the coordinate sequences. Researchers have proposed a wide variety of features to discriminate genuine and forged signatures. However, no systematic study has been conducted on which of these features are reliable. A highly desirable property for any feature is that it must have high *consistency*. That is, the feature values from genuine signatures should be close to each other while the distances between genuine and forged features should be relatively large. It is commonly accepted that the dynamic information such as the speed, acceleration or pressure is difficult to forge and thus can be used to distinguish skilled forgeries. However, the dynamic features must consistent. Due to the

---

<sup>1</sup> Corresponding author. E-mail: hlei@buffalo.edu

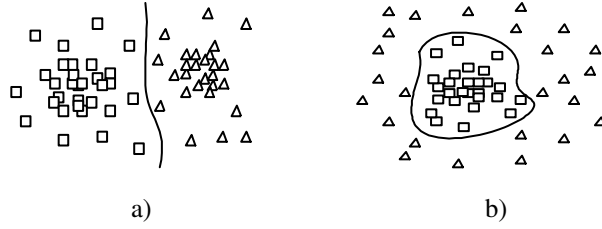


Fig. 1. Difference between two-category and one-category classification problem. a) Regular two-category problem has two clusterings with open decision boundary. b) Signature verification is a one-category problem with closed decision boundary around the clustering genuine signatures.

lack of a consistency measure and benchmark databases, no comparative study is available to evaluate the consistency of features.

In this paper, we describe a consistency model and compare the "consistency" of commonly used features in on-line signature verification. In section 2, we develop the consistency model. In section 3, we briefly describe a fairly exhaustive set of features that we compare. Section 4 presents experimental results and section 5 draws some conclusions.

## 2 Consistency Measure

Signature verification is essentially a one-category classification problem [10], where several genuine samples are available and counter-examples are very few, if any. Thus, it is different from the standard two-category problem where there are two clusters that can be separated by a decision boundary. The boundary is typically an *open* hyper-plane (Fig.1a). However, in the case of signature verification, there is only one cluster of genuine signatures, while forgeries have no clustering characteristic because they have no reason to be close to each other. Therefore, the decision boundary for the one-category classification is a *closed* hyper-plane.

Given few samples of genuine signatures (say, no more than 6), as is often the case in real applications, it is quite challenging to determine the decision boundary no matter what features are used. Because of limited training samples for genuine signatures, the consistency of features becomes extremely important. There are many potential features to choose from [1,4,7] and many new features are continually being invented. This is the main motivation for this paper.

## 2.1 Existing consistency model

A simple consistency measure has been previously defined as [4]:

$$d_i(a) = \frac{|m(a, i) - m(f, i)|}{\sqrt{\sigma^2(a, i) + \sigma^2(f, i)}}, \quad (1)$$

where  $d_i(a)$  denotes the consistency of feature  $i$  for subject  $a$ ;  $m(a, i)$  and  $m(f, i)$  are, respectively, the sample mean among feature  $i$  of genuine and forged signatures;  $\sigma^2(a, i)$  and  $\sigma^2(f, i)$  are the corresponding sample variations of feature  $i$ .

This model has an intuitive geometric meaning. It captures the separability between two clusters of features that are assumed to have Gaussian distribution. While the model is suitable for standard two-category classification, it is inadequate for signature verification primarily because the model is based on features. There are many kinds of features used in signature verification; some are scalar and some are sequential of variable length. The mean  $m(a, i)$  or  $m(f, i)$  of sequential features can not be computed directly. For example, if we define the coordinate sequence itself ( $X$ ,  $Y$  or  $[X, Y]$ ) as a feature, it is difficult to obtain the mean because the sequences are of different lengths. Although sequences can be uniformly re-sampled [1] to be of the same length, the mean of sequences does not carry a practical sense. Moreover, each feature is associated with a certain distance measure, not limited to Euclidean norm. Dynamic Time Warping (DTW) [3] and correlation [1] are also widely used. The existing model implicitly assumes that the features are scalar or vectors of the same length and the distance measure is the Euclidean norm.

## 2.2 Proposed consistency model

The model described above is feature-based, which restricts its application to Euclidean space. If we transform its description from "feature-based" to "distance-based", the model can be enhanced. When we extract and select features to identify true or false objects, we implicitly use the distances between features to distinguish objects instead of using the features themselves. Thus, a distance measure is in some form involved in classification. We generalize model (1) as follows:

$$d_i(a) = \frac{M_{DM_i}(g, f) - M_{DM_i}(g, g)}{\sqrt{\sigma_{DM_i}^2(g, g) + \sigma_{DM_i}^2(g, f)}}, \quad (2)$$

where  $DM_i$  is the distance measure associated with feature  $i$ . Different features may have different distance measures, such as Euclidean norm, DTW and correlation.

$M_{DM_i}(C_1, C_2)$  is the mean of the feature distances (by distance measure  $DM_i$ ) between pairwise objects in class  $C_1$  and class  $C_2$ . Formally,

$$M_{DM_i}(C_1, C_2) = \frac{1}{|C_1||C_2|} \sum_{c_1 \in C_1, c_2 \in C_2, c_1 \neq c_2} DM_i(c_1, c_2), \quad (3)$$

where  $DM_i(c_1, c_2)$  denotes the distance of feature  $i$  between object  $c_1$  and  $c_2$ .  $g$  represents a set of genuine signatures and  $f$  represents set of forged signatures.  $\sigma_{DM_i}^2(g, g)$  denotes the variation of the feature distances (by distance measure  $DM_i$ ) within genuine signatures and  $\sigma_{DM_i}^2(g, f)$  denotes variation of the feature distances (by distance measure  $DM_i$ ) between genuine and forged signatures.

Our model takes into account the fact that the features always come with a certain kind of distance measure. So, we use  $DM_i$  to denote the distance measure associated with feature  $i$ . We compute the mean of feature distances instead of the features themselves because it is the distance that discriminates objects. The inter-class mean distance  $M_{DM_i}(g, f)$  is assumed to be larger than the intra-class mean distance  $M_{DM_i}(g, g)$ . The larger the difference between  $M_{DM_i}(g, f)$  and  $M_{DM_i}(g, g)$ , the higher the discriminability feature  $i$  has. The intra-class distance variance  $\sigma_{DM_i}^2(g, g)$  is an indicator of the consistency of feature  $i$ . Inter-class distance variance is  $\sigma_{DM_i}^2(g, f)$ . If both the inter-class and intra-class distances do not vary much, then it indicates that the feature  $i$  is reliable. As a whole, the value of  $d_i(a)$  measures the consistency of feature  $i$  on subject  $a$ . It should be noted that the same feature can have different consistency values across subjects. Given a data set which consists of a set of subjects, we need to calculate the mean and the standard deviation of the consistency values from equation (2).

### 3 Commonly Used Features

We compare the consistency of 22 commonly used features. We briefly discuss these features and their associated distance measures. Readers are referred to the origin papers for details.

- *Coordinate sequences.*  $X, Y, [X, Y]$  are often called *featureless features*. The lengths of these features are variable. Usually, DTW is used as distance measure because DTW allows elastic matching [3]. It is believed that re-sampling the sequences with uniform arc-length makes the matching more

robust [1]. However, this belief is lack of solid support either theoretically or empirically. We will compare the consistency of the coordinates sequences with and without re-sampling using our proposed model.

- *Speed sequences.*  $V$  (Speed),  $V_x$  (speed of the  $X$ -coordinate) and  $V_y$  (speed of  $Y$ -coordinate) can be derived from sequence  $[X, Y]$  directly by subtracting neighboring points. From the speed sequence, acceleration  $V_a$  can be further derived.
- *Pressure, Altitude, Azimuth.* Pressure is a typical dynamic feature in on-line signatures. Some digitizing devices can capture additional information, such as the Azimuth (the clockwise rotation of cursor about the  $z$ -axis) and Altitude (the upward angle towards the positive  $z$ -axis) [6].
- *Center of Mass  $\bar{x}(l)$  and  $\bar{y}(l)$ , Torque  $\bar{T}(l)$ , Curvature-ellipse  $s_1(l)$  and  $s_2(l)$ .* These five features were defined in [1]. Center of Mass is actually a Gaussian smoothed coordinate sequence. Torque measures the area swept by the vector of the pen movement.  $s_1(l)$  and  $s_2(l)$  measure the curvature ellipses based on moments. The associated distance measure is the cross-correlation (Pearson's  $r$ ) weighted by the consistency of points.
- $\bar{V}$  (average speed),  $\bar{V}_{x+}$  (average positive speed on  $X$ -axis),  $\bar{V}_{y+}$  (average positive speed on  $Y$ -axis),  $T_s$  (total signing duration). Lee et al. [4] list two sets of scalar features. The four features have the highest preference in the first set. The distance measure is the Euclidean norm.
- $\cos(\alpha)$ ,  $\sin(\alpha)$ , Curvature  $\beta$ .  $\alpha$  is the angle between the speed vector and the  $X$ -axis. The three features were proposed in [7]. Two other features  $\delta x$  and  $\delta y$  were also proposed there.  $\delta x$  and  $\delta y$  are the coordinate sequence difference, same as feature # 5 and # 6 in Table 1 respectively.

The features above and the corresponding distance measures are summarized in Table 1. These features are only a small subset of features used in on-line signature verification.

#### 4 Comparison of consistency

The listed SVC database [6] was used to calculate the consistency of the features in Table 1. SVC has two sets of signatures, namely task 1 and task 2. Each signature is represented as a sequence of points, which contains  $X$  coordinate,  $Y$  coordinate, time stamp and pen status (pen-up or pen-down). In task 2, additional information like Azimuth, Altitude and Pressure are available. The information of time stamp and pen status was ignored in our experiments. There are 40 subjects in each task with 20 genuine signatures and 20 skilled forgeries for each subject. Examples of signatures from SVC are shown in Fig. 2.

To compare the consistency of different features, we normalize the raw sig-

Table 1  
Commonly used features and corresponding distance measures.

#	Feature	Dist. Measure
1	$X$ -coordinate: $X$	DTW
2	$Y$ -coordinate: $Y$	DTW
3	Coordinates: $[X, Y]$	DTW
4	Speed: $V$	DTW
5	Speed $X$ : $V_x$	DTW
6	Speed $Y$ : $V_y$	DTW
7	Pressure: $P$	DTW
8	Acceleration: $V_a$	DTW
9	Altitude: $A_l$	DTW
10	Azimuth: $Z_u$	DTW
11	Center of Mass $X$ : $\bar{x}(l)$	Weighted $r$
12	Center of Mass $Y$ : $\bar{y}(l)$	Weighted $r$
13	Torque: $T(l)$	Weighted $r$
14	Curvature-ellipse: $s_1(l)$	Weighted $r$
15	Curvature-ellipse: $s_2(l)$	Weighted $r$
16	Average speed: $\bar{V}$	Euclidean
17	Average positive $V_x$ : $\bar{V}_{x+}$	Euclidean
18	Average positive $V_y$ : $\bar{V}_{y+}$	Euclidean
19	Total signing time: $T_s$	Euclidean
20	Curvature: $\beta$	DTW
21	Angle: $\sin(\alpha)$	DTW
22	Angle: $\cos(\alpha)$	DTW

natures by the same preprocessing: 1) smooth the raw sequence by Gaussian filter; 2) rotate if necessary [2]; 3) normalize the coordinates of each as follows:

$$X = \frac{X - \min(X)}{\max(X) - \min(X)}, Y = \frac{Y - \min(Y)}{\max(Y) - \min(Y)} \quad (4)$$

The feature distances are also normalized. Suppose we apply  $DM_i$  to class  $g$  (genuine signatures) and class  $f$  (forgeries). We first calculate all the pairwise feature distances within class  $g$  by  $DM_i$ . Then, we calculate all the pairwise

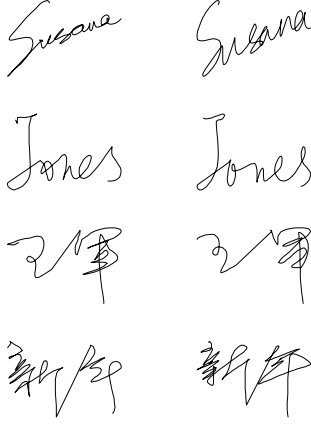


Fig. 2. Examples of signatures from SVC. Signatures in the first column are genuine and those in the second column are forgeries.

feature distances between class  $g$  and  $f$ . Let the maximum feature distance be  $D_{max}$ . Every  $dist$  is normalized by  $e^{\frac{-dist}{2 * D_{max}}}$  to map to a value between 0 and 1. The larger the distance, the closer the mapped value is to 1.

It is possible that same feature may have different consistency values across subjects. Therefore, we calculated the average consistency of each feature and its standard deviation over subjects. Furthermore, to see the relation between feature consistency and its discriminatory ability in verification, the EER (Equal Error Rate) was calculate for each feature. We chose the first 5 genuine signatures from each subject for reference templates and utilized the remaining 35 signatures for verification testing. Given a test signature  $S$ , the distances between feature  $i$  of  $S$  and the feature  $i$  of the 5 genuine signatures were determined. The maximum normalized distance was returned as the similarity output. To determine the EER, the threshold was varied from 0% to 100%. EER is the point where the FRR (False Rejection Rate) equals the FAR (False Acceptance Rate). We have two models of operation: (i) *universal* threshold, we chose the same threshold for all subjects and averaged the error rate and (ii) *user-dependent* threshold, we chose the optimal threshold for each subject.

The results are summarized in Table 2 in the decreasing order of mean consistency value. We have the following observations:

- Although  $Z_u$ (Azimuth)and  $A_l$  (Altitude) have relatively high mean consistency values, they have high standard deviations (3.23 and 4.58 respectively), which means their discriminatory ability is not stable across subjects. Therefore,  $Z_u$  and  $A_l$  are not reliable features. The corresponding high EERs confirm this.
- Curvature-ellipse  $s_1(l)$  and  $s_2(l)$ , Torques  $T(l)$ , Center of Mass  $\bar{x}(l)$  and  $\bar{y}(l)$  are of low consistency and thus high EERs. For instance, the mean consistency of  $s_1(l)$  is only 0.32 and the EER with universal threshold is as

Table 2

Consistency (mean and standard deviation) and EER (universal threshold and user-dependent threshold) of the commonly used features.

Feature	Consistency		EER	
	Mean	Std.	Univ. T.	User. T.
$Z_u$	1.38	3.23	35.06%	26.63%
$V_y$	1.33	0.32	22.06%	11.38%
$V_x$	1.26	0.37	22.65%	16.81%
$V$	1.24	0.41	23.44%	17.63%
$T_s$	1.20	0.81	30.31%	28.18%
$A_l$	1.18	4.60	37.63%	29.06%
$\cos(a)$	1.12	0.28	26.72%	16.19%
$[X, Y]$	1.11	0.18	22.91%	16.83%
$\sin(a)$	1.10	0.25	29.56%	20.63%
$V_a$	1.10	0.32	29.29%	22.50%
$P$	1.06	0.38	36.86%	25.56%
$\beta$	0.99	0.18	28.90%	20.81%
$Y$	0.93	0.19	25.86%	18.68%
$X$	0.78	0.13	29.59%	25.13%
$\bar{y}(l)$	0.64	0.14	28.81%	19.19%
$\bar{x}(l)$	0.60	0.12	29.59%	23.00%
$T(l)$	0.53	0.15	33.63%	25.68%
$\bar{V}$	0.48	0.18	34.50%	31.94%
$\bar{V}_{y+}$	0.46	0.17	36.69%	33.88%
$\bar{V}_{x+}$	0.40	0.17	36.69%	34.00%
$s_2(l)$	0.36	0.03	43.86%	42.19%
$s_1(l)$	0.32	0.03	43.96%	42.13%
Average	0.89	0.57	31.30%	24.91%

- high as 43.96%. We can conclude that these features are not good enough for skilled forgeries, though they might be sufficient to detect random forgeries.
- Scalar features like  $\bar{V}_{x+}$  and  $\bar{V}_{y+}$  are too simple to carry enough discriminatory information. Their EERs are also relatively high. For  $\bar{V}_{x+}$ , the EER even with user-dependent threshold is 34.00%. Scalar feature  $T_s$  also has



high EER because its standard deviation of consistency is relatively high (0.82) compared to the average of all features (0.57). So, scalar features could be used to prune random forgeries but are not very reliable for accepting genuine ones.

- Pressure ( $P$ ) is one of the most commonly used dynamic features. Since pressure is invisible, it is difficult to forge. However, its mean consistency is not very high (1.06), which means that the pressure of the same person’s writing unconsciously varies. Its relatively high EER indicates usage of pressure comes as follows: large variation in pressure does not necessarily mean that the signature is forgery, but very similar pressure pattern is a strong indication that the signature is genuine.
- $V_y$  (speed of the  $Y$ -coordinates) and  $V_x$  (speed of the  $X$ -coordinates) are among the most consistent features. Their mean consistency are 1.33 and 1.26 respectively, while the EERs with universal threshold are both around 22%, which is low compared to the average (31.30%).  $V$  (speed),  $X$ ,  $Y$ ,  $[X, Y]$ ,  $\cos(a)$  and  $\sin(a)$  are also reliable.
- EER has a negative relation with the level of consistency, although the EER does not strictly increase while the mean consistency decreases. The standard deviation of consistency also impacts the verification performance. The consistency of a feature is positively related to its performance of verification.

It must be noted that all the EERs in Table 2 are not low enough for real applications because we used only one feature at a time for verification. How to combine these features optimally is still an open problem. Further experiments on real and larger signature databases are necessary to claim the consistency of any given feature.

There is an unsubstantiated claim in on-signature verification community that the curve of the signature should be re-sampled with uniform arc-length [1,2,7]. Experiments were conducted to verify the claim. We re-sampled all the signatures to be of length  $N$  and calculated the consistency of features and corresponding EERs. The results are summarized in Table 3 with  $N = 200$ . We varied  $N$  and found the results had no significant change. Comparing the consistency values and EERs in Table 2 and Table 3, we can see that re-sampling does not necessarily improve performance. On the contrary, the performance is jeopardized to some degree. For example, the mean consistency of  $V_y$  decreased from 1.33 to 0.88 and EER increased from 22.06% to 32.36% after re-sampling.

Table 3

Consistency and EER of features with uniform arc-length re-sampling.

Feature	Consistency		EER	
	Mean	Std.	Univ. T.	User. T.
$V_y$	0.88	0.24	32.36%	20.81%
$V_x$	0.87	0.30	30.44%	21.00%
$V$	1.14	0.31	27.94%	17.63%
$\cos(a)$	1.08	0.24	29.41%	20.25%
$[X, Y]$	1.05	0.58	26.63%	20.94%
$\sin(a)$	0.81	0.27	32.56%	23.69%
$\beta$	0.89	0.24	32.19%	24.75%
$Y$	0.87	0.27	29.94%	21.56%
$X$	0.79	0.44	34.03%	29.63 %

## 5 Summary

To study the consistency of a variety of features for on-line signature verification, we propose a new consistency model. Experiments were conducted to systematically compare the consistency of features based on the proposed model. Comparative results show that the consistency of feature is positively related to the feature’s performance in signature verification. The results also show that some features such as the *speed* ( $V_y$ ,  $V_x$ ,  $V$ ), the *coordinate sequence* ( $X$ ,  $Y$ ,  $[X, Y]$ ) and the angle  $\alpha$  ( $\cos(a)$ ,  $\sin(a)$ ) have high consistency and thus are reliable. We found that uniformly re-sampling the sequences does not necessarily increase verification performance.

## References

- [1] V. Nalwa. *Automatic on-line signature verification*. Proceedings of the IEEE, 85(2): pp. 213-239, 1997.
- [2] M. Munich, P. Perona. *Visual identification by signature tracking*. IEEE Trans. on Pattern Analysis and Machine Intelligence, 25(2): pp. 200 - 217, 2003.
- [3] R. Martens, L. Claesen. *On-line signature verification by dynamic time-warping*. In the 13th International Conference on Pattern Recognition, pp. 38-42, 1996.
- [4] L. Lee, T. Berger, E. Aviczer. *Reliable On-Line Human Signature Verification Systems*. IEEE Trans. on Pattern Analysis and Machine Intelligence , pp. 643-647, 1996.

- [5] T. Rhee, S. Cho, J. Kim. *On-Line Signature Verification Using Model-Guided Segmentation and Discriminative Feature Selection for Skilled Forgeries*. In the 6th International Conference on Document Analysis and Recognition, pp. 645 - 649, 2001.
- [6] SVC. *The First International Signature Verification Competition*. <http://www.cs.ust.hk/svc2004/>, 2004
- [7] A.Jain, F. Griess, S. Connell. *On-line Signature Verification*. Pattern Recognition, 35(12): 2963–2972, 2002.
- [8] F. Leclerc, R. Plamondon. *Automatic Signature Verification: the State of the Art 1989-1993*. International Journal of Pattern Recognition and Artificial Intelligence, 8(3): pp. 643-660, 1994.
- [9] R. Plamondon, F. Leclerc. *Automatic Signature Verification and Writer Identification the State of the Art*. Pattern Recognition, 22(2): pp. 107-131, 1989.
- [10] D. Tax. *One-class Classification*. PhD thesis, Delft University of Technology, the Netherlands, 2001.
- [11] R.Plamondon. *The Design of On-line Signature Verification System: From Theory to Practice*. International Journal of Pattern Recognition and Artificial Intelligence, 8(3), pp.795-811, 1994.
- [12] J. Brault and R. Plamondon. *Segmenting Handwritten Signatures at Their Perceptually Important Points*. IEEE Trans. On Pattern Analysis and Machine Intelligence, 15(9), pp.953-957, 1993.
- [13] T. Sebastian, P. Klein and B. Kimia. *On Aligning Curves*. IEEE Trans. On Pattern Analysis and Machine Intelligence, 25(1), 2003.
- [14] K. Huang and H. Yan, *On-Line Signature Verification Based on Dynamic segmentation and Global and Local Matching*, Optical Engineering, 34(12), pp. 3480-3487, 1995.
- [15] B. Wirtz. *Stroke-Based Time Warping for Signature Verification*. In the 1st International Conference on Document Analysis and Recognition, pp. 179-182, 1995.
- [16] Q. Wu, S. Lee and I. Jou. *On-line Signature Verification Based on Logarithmic Spectrum*. Pattern Recognition, 31(12), pp.1865-1871,1998.
- [17] H. Feng and C. Wah. *Online Signature Verification Using a New Extreme Points Warping Technique*. Pattern Recognition Letters, 24(16), pp. 2943-2951, 2003.
- [18] J. Lee, H. Yoon, J. Soh, B. Chun and Y. Chung. *Using Geometric Extrema for Segment-to-segment Characteristics Comparison in Online Signature Verification*. Pattern Recognition, 37(1), pp. 93-103, 2004.
- [19] H. Lei and V. Govindaraju. *ER<sup>2</sup>: an Intuitive Similarity Measure for On-line Signatuer Verification*. The 9th International Workshop on Frontiers in Handwriting Recognition, pp. 191-195, 2004