

A Study on the Consistency of Features for On-Line Signature Verification

Hansheng Lei and Venu Govindaraju

CUBS, Center for Unified Biometrics and Sensors
State University of New York at Buffalo, Amherst, NY 14260, USA
{hlei, govind}@cse.buffalo.edu

Abstract. A lot of different features have been proposed for on-line signature verification. By using these features, researchers implicitly believe they have high consistency as well as high discriminatory power. However, very little work has been done to measure the real consistency of these features. In this paper, we propose a model for consistency measure. Experiments were conducted to compare a comprehensive set of features commonly used for on-line signature verification.

1 Introduction and Motivation

Feature extraction and selection is the key for signature verification. A lot of work has been done on this [8, 9]. Different research groups use different features to discriminate genuine and forged signatures. The prerequisite for any feature is high *consistency*. That is, the feature from genuine signatures should be close to each other while the feature from forgeries should be far away. On-line signatures captured by digitizing device usually contains the information of the movement of pen (X -, Y -coordinates), pressure, altitude [8], etc. From the coordinate sequence, speed and acceleration can be derived. Among these information and potential features, which of them are reliable or consistent? Some researchers believe that the dynamic information such as the speed, acceleration or pressure are difficult to forge, thus they are able to distinguish skilled forgeries. However, the prerequisite is that the dynamic information from the authentic person should be consistent, otherwise false rejection will be incurred. Currently, we are lack of a solid support that the dynamic information are reliable or not. By using the extracted features, researchers implicitly believe they have high consistency as well as high discriminatory power. However, due to lack of consistency measure and benchmark databases, no experiment has been done to study the consistency of features proposed by a variety of research groups. Blind feature extraction can be avoided if there is a consistency measure model.

In this paper, we propose a consistency model and compare the consistencies of some commonly used features in on-line signature verification.

2 Consistency Measure

2.1 A Novel Consistency Model

Before we describe the consistency model, we need to know the basic requirements of on-line signature verification system.

Signature verification is a special two-category classification problem: true or false. We say it "special" because it is different from regular two-category problem. Fig. 1 shows the geometric difference. In the regular case, there are two clusters which can be discriminated by the decision boundary. The boundary is usually an *open* hyper-plane, as shown in fig. 1 a). However, in the case of signature verification, there is only one cluster, i.e., the set of genuine signatures, while forged signatures have no clustering characteristic because they have no reason to be close to each other. Therefore, the decision boundary must be a *closed* hyper-plane.

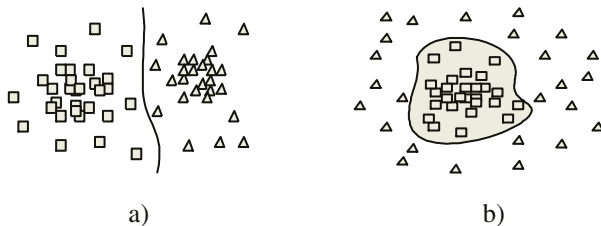


Fig. 1. Two different cases of two-category classification problem. a) Regular two-category problem has two clusterings with open decision boundary. b) Signature verification is a special two-category problem with closed decision boundary around the clustering genuine signatures.

Given few samples of genuine signatures (no more than 6 usually), it is very challenging to determine the decision boundary (threshold) no matter what kinds of features are used. Because of limited training samples, the consistency of feature becomes extremely important. There are many potential features to choose [7, 4, 1] and many new features are being invented. Thus, we are facing an increasing demand for a consistency model.

A simple consistency measure was defined by Lee etc [4] as:

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

where $d_i(a)$ means the consistency of feature i for subject a , $m(a, i)$ is the mean of feature i from genuine signatures, $m(f, i)$ is the mean of feature i from corresponding forgeries and $\sigma^2(a, i)$ or $\sigma^2(f, i)$ means the variation of feature i from genuine or forged signatures.

While this model is appropriate for regular two-category classification, it faces severe problems for signature verification:

- $\sigma^2(f, i)$ does not make sense. As shown in fig. 2 b), signature verification is a special two-category classification problem where the forgeries have no reason to be close to each other.
- The mean $m(a, i)$ can not be calculated directly. For example, if we define the coordinate sequence itself $(X, Y$ or $[X, Y])$ as feature, it is difficult to obtain the mean because the sequences are usually of different lengths. Even sequences can be re-sampled [1] to be of the same length, the mean of signature sequences actually does not mean anything.
- The calculation of distances between features are not limited to Euclidean norm. For example, if we define the sequence $[X, Y]$ as feature, we should use DTW (Dynamic Time Warping [3]) instead of Euclidean distance. Therefore, the calculation of $|m(a, i) - m(f, i)|$ and σ^2 should be generalized to incorporate different distance measures.

For above reasons, we need to tailor this model radically. Therefore, we define the consistency of feature i on subject a as:

$$Cons(a, i) = \frac{|M_{DM_i}(a, a) - M_{DM_i}(a, f)|}{\sqrt{\sigma_{DM_i}^2(a, a) + \sigma_{DM_i}^2(a, f)}}, \tag{2}$$

where the notions are described as follows:

DM_i : distance measure associated with feature i . We acknowledge that different features may have different distance measures, e.g., Euclidean norms, DTW, cross-correlation, etc.

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

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

where $DM_i(c_1, c_2)$ denotes the distance of feature i between object c_1 and c_2 .

a : a set of genuine signatures. f : a set of corresponding forged signatures.

$\sigma_{DM_i}^2(a, a)$: The variation of the feature distances (by distance measure DM_i) within genuine signatures.

$\sigma_{DM_i}^2(a, f)$: The variation of the feature distances (by distance measure DM_i) between genuine and forged signatures.

2.2 Discussion

In the consistency model above, we calculate the mean of feature distances instead of the feature itself because some kinds of features have meaningless "mean" (for example, the coordinate sequence). Also, feature is usually associated with a certain kind of distance measure. Above model takes this into account. Therefore, the model is applicable for all kinds of features proposed for on-line signature verification.

Note that the consistency of feature i as $Cons(a, i)$ is for a particular subject a . The same feature could have different consistency value on different subjects.

Given a data set which consists of a set of subjects, we will calculate the mean and the standard deviation of the consistency value.

3 Commonly Used Features

We subjectively choose over 20 features and compare their consistencies on the SVC signature database[6]. We briefly discuss these features here. Details are referred to corresponding papers.

- *Coordinate sequences.* X , Y , $[X, Y]$ are the most straightforward features. The lengths of these features from different signatures are usually different. Therefore, DTW is used as distance measure [3]. There exist some variants of sequence here. Some researchers propose that the sequence be re-sampled so that they have equal arc-length [1]. We will compare their consistency with/without re-sampling.
- *Speed sequences.* Speed V , speed of X coordinate V_x and speed of Y coordinate V_y can be derived from sequence $[X, Y]$ directly by subtracting neighboring points. From the speed, acceleration V_a can be further derived.
- *Pressure, altitude, azimuth.* Pressure is one of the most common dynamic information of on-line signature. Some devices can capture additional information, such as azimuth (the clockwise rotation of cursor about the z-axis) and altitude(the angle upward toward 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 $s_2(l)$.* The five features were defined in [1]. Center of Mass is actually the smoothed coordinate sequence by Gaussian filter. Torque measures the area swept by the vector of pen position. $s_1(l)$ and $s_2(l)$ measure the curvature ellipse based on moments. The distance measure used here is cross-correlation (Pearson's r) weighted by the consistency of points.
- *Average speed \bar{V} , average positive speed on X-axis \bar{V}_{x+} , average positive speed on Y-axis \bar{V}_{y+} , total signing duration T_s .* Lee etc[4] lists two sets of features. These four features have the highest preference in the first set. The distance measure is Euclidean norm.
- $\cos(\alpha)$, $\sin(\alpha)$, *Curvature β .* α is the angle between the speed vector and the X -axis. The three features are proposed by Jain etc[7]. It also proposes coordinate sequence difference δx and δy . Actually, δx and δy are the same as feature # 5 and # 6 in table 1 respectively.

All above features and corresponding distance measures are summarized in table 1. We have to notice that these features are only a small portions of the proposed features for signature. We choose them because we believe they are among the most promising ones according to our experience.

4 Comparison of Consistency

We used the released SVC database [6] to calculate the consistencies of the features in table 1. SVC has two sets of signatures, namely task 1 and task 2. Each

Table 1. Commonly used features.

#	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: V	Euclidean
17	Average positive V_x : \bar{V}_{x+}	Euclidean
18	Average positive V_y : \bar{V}_{y+}	Euclidean
19	Total signing time: Ts	Euclidean
20	Curvature: β	DTW
21	Angle: $\sin(\alpha)$	DTW
22	Angle: $\cos(\alpha)$	DTW

signature is represented as a sequence of point, 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. There are 40 subjects in each task with 20 genuine signatures and 20 forgeries for each subject.

To compare the consistency of different features fairly, we need to normalize the raw signatures as well the feature distances. We normalized each signature by the same preprocessing: 1) smooth the raw sequence by Gaussian filter; 2) rotate if necessary [2]; 3) normalize the coordinate of each signature Sig_i by:

$$X_i = \frac{X_i - \min(X_i)}{\max(X_i) - \min(X_i)}, Y_i = \frac{Y_i - \min(Y_i)}{\max(Y_i) - \min(Y_i)} \tag{4}$$

The feature distance must be normalized because we have different distance measures here, such as DTW, Euclidean norm and weighted cross-correlation. Normalization of distances is done as follows. Suppose we apply DM_i to class a (genuine signatures) and class f (forgeries). We first calculate all pairwise feature distances within class a by DM_i . We find the maximum feature distance (denoted as D_{max}). Then, we calculate all pairwise feature distances between class a and f . For each $dist$ among these distances, no matter within a or between

a and f , we normalize it by $e^{\frac{-dist}{2 * D_{max}}}$. In this way, all distances are mapped to a value between 0 and 1. The larger the distance, the closer the value to 1.

Same feature may have different consistency on different subjects. Thus, we calculated the mean of consistency of each feature as well as its standard deviation cross subjects.

In addition, to show the relation between feature consistency and its discriminatory power in verification, we calculated the EER (Equal Error Rate) for each feature. We randomly chose 5 genuine signatures from each subject and used the left 35 signatures for verification. Given a testing signature, we calculated the feature distance with all the 5 genuine signatures one by one and returned the maximum normalized distance as similarity output (except the weighted cross-correlation distance, where we output the minimum normalized distance). To determine the EER, we varied the threshold from 0% to 100% and found the point where the FRR (False Rejection Rate) equals the FAR (False Acceptance Rate). By *universal* threshold, we chose the same threshold for all subjects to calculate the total error rate. By *user-dependent* threshold, we chose the optimal threshold for each subject and took the average error rate.

The results are summarized in table 2 in the increasing order of mean consistency value.

From the results as shown in table 2, we have the following observations:

- Although *azimuth* and *altitude* have relatively high mean consistency, they have high standard deviations, which means their discriminatory ability are not stable cross subjects. The corresponding high EER confirms this.
- Some features, like curvature-ellipse $s_1(l)$ and $s_2(l)$, torques $T(l)$, center of mass $\bar{x}(l)$ and $\bar{y}(l)$ are not good enough for skilled forgeries, although they are sophisticated and might be enough for random forgeries. This means, complex features are not necessary better than simpler ones.
- Features like # 17 through # 19 are too simple to carry enough discriminatory information. Thus, they have high EERs. They could be used to prune random forgeries but non-reliable for accepting genuine ones.
- EER has negative relation with the level of consistency, although the EERs do not strictly increase with the consistency decreasing. The variation of consistency also acts to affect the verification performance. This confirms the consistency of feature is directly related to its performance on verification.

Here we have to emphasize that all the EERs in table 2 are high because we used only one feature each time for verification. How to combine these features optimally is still an open problem. We also have to mention that further experiments on real and larger signature databases are necessary to claim the consistency of any given feature.

There is a belief in on-signature verification community that the curve of the signature should be re-sampled with uniform equal arc-length [7, 1, 2]. Is this necessarily true? We conducted experiments to answer this question. We re-sampled all the signature to be of length N and calculated the consistencies of features and corresponding EERs. The results are summarized in table 3 with $N = 200$ (we varied N and found the results had no much difference). Note that

Table 2. Consistencies of features (mean and standard deviation) and EERs (by universal threshold and user-dependent threshold).

Feature	Consistency		EER	
	Mean	Std.	Univ. T.	User. T.
Z_u	1.3789	3.2333	35.06%	26.63%
V_y	1.3255	0.3231	22.06%	11.38%
V_x	1.2644	0.3659	22.65%	16.81%
V	1.2374	0.4143	23.44%	17.63%
T_s	1.2025	0.8128	30.31%	28.18%
A_l	1.1829	4.578	37.63%	29.06%
$\cos(a)$	1.1199	0.2845	26.72%	16.19%
$[X, Y]$	1.1061	0.1795	22.91%	16.83%
$\sin(a)$	1.0997	0.2491	29.56%	20.63%
V_a	1.0966	0.3199	29.29%	22.50%
P	1.0647	0.375	36.86%	25.56%
β	0.985	0.1767	28.90%	20.81%
Y	0.9252	0.1921	25.86%	18.68%
X	0.7784	0.1313	29.59%	25.13%
$\bar{y}(l)$	0.637	0.1397	28.81%	19.19%
$\bar{x}(l)$	0.6023	0.122	29.59%	23.00%
$T(l)$	0.5291	0.1504	33.63%	25.68%
V	0.4814	0.1761	34.50%	31.94%
\bar{V}_{y+}	0.4587	0.1662	36.69%	33.88%
\bar{V}_{x+}	0.3983	0.1717	36.69%	34.00%
$s_2(l)$	0.3158	0.0295	43.86%	42.19%
$s_1(l)$	0.3158	0.0292	43.96%	42.13%

Table 3. Consistencies of features and EER with uniform arc-length re-sampling.

Feature	Consistency		EER	
	Mean	Std.	Univ. T.	User. T.
V_y	0.8841	0.2425	32.36%	20.81%
V_x	0.8714	0.3025	30.44%	21.00%
V	1.1452	0.3066	27.94%	17.63%
$\cos(a)$	1.082	0.2352	29.41%	20.25%
$[X, Y]$	1.0546	0.5839	26.63%	20.94%
$\sin(a)$	0.8093	0.2661	32.56%	23.69%
β	0.8941	0.2360	32.19%	24.75%
Y	0.8683	0.2667	29.94%	21.56%
X	0.7893	0.4407	34.03%	29.63%

some features in table 1 have no sense or difference with the signatures being re-sampled. Compared the consistencies and EERs with/without re-sampling, we can see that re-sampling does not necessarily improve performance. On the contrary, the performance is damaged to some degree.

5 Conclusion Remarks

A novel consistency model tailored for on-line signature verification is proposed in this paper. The consistency of feature is directly related to the feature's performance on signature verification. We conducted experiments to calculate the consistencies of a set of features. The results summarized in table 2 show that some features such as *speed*, *coordinate sequence*, angle α have relatively high consistency, while some others like *azimuth*, *altitude*, curvature-ellipse $s_1(l)$ and $s_1(l)$ are non-reliable. Also, we found that the re-sampling with uniform arc-length does not necessarily increase performance.

References

1. V. Nalwa. *Automatic on-line signature verification*. Proceedings of the IEEE, 85(2): pp. 213-239, Feb. 1997.
2. M. Munich, P. Perona. *Visual identification by signature tracking*. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2003.
3. R. Martens, L. Claesen. *On-line signature verification by dynamic time-warping*. In Proc. 13th Int. Conf. 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 Int. Conf. on Document Analysis and Recognition (ICDAR), 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.