Reliable On-Line Human Signature Verification Systems
Luan L. Lee, Toby Berger, and Erez Avizier

Abstract—On-line dynamic signature verification systems were designed and tested. A data base of more than 10,000 signatures in (x(t), y(t)) form was acquired using a graphics tablet. We extracted a 42-parameter feature set at first, and advanced to a set of 49 normalized features that tolerate inconsistencies in genuine signatures while retaining the power to discriminate against forgeries. We studied algorithms for selecting and perhaps orthogonalizing features in accordance with the availability of training data and the level of system complexity. For decision making we studied several classifiers types. A modified version of our majority classifier yielded 2.5% equal error rate and, more importantly, an asymptotic performance of 7% false acceptance rate at zero false rejection rate, was robust to the speed of genuine signatures, and used only 15 parameter features.

Index Terms—Signature verification, human signature verification, dynamic signature verification, on-line signature verification, point-of-sale, point-of-delivery, forgery.

1 INTRODUCTION

The design and implementation of an on-line dynamic signature verification system involves data acquisition, feature extraction, feature selection, decision making and performance evaluation. Such problems have been discussed both in the by-now classic survey paper of Plamondon and Lorette [1] and, more recently, in the sequel thereto by Leclerc and Plamondon [2].

In this correspondence we describe an approach to designing reliable and effective on-line signature verification which includes: construction of a reliable data base, selection of optimum feature sets with or without forgery data available, finding classifier-independent feature selection procedures, obtaining reliable asymptotic global and individual performance, obtaining a suitable statistical model for signatures, minimizing the effects both of the inconsistency of genuine signatures and of the variety of forgeries, and adapting to practical limitations such as on-line response and limited memory size. We introduce new techniques for on-line human signature verification that are responsive to all these issues and yield performance satisfactory for point-of-sale (POS) applications. Among articles that have appeared subsequent to the submission of this correspondence that overlap somewhat with our results are works by Dimouro et al. [3], Plamondon [4], Nelson et al. [5], Fairhurst and Brittan [6], and Yang et al. [7].

2 DATA ACQUISITION

A total of 3,603 genuine signatures were collected from a population of 105 human subjects which included 22 women and five left-handed writers. Some subjects contributed as few as 13 genuine signatures; one subject wrote his signature more than 1,000 times. About 90% of the genuine signatures were collected under “normal” writing conditions. Because a signature verification system must be robust with respect to variations in writing speed, we also collected a set of 240 “fast” signatures from nine subjects who were asked to write their genuine signatures as fast as possible. The percentage reduction in writing time from normal to fast signing ranged from 10% to 50%. With respect to the size of signatures, the smallest rectangles that fit the largest and smallest signatures, respectively, measured 16 cm by 15 cm and 2 cm by 0.5 cm. Signature writing times ranged from 1 s to 14 s.

The construction of a meaningful forgery data base requires careful planning. We employed three kinds of forgeries, namely: simple [1], statically skilled, and timed forgeries. Although other kinds of forgeries have been suggested in the literature [1], we believe that the three kinds forgeries we used are the sorts most likely to be encountered in POS applications. In simple forgery, we assumed that forgers know how to spell the genuine signatures. A total of 1,148 simple forgeries were collected. A forgery is considered statically skilled when, in addition to possessing all spatial information about the genuine signature, the forger is allowed to practice imitations both on paper and on the tablet. A total of 3,466 statically skilled forgeries were collected. For timed forgery, in addition to the static information about genuine signatures, the forger is provided during practice with information about the average genuine writing time. A total of 1,148 timed forgeries were collected.

Additionally we collected 248 Chinese signatures from 23 subjects, 26 Arabic signatures from two subjects, 13 Tamil signatures from one subject, 13 Korean signatures from one subject, and 13 Hebrew signatures from one subject.

3 FEATURE EXTRACTION

The first feature set consists of 42 personalized parameter features—13 static and 29 dynamic—as described in Appendix A. Some of the 42 features were inspired by the literature of handwriting verification [8], [9], [10], [15], after being appropriately adapted to our signature verification application. Others of the 42-feature set are our own contributions based on experience acquired from the signature collection procedure. The only preprocessing required here is to minimize the effect of the spatial resolution of the graphics tablet by eliminating one of any two consecutive sample points that were separated by only one or two basic spatial resolution units. Spatial resolution of the tablet can seriously affect instantaneous features which involve only a few sample points. We eliminated the problem of signature rotation and shifting by providing a horizontal line on paper taped over the tablet surface as a reference for writing.

Preliminary experiments showed that the 42-feature set is highly sensitive to variations in size and speed of genuine signatures. Therefore, a second set consisting of 49 normalized features was constructed with the objective of rectifying these deficiencies [13]. The feature normalization procedure can be as complex as that proposed in [11], [16] or as simple as linear normalization. The effectiveness of the linear normalization procedure we employed depends on the extent to which the temporal assumption and the spatial assumption below are valid. We did not directly examine the degree of validity of these assumption, but our positive verification results strongly suggest that reality conforms well to these assumptions.

Temporal Assumption: An instance of an event will occur at roughly the same fraction of time of the writing duration of a signature regardless of the overall signing speed. Spatial Assumption: linear scaling of the horizontal and vertical displacements, by possibly different scaling constants, will restore genuine signatures written larger or smaller than usual to the standard shape.

0162-8828/96$05.00 © 1996 IEEE
since the normalization factors—total writing time, total horizontal displacement, and total vertical displacement—were in the original 42-feature set. A detailed description of the normalized feature set appears in Appendix B.

4 Feature Selection

Relatively few attempts have been made to select an optimal subset of features from a larger feature set for automatic human signature verification [8], [10]. Moreover, these approaches either depended on the classifier being used or were not effectively tested by forgery data. Three feature selection algorithms are now described. The first selects features assuming availability of genuine signatures only. Let \( m(a, i) \) and \( \sigma^2(a, i) \) be the sample mean and the sample variance of feature \( i \) computed from the reference data base of subject \( a \). Then the distance measure for feature \( i \) between subject \( a \) and subject \( b \) is defined as

\[
d_i(a, b) = \frac{|m(a, i) - m(b, i)|}{\sqrt{\sigma^2(a, i) + \sigma^2(b, i)}}.
\]

We say that feature \( i \) has a higher order, or degree, of importance for subject \( a \) than does feature \( j \) in population \( P \) if

\[
d_i(a) = \min_{b \neq a} d_i(a, b) > \min_{b \neq a} d_j(a, b) = d_j(a).
\]

In other words, we order features for subject \( a \) in terms of their "maximin" distance from the rest of the entire population \( P \). Selection of the \( k \) best features is equivalent to selection of those \( k \) features (i.e., those \( k \) values of \( i \)) for which the distances \( d_i(a) \) defined by (1) are largest.

When both genuine and forgery data are available, we replace (1) by

\[
d_i(a) = \frac{|m(a, i) - m(j, i)|}{\sqrt{\sigma^2(a, i) + \sigma^2(j, i)}}.
\]

Here \( m(f, i) \) and \( \sigma^2(f, i) \) are, respectively, the sample mean and the sample covariance of feature \( i \) computed from the database of forgeries of subject \( a \)'s signature. When forgery data are available feature \( i \) is considered more important than feature \( j \) if \( d_i(a) > d_j(a) \). An extensive list of the individually ordered 42 features for each of 22 subjects is available in [13]. Features should be selected for incorporation in a classifier not just on the basis of such orderings but also with regard to how they are correlated with one another; see Section 4.2.

Test results confirm that better performance can be achieved by using an optimum individualized subset of features instead of using the whole feature set. However, individually optimized subsets may be precluded by limitation of fast response time and memory size. A feasible suboptimal alternative to the problem of finding an ideal feature set is to identify a so-called common feature set consisting of features that are good for most persons in the population. A common feature set is attractive if the procedure for finding it is simple and the degradation in performance is small enough relative to using individualized feature sets. Our \( m \)-feature common set consists of those \( m \) features with the highest relative frequencies of appearances in the subjects' lists of \( m \) best individualized features. Table 1 shows the selected common set of 10 feature frequencies of appearances and the order of preference among them for the population of 22 subjects.

<table>
<thead>
<tr>
<th>TABLE 1 COMMON 10 FEATURES SETS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>The 42 Feature Set</strong></td>
</tr>
<tr>
<td>Feature #</td>
</tr>
<tr>
<td>Frequency</td>
</tr>
<tr>
<td><strong>The 49 Feature Set</strong></td>
</tr>
<tr>
<td>Feature #</td>
</tr>
</tbody>
</table>

4.1 Classifier Designs

The main task in classifier design is to deduce a distance measure between signatures that reflects strong separation between the class of forgeries and the class of genuine signatures for each subject. However, the joint probability distribution of the features usually is unknown and difficult to estimate. Accordingly, many popular methods which involve probabilistic distance measures cannot be used and all approaches are more or less ad hoc. Majority classifiers, which implement the majority decision rule described below, have the advantage of being simple to implement while providing performance satisfactory for POS applications.

Let \( m_i \) and \( \sigma_i \) represent, respectively, the sample average and sample standard deviation of feature \( i \) in the ensemble of an individual's genuine signatures. Let \( n \) denote the total number of features used in decision process, let \( \alpha \) be a fixed threshold, and \( t_i \) be the value of feature \( i \) for the candidate signature (7) being tested. Define

\[
N_0 = \left| t_i - m_i \right| / \sigma_i \leq \alpha.
\]

The majority decision rule is "T is declared a genuine signature if \( N_0 \geq n/2 \) and a forgery if \( N_0 < n/2.2 \)." Note that the majority rule is highly nonlinear, the decision region in the \( n \)-dimensional normalized feature space for accepting a signature as genuine being unbounded and, roughly speaking, consisting of \( n \) circular infinite cylinders centered in the origin, each with its axis parallel to one of \( n \) feature coordinate axes.

The problem of inconsistency of genuine signatures can be attacked by using normalized features. This reduces the false rejection rate of the majority classifier for fixed \( \alpha \) and a fixed set of \( n \) features but also reduces the forgery rejection rate. Accordingly, tradeoffs must be studied carefully to assess the utility of normalization. Intuition suggests that, if we do not altogether destroy the information that normalization removes from the raw signature data, we may be able to use it to make the majority classifier perform better than when it uses only normalized features. Two such modified decision procedures, designed to work in conjunction with time-normalized features, called "presort majority decision" and "prefinal majority decision," are described below.

The presort majority decision procedure consists of comparing the writing time, \( t_{wt} \), of the signature being tested to the average writing time, \( \bar{t}_{wt} \), computed from the subject's genuine reference set. If \( t_{wt} - \bar{t}_{wt} > \beta \mu_{wt} \), where a nominal value for parameter \( \beta \) is 0.2, do not normalize the dynamic features in the 49-feature set; otherwise, normalize them as usual. Finally, execute the conventional majority decision rule using the resulting 49 features.

The presort majority classifier penalizes signatures possessing highly deviant writing time by removing the normalization factor of signature writing time from the dynamic features. This is justified because considerable consistency in signature writing time was observed among genuine signatures. However, there still is hope for a genuine signature with an anomalous \( t_{wt} \) to be classified correctly if enough features in the 49-feature set behave satisfactorily. Similarly, some forgeries might escape detection even if they fail to pass the predecision procedure.

The prehard majority classifier penalizes signatures with large \( t_{wt} - \bar{t}_{wt} > \beta \mu_{wt} \) more heavily by declaring a signature to be a forgery if \( t_{wt} - \bar{t}_{wt} > \beta \mu_{wt} \).
Karhunen-Loeve (K-L) representation. Toward this end we first
diagonized the sample covariance matrix of the feature set for the
training sample of genuine signatures; usually only five or six of
the resulting uncorrelated linear combination of features were
needed to account for most (80% to 95%) of the overall variance in
individuals' genuine signatures. Next, we did the same for the
training ensembles of "forgeries," finding that we usually needed
about 20 K-L features to account for the bulk of the variance.
Detailed description of extensive experiments on K-L features and
their results can be found in [12].

We have also attacked the feature selection problem via neural
nets and jackknife statistics. A single-neuron classifier yields an
improvement in performance [17]; more general nets naturally
yield further improvement [18]. Also, a bottom-up jackknife fea-
ture selection algorithm can design majority classifiers with small
feature sets and good Type I vs Type II tradeoffs [19].

5 Statistical Model
A good statistical model for signatures would advance signature
technology by quantifying interclass and intraclass variability
among signatures [11]; forgery classes are particularly difficult to
characterize. In attempts to address this problem, this method, and a
recently proposed a statistical model for signature verification by
computer which provides good results for relatively consistent
signatures. We make no attempt to solve the complex and funda-
mentally the problem of developing a statistical model for human
signatures. However, by introducing a statistical model for the pa-
rameter feature sets of genuine signatures, we were able to gener-
ate enough simulated feature vectors to estimate the asymptotic
performance of our signature verification systems and to check the
estimate using data from the subject who signed 1,000 times. We
postulated an additive model, \( X = M + N \), where \( X = (X_1, X_2, \ldots, X_n) \)
is an \( n \)-dimensional random vector representing the signature's
parameter feature set, \( M \) is an \( n \)-dimensional constant vector and
\( N \) is an \( n \)-dimensional Gaussian vector with zero mean and covari-
ance matrix \( C \).

6 Experimental Results and Conclusion
Type I error (false rejection) and Type II error (false acceptance)
are used to evaluate the performance of a signature verification
system [11]. The key performance requirement for POS applications
is that the Type II error must stay small (say, \( \leq 25\% \)) as the deci-
sion threshold is adjusted to drive the Type I error to zero. We
now present experimental results on majority classifiers which
show their suitability for POS applications. Additional results can
be found in [13].

Fig. 1a shows the performance measure of majority classifiers
which employ four different 10-feature sets and the 13 static fea-
tures set, all selected from the 42-feature set. The solid line and
dash-dotted line are for subsets of 10 best individualized features
selected, respectively, with and without reference forgery data
available. The solid line with \( * \) is for the set of 13 static features.
The dashed line is the case of the common 10-feature set selected
from the reference database, i.e., the Basic Data Base of six genu-
ine signatures and eight statically skilled forgeries for each subject.
The performance curve was obtained by evaluating the testing
data base which has five genuine signatures and 22 statically
skilled forgeries for each subject. Finally the dotted line shows
the performance of the same common 10-feature set when not only
the six genuine reference signatures but also the five genuine testing
signatures were included in the procedure to search for a common
feature set, and simultaneously the genuine testing data base also
consists of the same 11 genuine signatures.

We conclude from Fig. 1a: 1) that using a large genuine data
base when selecting common feature results in better classifier
performance, 2) that our common feature sets achieve perform-
ance satisfactory for POS application, 3) that our feature selection
algorithms provide optimum individualized subsets of 10 features
that yield excellent discrimination, and 4) on-line signature verifica-
tion systems that include dynamic features are superior, especi-
ally in POS applications where near-zero Type I error is needed.
The majority classifier with the optimal individualized 34-
feature set provided the overall optimum performance relative to
any size \( N \) of the subset of features selected from our 42 features
when the feature selection algorithm used only genuine signa-
tures. When both genuine and forgery data are available, the per-
formance of subsets of 24 individually selected features is the best
among subsets of features of any dimension selected from our 42
feature set. These results shows that forgery data are desirable but
not essential for feature selection; forgery data avail us of im-
proved performance and simpler verifier structure.

![Fig. 1. Performance comparison of the majority classifiers. (a) some
individual and common majority classifiers; (b) real vs. simulated data;
(c) 42 and 49 feature sets; (d) error rate vs. threshold; (e) preshift ma-
jority classifiers; (f) pekhard majority classifiers.](image-url)
feasible tool not only for computing individual asymptotic performance but also for testing new classifier designs and reducing the burden of collecting extensive samples of raw signatures.

Fig. 1c shows the performance of the majority classifier using subsets of 15 best features selected from the 42 and 49 feature sets, and tested by the fast genuine signatures. The comparison strongly suggests that reliable signature verification systems require normalized features in order to provide satisfactory performance in the presence of significant variations in writing speed. Perhaps most importantly, the performance curves in Fig. 1c support the validity of the assumptions made in Section 3 when we introduced the normalized feature set.

Fig. 1d shows the performance of a majority classifier versus decision threshold for the Fast Data Base and the Global Data Base using the best individualized 15 features selected from the 49-feature set. For POS applications we choose a decision threshold of 2.5; this yielded a Type I error of less than 1% for the normal speed signatures(N) and a Type II error of 20%. Note that the same decision threshold results in 5% Type I error(S) for the fast signatures. It is worth mentioning that in practice we would have a degradation much smaller than 4% because subjects write their signatures fast relatively rarely. Also, we expect that they will be willing to tolerate a somewhat higher rate of rejection of signatures they scrabbled in haste. We conclude from Fig. 1d that using a threshold in the interval from 2.5 to 3.0 for the best 15 features from the set of 49 normalized features results in a majority classifier that yields Type I and Type II errors suitable for POS applications robustly with respect to the speed of genuine signature.

The experimental results reveal that only 0.5% of the genuine signatures had writing durations that deviated from the subject's average writing time by more than 18% (β = 0.18). Under normal conditions few if any people have signature writing times that deviate from their nominal value by more than 20%. Fig. 1e and Fig. 1f show, respectively, the performance of the presoft and prehard majority classifier of Section 4.1. Comparing Fig. 1e and Fig. 1f to Fig 1c shows that notable improvements are achieved using modified majority decision rules. The prehard majority classifier provides the best performance among those studied in this work: 2.5% equal error rate and, more importantly for POS, an asymptotic performance of 7% false acceptance rate at “zero” false rejection rate using only 15 individualized parameter features selected from the 49-feature set. Moreover, this was achieved despite the unrealistically demanding condition that many of the forgers were permitted to produce imitations that matched almost perfectly in signing time in addition to being given the opportunity to practice static forgeries extensively. In particular, the presoft majority classifier detected about 50% of the timed forgeries while successfully classifying 99% of the genuine signatures as shown by the performance curve in Fig. 2.

Our findings provide the basis for an adaptive method and system for real-time verification of dynamic human signature [14] suitable for POS applications in regard to performance, memory and cost; the mean genuine feature values are easily adapted to changes in the signature of an individual with time. Tests on our verification system show that it is suitable for on-line implementation. Using a PC with a 486-processor, the verification algorithms require less than 1.5 s verification time, even without any effort at optimizing the algorithm code. Chip-based special purpose hardware in a commercial realization of the system would run considerably faster.

APPENDIX A – 42 FEATURE SET

Before listing the 42 and 49 feature sets for convenience we introduce the following definition and notations:

- \( x_0 \) = \( x(1\text{st pen down}) \)
- \( y_{nd} \) = \( y(\text{last pen up}) \)
- \( y_{max} \) = maximum \( y \)
- \( d_x \) = \( |y_{max} - y_{min}| \)
- \( y_1 \) = \( y(1\text{st pen down}) \)
- \( x_{max} \) = maximum \( x \)
- \( y_{min} \) = minimum \( y \)
- \( d_y \) = \( |y_{max} - y_{min}| \)
- \( x_{nd} \) = \( x(\text{last pen up}) \)
- \( x_{min} \) = minimum \( x \)
- \( \Delta_x \) = total shift of \( x \) in pen downs
- \( \Delta_y \) = total shift of \( y \) in pen downs

1. Avg. writing speed (\( \bar{v} \))
2. Max. writing speed (\( v_{max} \))
3. Time of max speed (\( t_{v_{max}} \))
4. \( t(1\text{st pen move}) - t(1\text{st pen down}) \)
5. Total signing duration (\( T_f \))
6. Total pen down duration (\( T_d \))
7. Min horiz. writing speed
8. Time of Feature 7
9. Total dots recorded
10. Average dot execution time
11. Number pen ups
12. Time of 2nd pen down
13. Initial direction
14. Dir. from 1st to 2nd pen down
15. Dir. of 1st pen down to 2nd pen up
16. Init. dir. after 2nd pen down
17. Dir. from 1st pen down to last pen up
18. Duration of \( v_i > 0 \)
19. Duration of \( v_i < 0 \)
20. Duration of \( v_i > 0 \)
21. Duration of \( v_i < 0 \)
22. Average positive \( v_i \)
23. Average negative \( v_i \)
24. Average positive \( v_i \)
25. Average negative \( v_i \)
26. Total \( v_i = 0 \) events recorded
27. Total \( v_i = 0 \) events recorded
28. Max \( v_i - \bar{v} \)
29. Max \( v_i - \bar{v} \)
30. Max \( v_i - \min v_i \)
31. Max \( v_i - \min v_i \)
32. Max $v_r - \min v_r$
33. $(x_{max} - T)_{T_o}$
34. $(v_{max})_{T_o}$
35. $(x_{max} - x_{min}) \times (y_{max} - y_{min}) = A_{max}$
36. Signature length $/ A_{max}$
37. $X_i - x_{max}$
38. $x_{max} - x_{max}$
39. $x_{max} - x_{max}$
40. $(x_{max} - x_{min}) / (y_{max} - y_{min})$
41. Standard deviation of $x$
42. Standard deviation of $y$

**APPENDIX B—49 NORMALIZED FEATURE SET**

1. $T_i / T_o$
2. Time of max $v / T_o$
3. Average $v / \text{maximum } v$
4. Duration of $v_o > 0 / T_o$
5. Duration of $v_o < 0 / T_o$
6. Duration of $v_o > 0 / T_o$
7. Duration of $v_o < 0 / T_o$
8. Duration of $v_o > 0$ in pen ups $(T_i - T_o)$
9. Duration of $v_o < 0$ in pen ups $(T_i - T_o)$
10. Duration of $v_o > 0$ in pen ups $(T_i - T_o)$
11. Duration of $v_o < 0$ in pen ups $(T_i - T_o)$
12. Normalized initial direction $(d_i / d_o)$
13. Dir. from 1st to 2nd pen down $(d_i / d_o)$
14. Dir. 1st pen down to 2nd pen up $(d_i / d_o)$
15. Dir. after 2nd pen down $(d_i / d_o)$
16. Dir. before last pen up $(d_i / d_o)$
17. Dir. 1st pen down to last pen up $(d_i / d_o)$
18. Total dots recorded
19. Number of pen ups
20. Time of 2nd pen down $/ T_o$
21. Total dot execution time $/ T_o$
22. Time of max $v_r / T_o$
23. Time of min $v_r / T_o$
24. Time of max $v / T_o$
25. Time of min $v / T_o$
26. Total $v_i = 0$ events recorded
27. Total $v_r = 0$ events recorded
28. Number of quadrant slope changes
29. $F / \text{max } v$
30. $F / \text{max } v$
31. Min. $v / F$
32. Min. $v / F$
33. First time instance of $v = 0$
34. $A_{max} / (\Delta x \Delta y)$
35. Signature length $/ A_{max}$
36. $(x_i - x_{max}) / \Delta x$
37. $(x_i - x_{max}) / \Delta x$
38. $(x_{min} - x_{max}) / \Delta x$
39. $(x_{max} - x_{min}) / \Delta y$
40. $(y_i - y_{max}) / \Delta y$
41. $(y_i - y_{max}) / \Delta y$
42. $(y_{max} - y_{min}) / \Delta y$
43. $(y_{max} - y_{min}) / \Delta y$
44. $(x_{max} - x_{min}) / (y_{max} - y_{min}) \times (x_{max} - x_{max}) / \Delta y$
45. Standard deviation of $x / \Delta x$
46. Standard deviation of $y / \Delta y$

**ACKNOWLEDGMENT**

This work was supported in part by the Brazilian Science Agency (CNPq), the State University of Campinas (UNICAMP), and the National Science Foundation under grant NCR-8903288.

**REFERENCES**


