Machine Learning for Signature Verification

Harish Srinivasan[1], Sargur N. Srihari[1], and Matthew J. Beal

Department of Computer Science and Engineering
University at Buffalo, The State University of New York, Buffalo NY, USA
[1]Center of Excellence for Document Analysis and Recognition (CEDAR), Buffalo NY

Abstract. Signature verification is a common task in forensic document analysis. It is one of determining whether a questioned signature matches known signature samples. From the viewpoint of automating the task it can be viewed as one that involves machine learning from a population of signatures. There are two types of learning to be accomplished. In the first, the training set consists of genuines and forgeries from a general population. In the second there are genuine signatures in a given case. The two learning tasks are called person-independent (or general) learning and person-dependent (or special) learning. General learning is from a population of genuine and forged signatures of several individuals, where the differences between genuines and forgeries across all individuals are learnt. The general learning model allows a questioned signature to be compared to a single genuine signature. In special learning, a person's signature is learnt from multiple samples of only that person's signature– where within-person similarities are learnt. When a sufficient number of samples are available, special learning performs better than general learning (5% higher accuracy). With special learning, verification accuracy increases with the number of samples.

Keywords: machine learning, forensic signature examination, biometrics, signature verification, digital document processing.

1 Introduction

The most common task in the field of forensic document analysis [1, 2, 3, 4, 5] is that of authenticating signatures. The problem most frequently brought to a document examiner is the question relating to the authenticity of a signature: *Does this questioned signature (Q) match the known, true signatures (K) of this subject?* [6] A forensic document examiner– also known as a questioned document (QD) examiner–uses years of training in examining signatures in making a decision in case work.

The training of a document examiner involves years of learning from signatures that are both genuine and forged. In case-work, exemplars are usually only available for genuine signatures of a particular individual, from which the characteristics of the genuine signature are learnt.

Algorithms for visual signature verification are considered in this paper. The performance task of signature verification is one of determining whether a questioned signature is genuine or not.

P. Kalra and S. Peleg (Eds.): ICVGIP 2006, LNCS 4338, pp. 761–775, 2006.

Visual signature verification is naturally formulated as a machine learning task. A program is said to exhibit machine learning capability in performing a task if it is able to learn from exemplars, improve as the number of exemplars increase, etc. [7]. Paralleling the learning tasks of the human questioned document examiner, the machine learning tasks can be stated as *general learning* (which is person-independent) or *special learning* (which is person-dependent) [8].

In the case of general learning the goal is to learn from a large population of genuine and forged signature samples. The focus is on differentiating between genuine-genuine differences and genuine-forgery differences. The learning problem is stated as learning a two-class classification problem where the input consists of the difference between a pair of signatures. The verification task is performed by comparing the questioned signature against each known signature. The general learning problem can be viewed as one where learning takes place with near misses as counter-examples [9].

Special learning focuses on learning from genuine samples of a particular person. The focus is on learning the differences between members of the class of genuines. The verification task is essentially a one-class problem of determining whether the questioned signature belongs to that class or not.

There is scattered literature on automatic methods of signature verification [10, 11, 12, 13, 14]. Automatic methods of writer verification– which is the task of determining whether a sample of handwriting, not necessarily a signature, was written by a given individual– are also relevant [15]. Identification is the task of determining as to who among a given set of individuals might have written the questioned writing. The handwriting verification and identification tasks parallel those of biometric verification and identification for which there is a large literature. The use of a machine learning paradigm for biometrics has been proposed recently [16].

The rest of this paper is organized as follows. Section 2 describes feature extraction in general. Section 3 describes the two methods of learning. Section 4 deals with how the learnt knowledge is used in evaluating a questioned signature (called the performance task). A comparison of the accuracies of the two strategies on a database of genuines and forgeries, along with the particular feature description is described in Section 5. Section 6 is a paper summary.

2 Feature Extraction and Similarity Computation

Signatures are relied upon for identification due to the fact that each person develops unique habits of pen movement which serve to represent his or her signature. Thus at the heart of any automatic signature verification system are two algorithms: one for extracting features and the other for determining the similarities of two signatures based on the features. Features are elements that capture the uniqueness. In the QD literature such elements are termed *discriminating elements* or *elements of comparison.* A given person's samples can have a (possibly variable) number of elements and the combination of elements have greater discriminating power.

A human document examiner uses a chart of elemental characteristics [6]. Such elements are ticks, smoothness of curves, smoothness of pressure changes, placement, expansion and spacing, top of writing, base of writing, angulation/slant, overall pressure, pressure change patterns, gross forms, variations, connective forms and micro-forms. Using the elemental characteristics such as speed, proportion, pressure and design are determined. These in turn allow rhythm and form and their balance are determined.

Automatic signature verification methods described in the literature use an entirely different set of features. Some are based on image texture such as wavelets while others focus on geometry and topology of the signature image. Types of features used for signature verification are wavelet descriptors [17], projection distribution functions [18,14,19], extended shadow code [18] and geometric features [20].

The features are considered representative characteristics of the signature. In order to compare two signatures and to quantify their similarity, a similarity measure or a distance measure is used to compute a score that signifies the strength of match between the features of the two samples. Eventually, irrespective of the method used, one can arrive at a distance space representation of the data that characterizes the strength of match between two signatures. It is usefull to note here that, the learning strategies that ensue are general and are applicable not just to signature veriication but to any bio-metric. As long, as there exist a similarity measure that maps the feature values between a pair of samples, to a score, the below mentioned learning strategies can be used. The particular set of features used for signature verification are described in detail in the experiment and result section 5.

3 Learning Strategies

Person-independent or general learning is a one-step approach that learns from a large population of genuine and forged samples. On the other hand person-dependent(person specific) learning focuses on learning from the genuine samples of a specific individual.

3.1 Person-Independent (General) Learning

The general learning approach uses two sets of signature pairs: genuine-genuine and genuine-forgery. Forgeries in forensic document examination can be either simulated or traced. In this sense task is analogous to learning from near misses in the machine learning literature.

Features are extracted for each pair of signatures and a similarity measure is used to compute the distance between each pair. Let $\boldsymbol{D_S}$ denote the vector of distances between all pairs in set *one*, which represents the distribution of distances when samples truly came from the same person. Similarly let $\boldsymbol{D_D}$ denote the vector of distances between all pairs in set *two*, which represents the distribution of distances when samples truly came from different persons. These distributions can be modeled using known distributions such as Gaussian or gamma. The Gaussian assigns non-zero probabilities to negative values of distance although such values are never encountered. Since this problem in not there with the gamma it is to be preferred. The probability density function of the gamma distributions is as follows: $Gamma(x) = \frac{x^{\alpha-1}\exp^{(-x/\beta)}}{(\Gamma(\alpha))\beta^{\alpha}}$ Here α and β are gamma parameters which can be evaluated from the mean and variance as follows $\alpha = \mu^2/\sigma^2$ and $\beta = \sigma^2/\mu$. 'α' is called the shape parameter and 'β' is the scale parameter. The parameters that need to be learnt for such a model are typically derived from the sufficient statistics of the distribution, and are namely μ (mean) and σ (variance) for a Gaussian, or α (shape) and β (width) for a gamma. These distributions are referred to as genuine-genuine and genuine-impostor distributions in the domain of biometrics.

3.2 Person-Dependent Learning (Person Specific Learning)

In questioned document case work there are typically multiple genuine signatures available. They can be used to learn the variation across them– so as to determine whether the questioned signature is within the range of variation. First, pairs of known samples are compared using a similarity measure to obtain a distribution over distances between features of samples — this represents the distribution of the variation/similarities amongst samples — for the individual. The corresponding classification method involves comparing the questioned sample against all available known samples to obtain another distribution in distance space. The Kolmogorov-Smirnov test, KL-divergence and other information-theoretic methods can be used to obtain a probability of similarity of the two distributions, which is the probability of the questioned sample belonging to the ensemble of knowns. These methods are discussed below.

Within-person distribution. If a given person has N samples, $\binom{N}{2}$ defined as $\frac{N!}{N!(N-r)!}$ pairs of samples can be compared as shown in Figure 1. In each comparison, the distance between the features is computed. This calculation maps feature space to distance space. The result of all $\binom{N}{2}$ comparisons is a $\{\binom{N}{2}\times 1\}$ distance vector. This vector is the distribution in distance space for a given person. For example, in the signature verification problem this vector is the distribution in distance space for the ensemble of genuine known signatures for that writer. A key advantage of mapping from feature space to distance space is that the number of data points in the distribution is $\binom{N}{2}$ as compared to N for a distribution in feature space alone. Also the calculation of the distance between every pair of samples gives a measure of the variation in samples for that writer. In essence the distribution in distance space for a given known person captures the similarities and variation amongst the samples for that person. Let N be the total number of samples and $N_{WD} = \binom{N}{2}$ be the total number of comparisons that can be made which also equals the length of the within-person distribution vector. The within-person distribution can be written as

$$\boldsymbol{D_W} = (d_1, d_2, \ldots, d_{N_{WD}})^\top \tag{1}$$

where $\top$ denotes the transpose operation and d_j is the distance between the pair of samples taken at the j^{th} comparison, $j \in \{1, \ldots, N_{WD}\}$.

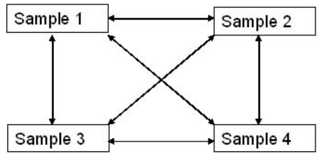


Fig. 1. Person-dependent (special) learning involves comparing all possible genuine-genuine pairs, as shown for four genuine samples, to form the vector $\boldsymbol{D_W}$, which in this example is of length $N_{WD} = \binom{4}{2} = 6$ comparisons

4 Performance Task

The performance task of signature verification is to answer the question whether or not a questioned signature belongs to the genuine signature set. The person-independent method uses knowledge from a general population to determine whether two samples, one a questioned and the other a genuine, belong to the same person. This task is called 1 : 1 verification. Person-dependent classification tasks involves matching one questioned sample against multiple known samples from the person. Details of the two performance algorithms are given below.

4.1 Person-Independent Classification

The process of 1 : 1 verification(one input sample compared with one known sample) starts with feature extraction and then computing the distance d between the features using a similarity measure. From the learning described in Section 3.1, the likelihood ratio defined as $\frac{P(D_S|d)}{P(D_D|d)}$ can be calculated, where $P(D_S|d)$ is the probability density function value under the D_S distribution at the distance d and $P(D_D|d)$ is the probability density function value under the D_D distribution at the distance d. If the likelihood ratio is greater than 1, then the classification answer is that the two samples do belong the same person and if the ratio is less than 1, they belong to different persons. If there are a total of N known samples from a person, then for one questioned sample N, 1 : 1 verifications can be performed and the likelihood ratios multiplied. In these circumstances it is convenient to deal with log likelihood-ratios rather than with just likelihood ratios. The log-likelihood-ratio (LLR) is given by $\log P(D_S|d) - \log P(D_D|d)$. The decision of same-person is favored if $\log P(D_S|d) - \log P(D_D|d) > 0$, and the decision of different-person chosen if $\log P(D_S|d) - \log P(D_D|d) < 0$. When N of these 1 : 1 verifications are performed these LLR's are summed and then the decision is taken.

4.2 Person-Dependent Classification

When multiple genuines are available then the within-person distribution is obtained in accordance with equation 1. A questioned can be compared against the ensemble of knowns for verification. The classification process consists of two steps.

(i) obtaining questioned *vs* known distribution; and

(ii) comparison of two distributions: questioned *vs* known distribution and within-person distribution.

Questioned vs *Known Distribution.* In Section 3.2 and with equation 1 the within-person distribution is obtained by comparing every possible pair of samples from within the given persons samples. Analogous to this, the questioned sample can be compared with every one of the N knowns in a similar way to obtain the questioned vs known distribution. The questioned vs known distribution is given by

$$D_{QK} = (d_1, d_2, \ldots, d_N)^\top \ , \tag{2}$$

where d_j is the distance between the questioned sample and the j^{th} known sample, $j \in \{1, \ldots, N\}$.

Comparing Distributions. Once the two distributions are obtained, namely the within-person distribution, denoted D_w (Section 3.2, equation 1), and the Questioned Vs Known distribution, D_{QK} (Section 4.2, equation 2), the task now is to compare the two distributions to obtain a probability of similarity. The intuition is that if the questioned sample did indeed belong to the ensemble of the knowns, then the two distributions must be the same (to within some sampling noise). There are various ways of comparing two distributions and these are described in the following sections.

Kolmogorov-Smirnov Test. The Kolmogorov-Smirnov (KS) test can be applied to obtain a probability of similarity between two distributions. The KS test is applicable to unbinned distributions that are functions of a single independent variable, that is, to data sets where each data point can be associated with a single number [21]. The test first obtains the cumulative distribution function of each of the two distributions to be compared, and then computes the statistic, D, which is a particularly simple measure: it is defined as the maximum value of the absolute difference between the two cumulative distribution functions. Therefore, if comparing two different cumulative distribution functions $S_{N1}(x)$ and $S_{N2}(x)$, the KS statistic D is given by $D = \max_{-\infty<x<\infty} |S_{N1}(x) - S_{N2}(x)|$. The statistic D is then mapped to a probability of similarity, P, according to equation 3

$$P_{KS} = Q_{KS}\left(\sqrt{N_e} + 0.12 + (0.11/\sqrt{N_e})D\right) \ , \tag{3}$$

where the $Q_{KS}(\cdot)$ function is given by (see [21] for details):

$$Q_{KS}(\lambda) = 2\sum_{j=1}^{\infty}(-1)^{j-1}e^{-2j^2\lambda^2} \ , \text{such that} : Q_{KS}(0) = 1 \ , Q_{KS}(\infty) = 0 \ , \tag{4}$$

and N_e is the effective number of data points, $N_e = N_1 N_2 (N_1 + N_2)^{-1}$, where N_1 is the number of data points in the first distribution and N_2 the number in the second. The following sections discuss other methods of comparing two distributions.

Kullback-Leibler Divergence and other methods. The Kullback-Leibler (KL) divergence is a measure that can be used to compare two binned distributions. The KL divergence measure between two distributions is measured in bits or nats. An information theoretic interpretation is that it represents the average number of bits that are wasted by encoding events from a distribution P with a code which is optimal for a distribution Q (i.e. using codewords of length $-\log q_i$ instead of $-\log p_i$). Jensen's inequality can be used to show that $D_{KL} = KL(P\|Q) \geq 0$ for all probability distributions P and Q, and $D_{KL} = KL(P\|Q) = 0$ iff $P = Q$. Strictly speaking, the KL measure is a *divergence* between distributions and not a distance, since it is neither symmetric nor satisfies the triangle equality). The KL divergence so obtained can be converted to represent a probability by $\exp(-\zeta D_{KL})$ (for the sake of simplicity we set $\zeta = 1$ in this article). If the divergence D_{KL} is 0, then the probability is 1 signifying that the two distributions are the same. In order to use this method and other methods discussed in the following sections it is first necessary to convert the two unbinned distributions to binned distributions with a probability associated with each bin. The KL divergence between two distributions is given in equation 5 below, where B is the total number of bins, P_b and Q_b are the probabilities of the b^{th} bin of two distributions respectively. P_{KL} denotes the probability that the two distributions are the same. Other related measures between distributions P and Q that we will examine are given in equations 6, 7 and 8

$$\text{Kullback-Leibler:} \quad D_{KL} = \sum_{b=1}^{B} P_b \ \log(\frac{P_b}{Q_b}) \quad P_{KL} = e^{-\zeta D_{KL}} \tag{5}$$

$$\text{Reverse KL:} D_{RKL} = KL(Q\|P) = \sum_{b=1}^{B} Q_b \ \log(\frac{Q_b}{P_b}) P_{RKL} = e^{-\zeta D_{RKL}} \tag{6}$$

$$\text{Symmetric KL:} \quad D_{HKL} = \frac{1}{2} KL\,(P\|Q) + \frac{1}{2} KL\,(Q\|P) = \frac{D_{KL} + D_{RKL}}{2} \qquad P_{HKL} = e^{-\zeta D_{HKL}} \tag{7}$$

$$\text{Jensen-Shannon KL:} \qquad D_{JS} = \frac{1}{2} KL \left(P \middle\| \frac{P+Q}{2} \right) + \frac{1}{2} KL \left(Q \middle\| \frac{P+Q}{2} \right) \qquad P_{JS} = e^{-\zeta D_{JS}} \tag{8}$$

Combined KL and KS measure. A combination of the Kolmogorov-Smirnov and Kullback-Leibler measure, denoted KLKS, has been found to outperform the individual measures as will be analyzed in the performance evaluation section following this. The method to combine is very simple and is obtained by averaging the probabilities defined in equations 3 and 5.

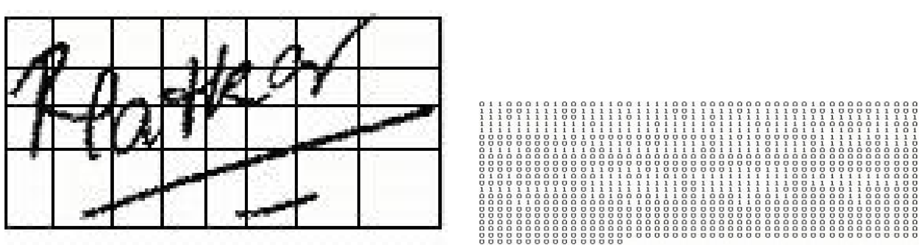
$$P_{KLKS} = \frac{P_{KL} + P_{KS}}{2} \tag{9}$$

5 Performance Evaluation

The particular set of features used for signature verification are mentioned below.

5.1 Multiresolution Features

A quasi-multiresolution approach for features are the Gradient, Structural and Concavity, or GSC, features [22, 23]. Gradient features measure the local scale characteristics (obtained from the two-dimensional gradient of the image), structural features measure the intermediate scale ones (representing strokes), and concavity can measure the characteristics over the scale of whole image (representing concavities and topology). Following this philosophy, three types of feature maps are drawn and the corresponding local histograms of each cell is quantized into binary features. Fig. 2(a) shows an example of a signature, which has a 4x8 grid imposed on it for extracting GSC features; rows and columns of the grid are drawn based on the black pixel distributions along the horizontal and vertical directions. A large number of binary features have been extracted from these, as shown in Fig. 2(b), which are *global word shape features* [24]; there are 1024 bits which are obtained by concatenating 384 gradient bits, 384 structural bits and 256 concavity bits.



(a) Variable size grid (b) 1024-bit binary feature vector

Fig. 2. Signature feature computation using a grid: (a) variable size 4x8 grid, and (b) binary feature vector representing gradient, structural and concavity features

A similarity or distance measure is used to compute a score that signifies the strength of match between two signatures. The similarity measure converts the pairwise data from feature space to *distance* space.

Several similarity measures can be used with binary vectors, including the well-known Hamming distance. Much experimentation with binary-valued GSC features, has led to the *correlation* measure of distance as yielding the best asccuracy in matching handwriting shapes [25]. It is defined as follows. Let S_{ij} $(i, j \in \{0, 1\})$ be the number of occurrences of matches with i in the first vector and j in the second vector at the corresponding positions, the dissimilarity D between the two feature vectors X and Y is given by the formula:

$$D(X,Y) = \frac{1}{2} - \frac{S_{11}S_{00} - S_{10}S_{01}}{2\sqrt{(S_{10}+S_{11})(S_{01}+S_{00})(S_{11}+S_{01})(S_{00}+S_{10})}}$$

It can be observed that the range of $D(X,Y)$ has been normalized to $[0,1]$. That is, when $X = Y$, $D(X,Y) = 0$, and when they are completely different, $D(X,Y) = 1$.

A refined method to compute the features and obtain the distance values is discussed in [26].

5.2 Experiments

A database of off-line signatures was prepared as a test-bed [13]. Each of 55 individuals contributed 24 signatures thereby creating 1320 genuine signatures. Some were asked to forge three other writers' signatures, eight times per subject, thus creating 1320 forgeries. One example of each of 55 genuines are shown in Figure 3. Ten examples of genuines of one subject (subject no. 21) and ten forgeries of that subject are shown in Figure 4. Each signature was scanned at 300 dpi gray-scale and binarized using a gray-scale histogram. Salt pepper noise removal and slant normalization were two steps involved in image preprocessing. The database had 24 genuines and 24 forgeries available for each writer as in Figure 4. For each test case a writer was chosen and N genuine samples of that writer's signature were used for learning. The remaining $24-N$ genuine samples were used for testing. Also 24 forged signatures of this writer were used for testing. Figure (Fig. 5) shows the image of a questioned signature is matched against multiple images of known signatures in figure.

Fig. 3. Genuine signature samples

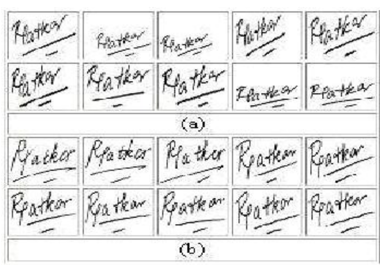


Fig. 4. Samples for one writer: (a) genuines and (b) forgeries

Two different error types can be defined for any biometric person identification problem. False reject rate (Type 1) is the fraction of samples classified as not belonging to the person when truly there were from that person. False acceptance rate (Type 2) is the fraction of samples classified as belonging to the person when truly the samples were not from that person. In the domain of signatures, Type 1 is the fraction of samples classified as forgeries when truly they were genuine and Type 2 the fraction of samples classified as genuine when truly they were forgeries.

5.3 Person-Independent(General) Method

The classification decision boundary discussed in Section 4.1 is given by the sign of the log likelihood-ratio, LLR, $\log P(D_S|d) - \log P(D_D|d)$. A modified decision boundary can be constructed using a threshold α, such that $\log P(D_S|d) - \log P(D_D|d) > \alpha$. When α is varied, we can plot ROC curves as shown in Figure 6. The different subplots in the figure correspond to the ROC curves as the number of known samples is increased from 5 to 20. For each plot, the total error rate defined as (False acceptance+False reject)/2 is minimum at a particular value of α. This is the best setting of α for the specified number of known samples, denoted the *operating point*, and is indicated with an asterix '*'. When 20 samples are used for learning the error rate is approximately 79%. Figure 7 shows the distribution of LLRs when the questioned samples were genuine and when they were forgeries. A larger region of overlap indicates a higher error rate.

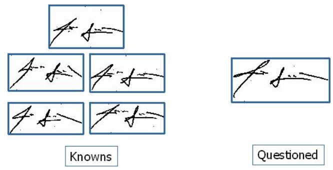


Fig. 5. Signature verification with multiple knowns

5.4 Person-Dependent Method

The person-dependent classification discussed in Section 4.1 mentioned six different statistics for comparing the two distributions to obtain a probability of

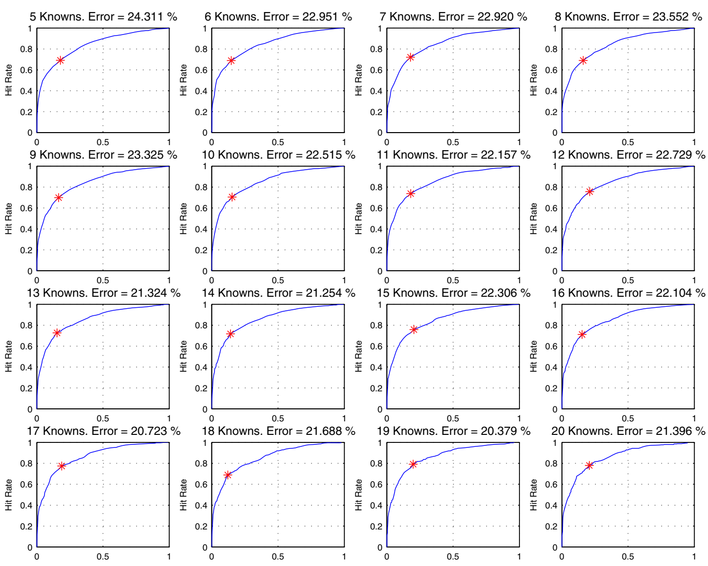


Fig. 6. ROC curves parameterized by α is varied. Each subplot is titled with the number of knowns used for training and the optimum error rate that is possible. The asterix '*' denotes the optimal operating point α for that model.

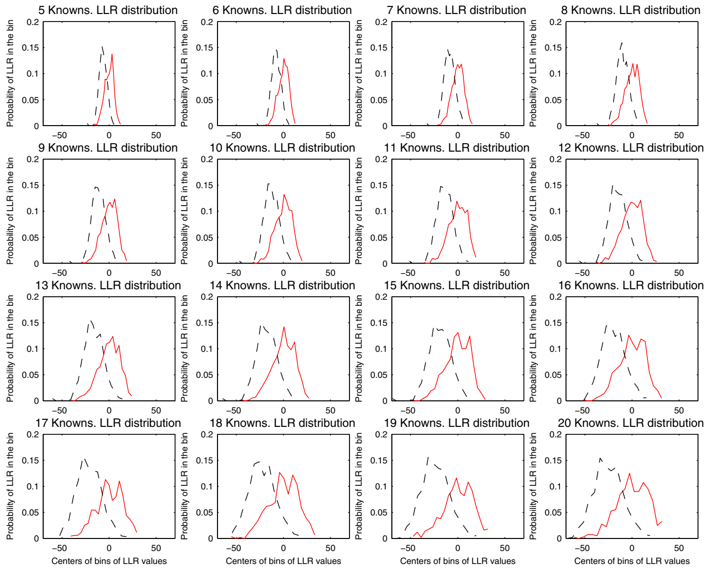
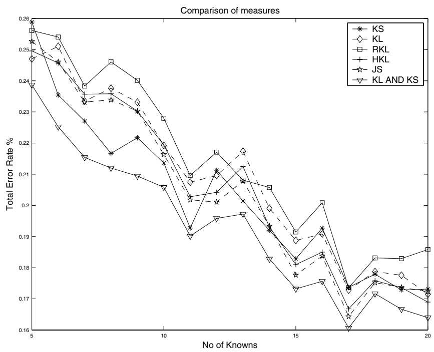


Fig. 7. LLR's obtained from each test case plotted as histograms. The probability (y-axis) that the LLR falls into a range of LLR values (x-axis) is shown for the results of truly genuine (solid) and forgery cases (dotted). Each subplot corresponds to training on a different number of knowns.

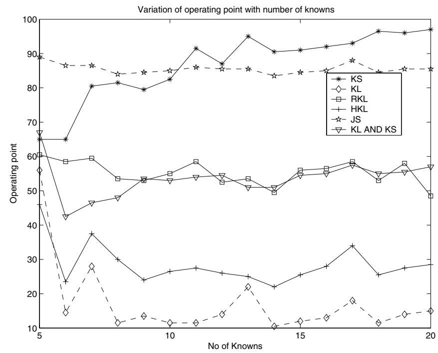
Table 1. Error rates for signature verification. Measures are Kolmogorov-Smirnov (**KS**), Kullback-Leibler (**KL**), reverse KL (**RKL**), symmetrized KL (**HKL**), Jensen-Shannon (**JS**), and combined KL and KS (**KL and KS**). These are graphed in Figure 8(a).

No. of Knowns	KS	KL	RKL	HKL	JS	KL and KS
5	25.88	24.70	25.61	24.96	25.26	23.87
6	23.54	25.10	25.40	24.57	24.60	22.52
7	22.71	23.35	23.83	23.57	23.31	21.54
8	21.67	23.76	24.60	23.58	23.39	21.20
9	22.17	23.31	24.01	23.03	23.03	20.94
10	21.36	21.93	22.79	21.94	21.63	20.58
11	19.27	20.74	20.96	20.28	20.18	19.02
12	21.13	20.96	21.71	20.42	20.10	19.58
13	20.14	21.73	20.81	21.25	20.78	19.72
14	19.06	20.03	20.84	19.46	19.33	18.41
15	18.28	18.88	19.15	18.10	17.76	17.32
16	19.27	19.08	20.08	18.50	18.38	17.56
17	17.37	17.28	17.36	16.68	16.43	16.07
18	17.79	17.88	18.31	17.58	17.52	17.17
19	17.39	18.09	18.42	17.75	17.37	16.97
20	17.31	17.15	18.58	16.90	17.23	16.40

match between the questioned sample and the ensemble of knowns. In order to measure error rates for this classificaton technique, once again a decision needs to be made based on the probability of whether or not the questioned sample belongs to the ensemble of knowns. If the probability of match $> \alpha$, then the decision is in favour of the questioned signature to be genuine, and if the probability



(a) Comparison of measures. The total error rate is plotted for the different measures as the number of knowns used for training is increased. The combined KL and KS measure (lowest trace) outperforms other measures (see text).



(b) Variation of operating point of the different measures, as a function of the number of knowns used for training. Operating point correponds to the best setting of the decision boundary (probability) for classifying samples as genuine or forgery.

Fig. 8.

of match $< \alpha$, the decision is in favor of a forgery (this α should not be confused with that used in the person-independent method). By varying the parameter α, once again ROC curves (False Accept vs. False Reject) can be plotted for each of the six measures. The best setting of α is termed as the *operating point.* This setting of α corresponds to the least total error rate possible. Note that the ROC curves are plotted for the test data set and the operating point determined on them. These test data set can be considered as a validation set that helps to determine the operating point. In the curve, the operating point is the point closest to the origin. Table 1 shows the least total error rate possible when different number of known samples were used for training for each of the 6 different measures. Figure 8(a) shows the same table as a graph comparing the different measures and it can be seen that the combined KL and KS measure performs the best. The reason for this can be intutively explained by the fact that KS statistic has low false accept rates whereas the KL statistic has low false reject rates. The combination of these two in the KL and KS measure works the best.

Figure 8(b) shows how the operating point (best setting of α) varies with the number of known samples used. It can be seen that in order to obtain the least total error rate, the value of α changes with the number of knowns for certain measures. The value of α explains a lot about what each statistic learns from the known samples. For example, the high value of α for the KS statistic when large numbers of known samples were used explains that the KS statistic focuses on learning the variation amongst the known samples. Presence of large known samples accounts for greater variation amongst them. Hence if KS focuses on

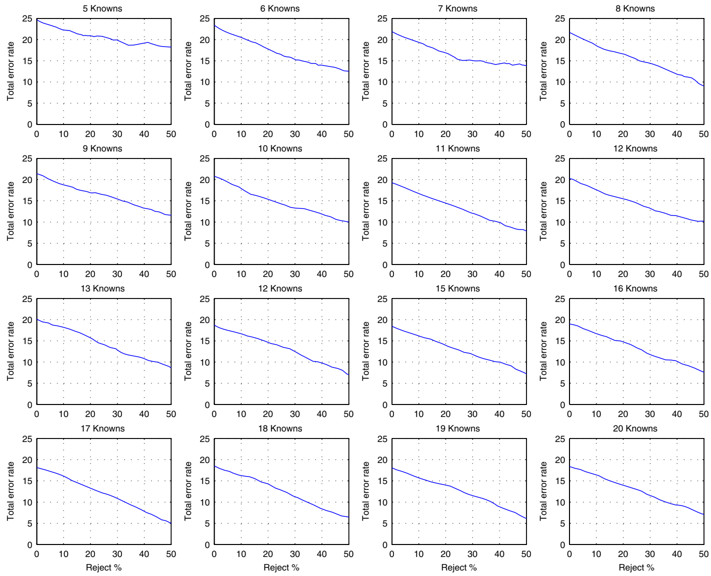


Fig. 9. Error rates as the percentage of allowed rejected (no decision) cases increases. The rejection rate is indirectly controlled by varying the β which assigns the probability region $50 - \beta$ and $50 + \beta$ where no decisions are made and considered as rejects. The different subplots show the plots for different number of knowns used for learning. We have plotted only the trend for the combined KL and KS measure.

learning the variation, then almost every questioned sample ends up receiving a high probability of match as the majority of questioned samples (genuines and forgeries) invariably fall within the variation. Thus by setting a high value of α the decision that a sample is truly genuine is made only if probability is really high. In simple terms this means that when more samples are used for training the KS statistic will declare a sample as genuine only if the probability of match is really high. In contrast to this measure the KL measure captures the similarities amongst the known samples a lot. This is evident by the low value of α for large number of knowns. Presence of large number of samples accounts for observing more similarities. The KL measure focuses on learning the similarities amongst the samples and it returns a high probability of match very rarely and only when every similarity that is learnt amongst the known samples is present in the questioned sample. Hence the majority of questioned sample receive a low probability of match by the KL measure. To counter this a low value of α ensures that the KL measure will declare a sample as forgery only if the probability of match is really low. Similar comments can be made about other measures and it is important to note that those measures for which the operating point does not vary with the number of knowns and those which are around the 50% mark can be a useful property. This basically shows that irrespective of the number of knowns used for training, one can make a decision using the same operating point, and also if the operating point is around the 50% mark there is an equal range of probabilties across which the two different decisions fall. And it is also intuitive that the combined KL KS measure has this fine property. It can be seen that the operating point for the combined KL and KS measure is closest to the 50% mark amongst other measure and is also independent of the number of known samples to some extent. Proceeding with the conclusion that the combined KL KS measure has a few desired properties and also outperforms other measures in terms total error rate, we can now consider allowing rejections to reduce the error rates even further. Consider probabilities between $.5 - \beta$ and $.5 + \beta$ for some $\beta > 0$ as the region for reject probabilities. No decision is made if $.5 - \beta < Probability < .5 + \beta$. This can significantly reduce the total error rate. Figure 9 shows the total error rate as it the rejection percentage is changed by changing the value of β. This analysis enables the operator to select a value of β that will induce a certain rejection rate and in turn result in a certain desired error rate. For example, in order to obtain a error rate of 10% with 20 knowns in this data set one should set β to .15 and that accounts for 35% reject rate. Similarly for an error rate of 5% for 20 knowns, β needs be set to .30 which accounts for 62% reject rate.

6 Summary and Discussion

Automatic signature verification is a task where machine learning can be used as a natural part of the process. Two different machine learning approaches, one involving genuines and forgeries in a general set and another involving only genuines for a particular case were described. The first approach is analogous to using counter-examples with near misses in the learning process. Both approaches involve using a similarity measure to compute a distance between features of two signatures. Special learning outperforms general learning particularly as the number of genuines increases. General learning is useful when the number of genuines is very small (less than four). A refined method of extracting features for signatures was also discussed which can further increase verification accuracy. Future work should consider combining the two types of learning to improve performance.

Acknowledgments. This work was supported in part by the National Institute of Justice grant 2004-IJ-CX-K030

References

1. Osborn, A.: Questioned Documents. Nelson Hall Pub (1929)
2. Robertson, E.W.: Fundamentals of Document Examination. Nelson-Hall (1991)
3. Bradford, R.R., Bradford, R.: Introduction to Handwriting Examination and Identification. Nelson-Hall (1992)
4. Hilton, O.: Scientific Examination of Questioned Documents. CRC Press (1993)
5. Huber, R., Headrick, A.: Handwriting Identification: Facts and Fundamentals. CRC Press (1999)
6. Slyter, S.A.: Forensic Signature Examination. Charles C. Thomas Pub (1995)
7. Mitchell, T.M.: Machine Learning. McGraw-Hill (1997)

8. Srihari, S.N., Xu, A., Kalera, M.K.: Learning strategies and classification methods for off-line signature verification, Proceedings of the Seventh International Workshop on Frontiers in Handwriting Recognition(IWHR), IEEE Computer Society Press (2004) 161–166
9. Winston, P.: Learning structural descriptions from examples. In Winston, P., ed.: The Psychology of Computer Vision. McGraw-Hill (1975) 157–210
10. Leclerc, F., Plamondon, R.: Automatic signature verification: the state of the art, 1989-1993. International Journal of Pattern Recognition and Artificial Intelligence **8** (1994) 643–660
11. Guo, J.K., Doermann, D., Rosenfield, A.: Local correspondences for detecting random forgeries, Proceedings of the International Conference on Document Analysis and Recognition (1997) 319–323
12. Plamondon, R., Srihari, S.N.: On-line and off-line handwriting recognition: A comprehensive survey. IEEE Transactions on Pattern Analysis and Machine Intelligence **22** (2000) 63–84
13. Kalera, M.K., Zhang, B., Srihari, S.N.: Off-line signature verification and identification using distance statistics. International Journal of Pattern Recognition and Artificial Intelligence **18** (2004) 1339–1360
14. Fang, B., Leung, C.H., Tang, Y.Y., Tse, K.W., Kwok, P.C.K., Wong, Y.K.: Off-line signature verification by the tracking of feature and stroke positions. Pattern Recognition **36** (2003) 91–101
15. Srihari, S.N., Cha, S., Arora, H., Lee, S.: Individuality of handwriting. Journal of Forensic Sciences (2002) 856–872
16. Srinivasan, H., Beal, M., Srihari, S.N.: Machine learning approaches for person verification and identification. Volume 5778., Proceedings of SPIE: Sensors, and Command, Control, Communications, and Intelligence Technologies for Homeland Security (2005) 574–586
17. Deng, P.S., Liao, H.Y., Ho, C., Tyan, H.R.: Wavelet-base off-line handwritten signature verification. Computer Vision Image Understanding **76** (1999) 173–190
18. Sabourin, R.: Off-line signature verification: Recent advances and perspectives. BSDIA (1997) 84–98
19. Coetzer, J., B.M.Herbst, du Preez, J.: Off-line signature verification using the discrete radon transform and a hidden markov model. Journal on Applied Signal Processing **4** (2004) 559–571
20. Ferrer, M.A., Alonso, J.B., Travieso, C.M.: Off-line geometric parameters for automatic signature verification using fixed-point arithmetic. IEEE Transactions on Pattern Analysis and Machine Intelligence **27** (2005) 993–997
21. Press, W.H., Flannery, B.P., Teukolsky, S.A., Vetterling, W.T.: Numerical Recipes in C: The Art of Scientific Computing. Cambridge University Press (1992)
22. Srikantan, G., Lam, S., Srihari, S.: Gradient based contour encoding for character recognition. Pattern Recognition **7** (1996) 1147–1160
23. Zhang, B., Srihari, S.N.: Analysis of handwriting individuality using handwritten words, Proceedings of the Seventh International Conference on Document Analysis and Recognition, IEEE Computer Society Press (2003) 1142–1146
24. Zhang, B., Srihari, S.N., Huang, C.: Word image retrieval using binary features. In Smith, E.H.B., Hu, J., Allan, J., eds.: SPIE. Volume 5296. (2004) 45–53
25. Zhang, B., Srihari, S.: Properties of binary vector dissimilarity measures. Cary, North Carolina (September, 2003)
26. Chen, S., Srihari, S.N.: A new off-line signature verification method based on graph matching, Proc. International Conference on Pattern Recognition (ICPR 2006) (August 2006)