Off-Line Signature Verification: A Comparison between Human and Machine Performance

J. Coetzer

Department of Mathematical Sciences, University of Stellenbosch, South Africa
jcoetzer@sun.ac.za

B.M. Herbst

Department of Mathematical Sciences, University of Stellenbosch, South Africa
herbst@dip.sun.ac.za

J.A. du Preez

Department of Electrical and Electronic Engineering, University of Stellenbosch, South Africa
dupreez@dsp.sun.ac.za

Abstract

When a large number of documents, e.g. bank cheques, have to be authenticated in a limited time, the manual verification of, say the authors' signatures, is often unrealistic. This led to the development of a wide range of automatic off-line signature verification systems. However, the value of such a system is rarely demonstrated by conducting a subjective test. We recently developed a novel off-line signature verification system [2, 3] that uses features that are based on the calculation of the Radon transform (RT) of a signature image. Each writer's signature is subsequently represented by a hidden Markov model (HMM). This paper is an extension of [3] and illustrates the value of our system by showing that it outperforms a typical human being. We conduct an experiment on a data set that contains 765 test signatures (432 genuine signatures and 333 skilled forgeries) from 51 writers.

Keywords: off-line signature verification, subjective test

1. Introduction

Handwritten signatures are socially and legally readily accepted as a convenient means of document authentication. Although handwritten signatures are not the most reliable means of personal identification, signature verification systems are inexpensive and non-intrusive.

Signature verification systems are categorised into on-line and off-line systems. In the *on-line* case a special pen is used on an electronic surface. A signature is captured dynamically and then stored as a function of time. The stored data is referred to as a *dynamic* signature and also contains information on pen velocity and acceleration. In *off-line* systems, a signature is digitised using a flat-bed scanner and then stored as an image. These images are called *static* signatures. On-line systems are of interest for "point-of-sale" and security applications, while off-line systems are of interest in scenarios where only hard copies of signatures are available, e.g. where a large number of documents need to be authenticated. Since off-line systems do not exploit any dynamic information, they are much less reliable than on-line systems.

One of the main incentives for developing an off-line system is the potential financial benefits that the automatic clearing of cheques will have for the banking industry. Banks still process millions of cheques daily. Typically, only those cheques of which the amount exceeds a certain threshold are verified manually by an operator. This is a cumbersome process that has to be completed within a limited time. It is therefore imperative that the performance of an automated system is comparable to that of a human being. At the same time, the processing requirements must be reasonable, so as to make the adoption of such a system economically viable.

We recently developed an HMM-based off-line signature verification system. In our paper [3] we explain that the value of this system is twofold:

(1) Although our system do not outperform all existing systems, these systems utilise a technique or features that are different from those used by our system. It is therefore very likely that a combination of any of these systems and ours, will result in a superior merged system. This makes their approaches complementary to ours. In our paper we included a brief survey of recent systems. Other surveys can be found in [8, 7, 9, 11]. An article by Guo, Doermann and Rosenfeld [6] also includes an extensive overview of previous work.

(2) Our system is computationally efficient. This makes the adoption of our system economically viable.

In this paper we show that, in addition to the above-mentioned attributes, our system also outperforms a typical human being. We therefore demonstrate that, as a substitute for human verification, our system is a viable option.

In order to make this paper self-contained, we give a concise description of our HMM-based off-line signature verification system in Section 2. In this paper we conduct an experiment on a data set that contains 765 test signatures (432 genuine signatures and 333 skilled forgeries) from 51 writers. A brief description of this data set is given in Section 3. The experimental setup and results are discussed in Sections 4 and 5, respectively.

2. System overview

In this section we give a brief overview of our off-line signature verification system. The feature extraction method is based on the calculation of the RT of a signature image, while each writer's signature is modelled by a ring-structured HMM. For a more detailed description, see [3].

2.1. Feature extraction

Our system first calculates the RT of a signature image. This implies that projections (shadows) of the signature are obtained at T different angles. These angles are equally distributed between 0^{o} and 360^{o}. The RT can be efficiently calculated with the algorithm described in [1]. In order to guarantee translational and scale invariance, the RT is subjected to further image processing (see [3]). Each processed projection represents a feature vector (observation) of dimension d. When all the angles are considered, an observation sequence of length T is therefore obtained. Since the RT is calculated at angles that range between 0^{o} and 360^{o}, each observation sequence is periodic.[1]

2.2. Signature modelling

For each writer, our system uses the appropriate training sequences to train a ring-structured HMM with N states and ℓ allotted forward links (see Fig. 1). Comprehensive tutorials on HMMs can be found in a paper by Rabiner [10] and the book by Deller [4].

2.2.1. Notation

We use the following notation for a sequence of T continuous observations,

$$\mathbf{X}_1^T = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T\}, \tag{1}$$

where $\mathbf{x}_i,\ i = 1, 2, \ldots T$ denotes the ith feature vector in the sequence.

[1] Although the last $T/2$ observations are reflected copies of the first $T/2$ observations, *all* the observations are used. As a result, each observation sequence is periodic. When a ring-structured HMM is therefore used, rotational invariance is assured.

The following notation is used for a continuous, first order HMM, λ:

(1) We denote the N individual states as

$$\mathbf{S} = \{s_1, s_2, \ldots, s_N\}, \tag{2}$$

and the state at time t as q_t.

(2) The initial state distribution is denoted by $\boldsymbol{\pi} = \{\pi_i\}$, where

$$\pi_i = P(q_1 = s_i),\ i = 1, \ldots N. \tag{3}$$

(3) The state transition probability distribution is denoted by $\mathbf{A} = \{a_{i,j}\}$, where

$$a_{i,j} = P(q_{t+1} = s_j | q_t = s_i),\ i, j = 1, \ldots, N. \tag{4}$$

(4) The probability density function (PDF), which quantifies the similarity between a feature vector $\mathbf{x}$ and the state s_j, is denoted by

$$f(\mathbf{x}|s_j, \lambda),\ j = 1, \ldots, N. \tag{5}$$

(5) The similarity between an observation sequence $\mathbf{X}$ and a model λ is denoted by

$$f(\mathbf{X}|\lambda). \tag{6}$$

2.2.2. HMM topology

Since the HMM is constructed in such a way that it is equally likely to enter the model at any state, the ring topology of our HMM and the periodic nature of each observation sequence guarantee that the signatures are not sensitive to rotational variations.

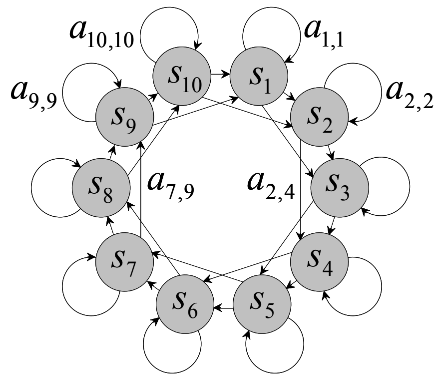

Figure 1. An example of an HMM with a ring topology. This model has ten states ($N = 10$) with two allotted forward links ($\ell = 2$).

2.2.3. Training

Each state is represented by a PDF of which only the mean vector is estimated. Each HMM is initialised and then trained, using the Viterbi re-estimation algorithm. The dissimilarity between an observation sequence $\mathbf{X}$ and an HMM λ is calculated as follows,

$$D(\mathbf{X}, \lambda) = -\ln(f(\mathbf{X}|\lambda)), \quad (7)$$

where $f(\mathbf{X}|\lambda)$ quantifies the match between $\mathbf{X}$ and λ.

When the training set for writer w is denoted by

$$\{\mathbf{X}_1^{(w)}, \mathbf{X}_2^{(w)}, \ldots, \mathbf{X}_{N_w}^{(w)}\}, \quad (8)$$

where N_w is the number of samples in the training set, the dissimilarity between every training sample and the HMM is used to determine the following statistic for the writer's signature,

$$\mu_w = \frac{1}{N_w} \sum_{i=1}^{N_w} D(\mathbf{X}_i^{(w)}, \lambda_w). \quad (9)$$

2.3. Verification

When a claim is made that a test sequence $\mathbf{X}_{\text{Test}}^{(w)}$ belongs to writer w, the sequence is matched with the appropriate HMM, λ_w, through Viterbi alignment. This match is quantified by $f(\mathbf{X}_{\text{Test}}^{(w)}|\lambda_w)$. The dissimilarity between the test pattern and the HMM is then calculated as follows,

$$D(\mathbf{X}_{\text{Test}}^{(w)}, \lambda_w) = -\ln(f(\mathbf{X}_{\text{Test}}^{(w)}|\lambda_w)). \quad (10)$$

Since each HMM state is only modelled by a mean vector, the dissimilarity value in Eq. 10 is based on an *Euclidean* distance measure. Therefore, in order to use a *global* threshold for all the writers in our data set, every dissimilarity value in Eq. 10 is normalised, using μ_w in Eq. 9,

$$D_{\text{Eucl}}(\mathbf{X}_{\text{Test}}^{(w)}, \lambda_w) = \frac{D(\mathbf{X}_{\text{Test}}^{(w)}, \lambda_w) - \mu_w}{\mu_w}. \quad (11)$$

We then use a *sliding* threshold τ, where $\tau \in [0, \infty)$. When $D_{\text{Eucl}}(\mathbf{X}_{\text{Test}}^{(w)}, \lambda_w) < \tau$, that is when

$$D(\mathbf{X}_{\text{Test}}^{(w)}, \lambda_w) < \mu_w(1 + \tau), \quad (12)$$

the claim is accepted, otherwise the claim is rejected.

3. Dolfing's data set

We conduct an experiment (Section 4) on signatures that are randomly selected from a data set that was originally captured on-line for Hans Dolfing's Ph.D. thesis [5]. Dolfing's data set contains 4800 signatures from fifty-one writers. Each of these signatures contains static and dynamic information captured at 160 samples per second. Each of these sample points contains information on pen-tip position, pen pressure, and pen tilt.

Static signature images are constructed from this data using only the pen-tip position, that is the x and y coordinates, for those sample points for which the pen pressure is non-zero. These signature images are therefore "ideal" in the sense that they contain virtually no background noise. This acquisition method also ensures a uniform stroke-width (five pixels) within each signature and throughout the data set. A more detailed discussion of this signature acquisition method can be found in [3].

Dolfing's data set contains different types of forgeries. This paper focuses on skilled forgeries. A *skilled forgery* is produced when the forger has unrestricted access to one or more samples of the writer's actual signature. Skilled forgeries can be subdivided into *amateur* and *professional* forgeries. A *professional forgery* is produced by an individual who has professional expertise in handwriting analysis. They are able to circumvent obvious problems and exploit their knowledge to produce high quality, spacial forgeries (see Fig. 2 (b)).

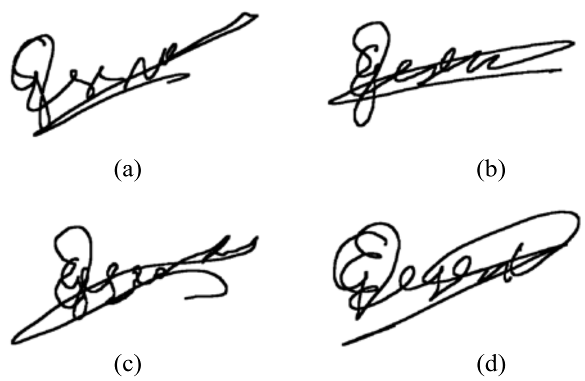

Figure 2. An example of a genuine signature and a professional forgery are shown in (a) and (b), respectively. Examples of amateur forgeries are shown in (c) and (d).

In the subsequent experiment we only consider amateur forgeries (see Fig. 2 (c) and (d)). We did not consider professional forgeries, since these forgeries are only available for a few writers in Dolfing's data set. For each writer in Dolfing's data set, there are 15 training signatures, 15 genuine test signatures and 60 amateur forgeries available, with the exception of two writers, for whom only 30 amateur forgeries are available.

4. Experimental setup

For each of the 51 writers in Dolfing's data set we construct a test set that consists of *only* 15 signatures. For each of the 51 writers, all the available training signatures (15 per writer) are used. Each test set contains a randomly selected number (any number between 0 and 15) of amateur forgeries. The remaining test signatures are randomly selected from the 15 genuine test signatures for the writer in question. It is therefore possible that a specific test set contains *only* genuine test signatures or *only* amateur forgeries. A verifier (human or machine) is therefore presented with a total of $15 \times 51 = 765$ test signatures. The *total* number of genuine test signatures and forgeries turns out to be 432 and 333 respectively.

4.1. Human verification

For our human verification experiment (subjective test) twenty-two individuals are each presented with the signatures from all 51 writers in the data set. These verifiers include faculty members, graduate students and departmental secretaries. We present each individual with a training set (15 signatures) and a corresponding test set (15 signatures) for all 51 writers. The training set and the corresponding test set for a specific writer are presented on two separate sheets of paper. An individual typically compares the test signatures, as a unit, with the corresponding training set and then decide which of the test signatures to reject. Each individual verifier was instructed not to ponder over a decision, so as to simulate what a bank clerk is likely to do. As a result, most individuals took approximately 45 minutes to 1 hour to verify all the signatures, that is approximately 3.5 to 4.7 seconds per signature.

4.2. Machine verification

The same training and test signatures, that are considered for human verification, are also considered for machine verification. For this experiment we use feature vectors of dimension $d = 512$ and observation sequences of length $N = 256$. Our HMMs have $N = 64$ states with one allotted forward link ($\ell = 1$).

5. Results

5.1. Error rates

When our HMM-based system is implemented on the randomly selected test signatures, described in Section 4, the receiver operating characteristic (ROC) graph (false acceptance rate (FAR) against false rejection rate (FRR)) in Fig. 3 is obtained.[2] Threshold values between $\tau = -0.1$ and $\tau = 1$ were used. The verification results (error rates) for the twenty-two human verifiers are indicated by circles.

From Fig. 3 it is clear that only four human verifiers performed better than our system, of which only one performed significantly better. These results are indicated by circles below the ROC graph. It is important to note that circles (manual verification results) further away from the diagonal (that is where the FAR is equal to the FRR) are less significant than circles closer to the diagonal. After all, a human verifier can ensure that he/she performs as well as our system by simply accepting or rejecting all the signatures. Also note that a human verifier has the advantage of being able to view *all* the test signatures for a specific writer *at once*, while our system only considers one test signature at a time.

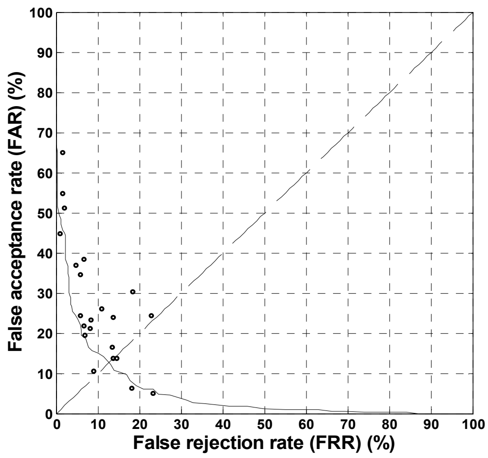

Figure 3. ROC graph when our HMM-based system is implemented on a randomly selected subset of Dolfing's data set. The results for the human verification experiment (when the same set of signatures are considered) are indicated by circles. Since four of these circles are below the ROC graph, it is clear that only four out of the twenty-two human verifiers outperformed our HMM-based system.

5.2. Significance test

The human verification experiment consists of 22 identical independent trials, where a trial represents a specific human verifier's attempt to "outperform" our HMM-based system. Each trial has two possible outcomes, "success" or "failure". The experiment therefore follows a binomial distribution.

The error rates for a human verifier are represented by a circle in Fig. 3. A trial is deemed successful when the circle is *below* the ROC graph obtained for our HMM-based system. The number of successes for the above experiment is therefore 4. Let us now assume that the probability of success is 0.5 and that this probability remains constant from one trail to another.

Let the random variable X count the number of successes in all the trials. Given a 0.5 probability of success, the probability that a human verifier will outperform our HMM-based system 4 times or less, in 22 trials, is therefore given by

$$P(X \in [0, 4]) = 0.5^{22} \sum_{k=0}^{4} \binom{22}{k} = 0.002172 \approx 0.22\%. \tag{13}$$

The mean and variance of the binomial distribution $X \sim \text{Bin}(22, 0.5)$ is given by $\mu_X = 11$ and $\sigma_X^2 = 5.5$, respectively. Since a count of 4 is almost 3 standard deviations away from the mean, the proba-

[2] Our HMM-based system achieves an equal error rate of approximately 12%.

bility of observing this count is extremely small.

It is therefore safe to conclude that the probability of success is significantly less than 0.5. This implies that our HMM-based system performs better than a typical human being.

6. Discussion and future work

In this paper we demonstrated that our HMM-based system outperforms most human verifiers. Therefore, as a substitute for human verification, our HMM-based system is a viable option.

It is clear from Fig. 3 that one human verifier, with a FRR of 9.0% and a FAR of 10.5%, performed significantly better than our HMM-based system. The equal error rate (EER) for our HMM-based system is 12.6%. In order to ascertain whether this result is significant or coincidental, it is necessary to repeat our human verification experiment several times, using the same 22 human verifiers, but using a different set of randomly selected test signatures each time. Due to time constraints, this was not carried out.

One should also be able to better simulate a real-life scenario by using bank officials, instead of academics, for our human verification experiment.

Acknowledgement

The authors wish to thank Hans Dolfing for making his signature database available to them.

References

[1] R.N. Bracewell, *Two-Dimensional Imaging.* Englewood Cliffs, NJ: Prentice Hall, pp. 505-537 (1995).

[2] J. Coetzer, *Off-line signature verification.* Ph.D. Thesis, University of Stellenbosch (2005).

[3] J. Coetzer, B.M. Herbst, and J.A. du Preez, "Offline Signature Verification Using the Discrete Radon Transform and a Hidden Markov Model", *Eurasip Journal on Applied Signal Processing - Special Issue on Biometric Signal Processing,* H. Bourland, I. Pitas, K.K. Lam, and Y. Wang, editors, vol. 2004, no. 4, pp. 559-571 (2004).

[4] J.R. Deller, J.G. Proakis, and J.H. Hansen, *Discrete-Time Processing of Speech Signals.* IEEE (1999).

[5] J.G.A. Dolfing, *Handwriting Recognition and Verification. A Hidden Markov Approach.* Ph.D. Thesis, Eindhoven University of Technology (1998).

[6] J.K. Guo, D. Doermann, and A. Rosenfeld, "Forgery Detection by Local Correspondence", *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 15, no. 4, pp. 579-641 (2001).

[7] F. Leclerc and R. Plamondon, "Automatic Signature Verification: The State of the Art, 1989-1993", *International Journal on Pattern Recognition and Artificial Intelligence: Special Issue on Signature Verification*, vol. 8, no. 3, pp. 643-660 (1994).

[8] R. Plamondon and S.N. Shihari, "On-Line and Off-Line Handwriting Recognition: A Comprehensive Survey", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 1 (2000).

[9] R. Plamondon and G. Lorette, "Automatic Signature Verification and Writer Identification - The Sate of the Art", *Pattern Recognition*, vol. 22, no. 2, pp. 107-131 (1989).

[10] L.R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition", *Proceedings of the IEEE*, vol. 77, pp. 257-286 (1989).

[11] R. Sabourin, R. Plamondon, and G. Lorette, "Off-Line Identification with Handwritten Signature Images: Survey and Perspectives", *Structured Document Image Analysis*, H. Baird, H. Bunke, and K. Yamamoto, editors, pp. 219-234, Berlin, Heidelberg, New York, Tokyo: Springer-Verlag (1992).