

# Evaluating the Biometric Sample Quality of Handwritten Signatures

Sascha Müller<sup>1</sup> and Olaf Henniger<sup>2</sup>

<sup>1</sup> Technische Universität Darmstadt, Darmstadt, Germany  
mueller@sec.informatik.tu-darmstadt.de

<sup>2</sup> Fraunhofer Institute for Secure Information Technology, Darmstadt, Germany  
henniger@sit.fraunhofer.de

**Abstract.** This paper addresses the problem of evaluating the quality of handwritten signatures used for biometric authentication. It is shown that some signature samples yield significantly worse performance than other samples from the same person. Thus, the importance of good reference samples is emphasized. We also give some examples of features that are related to the signature stability and show that these have no influence on the actual utility of the sample in a comparison environment.

## 1 Introduction

Not all biometric samples are equally well suited for the automated recognition of the persons from whom they are acquired. For example, one fingerprint sample may contain a very low number of utilisable minutiae, making biometric recognition difficult, while another one may contain more distinctive features and be better suited. In the case of handwritten signatures – which are widely accepted for authenticating purposes – this is specifically an issue since the stability of a signature varies greatly between individuals.

The question raises how useful an acquired biometric sample will be in a comparison environment. This can be expressed by a quality score that is assigned to that sample. The quality score of a genuine biometric sample is a quantitative expression of the predicted utility of the sample for telling genuine and forged samples apart. It can be used, for instance, for deciding whether a repetition of the data acquisition is necessary or for weighting results in multi-biometric systems.

Fields that can hold biometric sample quality scores have been introduced into the headers of several biometric data structures [1,2,3]. If a biometric sample quality score is reported, valid values are integers between 1 and 100. Values in the range 1–25 are supposed to indicate unacceptable quality, in the range 26–50 marginal quality, in the range 51–75 adequate quality, and in the range 76–100 excellent quality.

Quality assessment algorithms have been developed mainly for image samples such as fingerprint samples [4], but there are also proprietary algorithms in use for assessing the complexity of handwritten signature samples during the enrollment process. How to objectively assess biometric sample quality, such that

quality assessment algorithms of different suppliers would result in consistent sample quality scores is the topic of ongoing research activities. Results may enter the standardization process [5].

There are two approaches to evaluating the quality of a biometric sample [6]:

**A-posteriori approach.** Assessing the quality of a genuine biometric sample a posteriori means evaluating the utility of the sample by comparing it with the other samples from a database.

**A-priori approach.** Assessing biometric sample quality a priori means predicting the utility of the sample without comparing it with other samples from a database.

In order that the a-priori quality score of a biometric sample can be used as a predictor of the utility of that sample, the a-priori quality score must be assigned in such a way that it correlates with the a-posteriori quality score. This paper investigates the problem of quality evaluation in the case of handwritten signatures captured using a graphic tablet.

The quality of a biometric sample depends on the comparison algorithm used. So, a decision had to be taken on the comparison algorithm. In order to obtain results that are representative for practical scenarios of handwritten signature verification, we decided to apply the widely used comparison algorithm Dynamic Time Warping (DTW), which can be considered among the best approaches for online signature verification [7,8]. The algorithm determines the “distance” between two time series as a measure for their dissimilarity. As a common optimization a Sakoe/Chiba band [9] has been implemented, the width of which is set automatically to 10% of the length of the longer of the two time series that are to be compared.

A-posteriori biometric sample quality also depends on the database used for comparison. A publicly available subset of the database [10] has been used that consists of signature samples of 100 persons. For each person, there are 25 genuine samples and 25 skilled forgeries. Using skilled forgeries is important to estimate the resistance of the samples against forgery. In addition to the X and Y coordinates, each sample point record includes a numeric value representing the associated pen pressure.

The remainder of the paper is organized as follows: Section 2 deals with the a-posteriori assessment of the quality of handwritten signatures. Section 3 considers the a-priori quality assessment. Section 4 summarizes the results and gives an outlook.

## 2 A-Posteriori Quality Assessment

### 2.1 Obtaining Quality Scores From Comparison Results

Using the genuine sample to be assessed as biometric reference, its distance to the other genuine signatures of the same person and to the forgeries can be determined. Each of the 250 genuine signature samples has been compared both to the other 24 genuine samples of the same person and to the 25 forgery attempts. All distance values resulting from these comparisons have been recorded.

Two measures have been chosen to express the quality of a signature sample:

1. The sample equal error rate (sEER) that is achieved when comparing a sample with the corresponding genuine and forged samples. The lower it is, the better suited is the sample for keeping genuine and forged samples apart.
2. The mean value  $\mu$  of the distances of the sample to the other genuine samples. This measure depends only on the signature's stability. The lower it is, the more similar is the sample to the other genuine samples.

The mean value is strongly correlated to the standard deviation  $\sigma$  of the distances of the sample to the other genuine samples (correlation coefficient 0.8), which can therefore be used for signature stability considerations, too.

## 2.2 Results

The overall EER of the comparison algorithm over all reference samples is 4.82%, but it could be improved, as will be shown: Fig. 1 visualizes the sEER's of the signatures of the persons 60 to 79. Note, for instance, the outlier that occurs in the second sample of person 71. Using this sample as reference template causes an sEER that is significantly higher than the average sEER of the other genuine samples of the same person (3.57%). Visual inspection of the sample in question does not show any obvious aberration, and the mean value  $\mu$  of the distances of this sample to the other genuine samples is only marginally worse than that of the other samples (cf. Fig. 2). Outliers like this occur very often, so care has to be taken when choosing a reference sample during enrollment.

When all samples with an sEER of 20% or higher are prohibited as biometric reference – less than 4.5% of all authentic signatures fall into this category – then the overall EER lessens to 3.54%. Taking out the worst tenth of all signatures further reduces the overall EER to 2.71%. However, such strict requirements would make it impossible for some persons to use the biometric system at all because all of their samples would be rejected. A strong impact of outliers on the overall performance of biometric systems has been observed before for other systems [4].

A similar phenomenon, though less extreme, can be seen in the mean values of the distances of a sample to the other genuine samples (Fig. 2). Obviously, large distance values have a conspicuous impact on the sEER. For instance, the mean values of distances of genuine signatures of person 78 range from 41 to 114.<sup>1</sup> Accordingly, the sEER ranges from acceptable values to unusable ones of more than 80% (as seen in Fig. 1). This is plausible because the stability influences the sEER.<sup>2</sup> These results clearly show the importance of using a good (i.e., similar to other genuine signatures) reference signature. If the reference signature is one of the outliers, the recognition performance is bad even for the best possible threshold.

<sup>1</sup> These values are distances returned by the DTW algorithm, and thus depend on specific implementation details. In the implementation used for our research, the distances range from 0 for identical samples to about 250 for incomparable ones.

<sup>2</sup> In our tests, a correlation coefficient of 0.55 has been observed between the mean  $\mu$  of the distances of genuine signatures and the sEER.

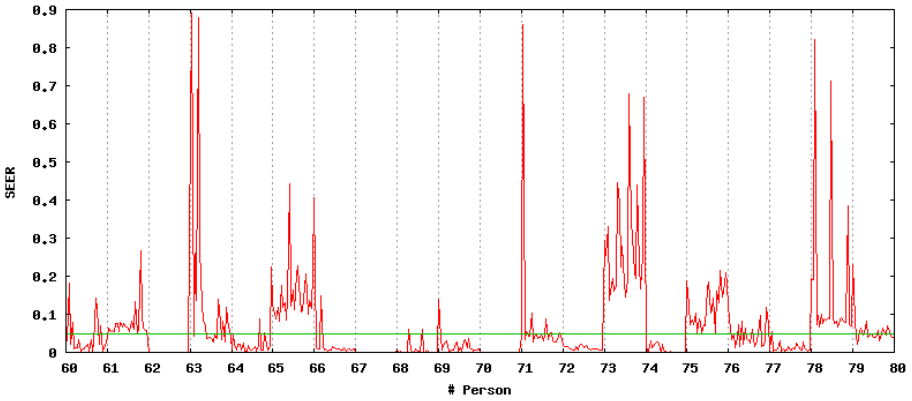


Fig. 1. sEER’s of the genuine signatures of the persons 60–79

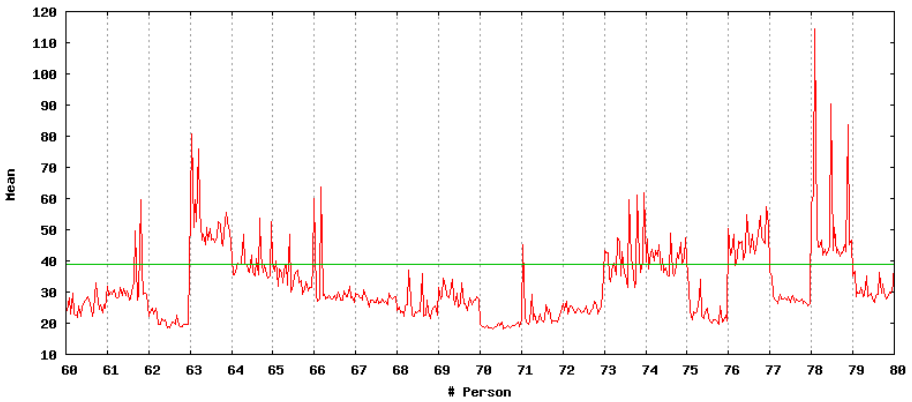


Fig. 2. Mean values  $\mu$  of the distances of the genuine signatures of the persons 60–79

### 3 A-Priori Quality Assessment

#### 3.1 Examined Features

In order to find correlations between global signature features and the a posteriori signature quality, the Pearson product-moment correlation coefficient was used to compare the distributions of the recorded quality measures (cf. Section 2) with certain features of the associated signature samples. Since only linear relationships can be detected this way, all features were also plotted and analyzed by visual inspection.

In the past a lot of global signature features have been considered for the analysis of signatures. [11], for example, uses a feature vector of 100 global features for this purpose. We selected a set of 12 global features that we considered obvious candidates for quality considerations and easy to extract from a given signature sample. All of the following global features were examined:

1. Length in the writing plane
2. Ratio of length to area
3. Average writing speed
4. Maximum writing speed
5. Average pen pressure in pen down strokes
6. Average acceleration
7. Maximum acceleration
8. Average absolute value of curvature
9. Number of pen down strokes
10. Number of extrema of x channel
11. Number of extrema of y channel
12. Ratio of width to height

### 3.2 Results

For lack of space, only a few representative results are given in detail, as well as a summary of all results. In most cases the correlation coefficients were very low, especially when looking at the sEER. Still, visual inspection allowed some interesting discoveries. For example, there is no linear correlation between the sEER and the maximum acceleration (coefficient  $-0.029$ ), but it can be seen in the plot (not given here) that all signatures that lead to very bad sEER (40% or higher) have very low maximum acceleration.

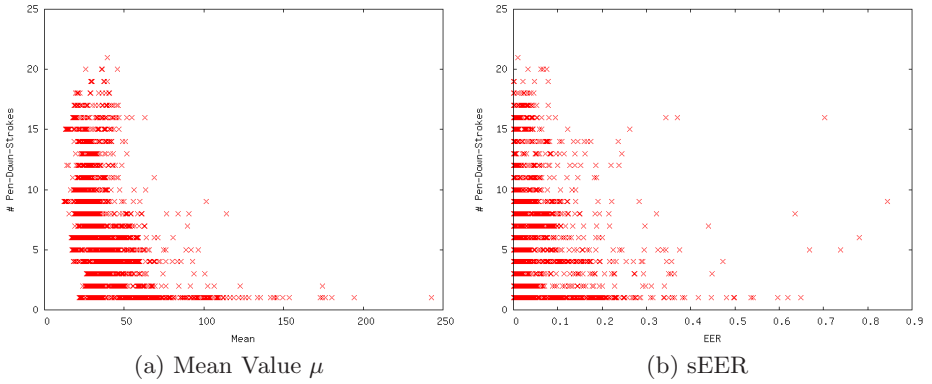
Some of the following examples have stronger correlations:

**Number of pen-down strokes.** With a correlation coefficient of  $-0.463$ , a slight inverse linear correlation can be assumed between the number of pen-down strokes of a signature sample and the mean value of the distances of the signature to the other genuine samples of the same person. This tells us that signatures with more pen-down strokes tend to be more stable than signatures consisting of only a few strokes. This can also be verified by examining Fig. 3a.

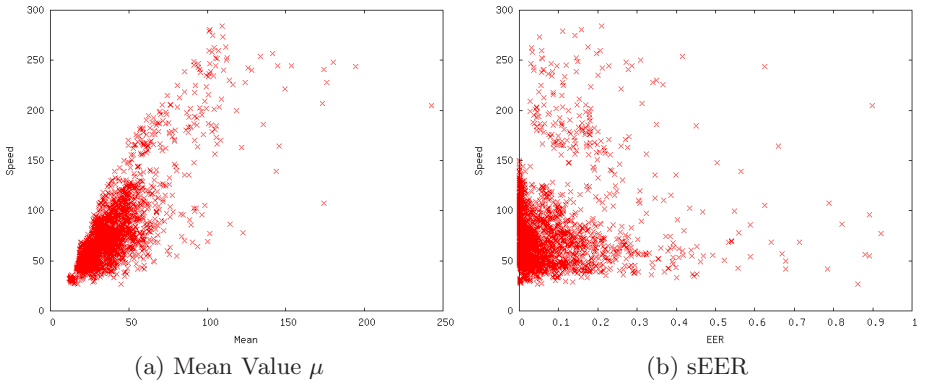
The connection between the number of pen-down strokes and the equal error rate (Fig. 3b) is a lot weaker, the correlation coefficient being only  $-0.223$ . This means, for the success of a forgery attempt it does not matter much of how many pen-down strokes the original signature consists.

**Average writing speed.** The correlation coefficient of the average writing speed and the mean value  $\mu$  of distances of genuine signatures is  $0.797$  – the highest value obtained in all tests. Surprisingly, the correlation with the sEER amounts to only  $0.130$ , meaning that there is no linear relationship. Obviously, slower signatures are more stable, but easier to forge. Fig. 4 shows the plots.

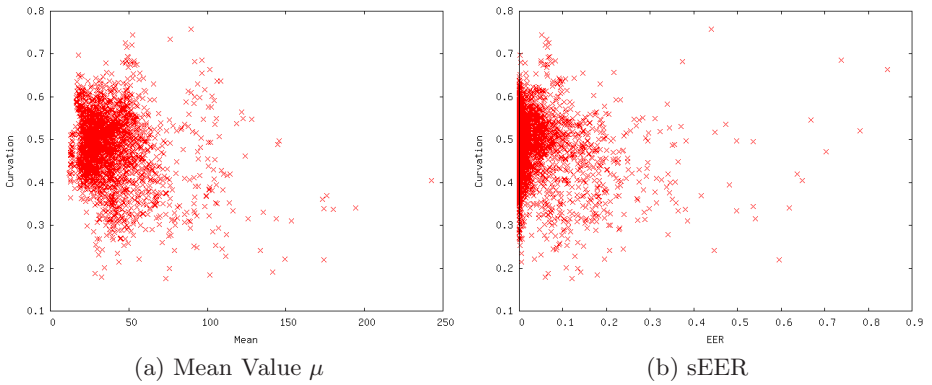
**Average absolute value of curvature.** The idea here is to measure the “shakiness” of the handwriting. The curvature is the difference of the gradients of adjacent sample points. Since gradient values can be very big, the tilt angle was used instead, which is  $\arctan(\frac{dy}{dx})$ . If  $dx = 0$ , some small value  $\epsilon$  was used instead. This is reasonable because  $\arctan(\frac{1}{\epsilon}) \approx \frac{1}{2}\pi = \arctan(\infty)$ . As the plot (Fig. 5) shows, this feature is neither correlated with sEER nor with  $\mu$ . Also, the correlation coefficients are very small ( $-0.223$ , resp.  $-0.144$ ).



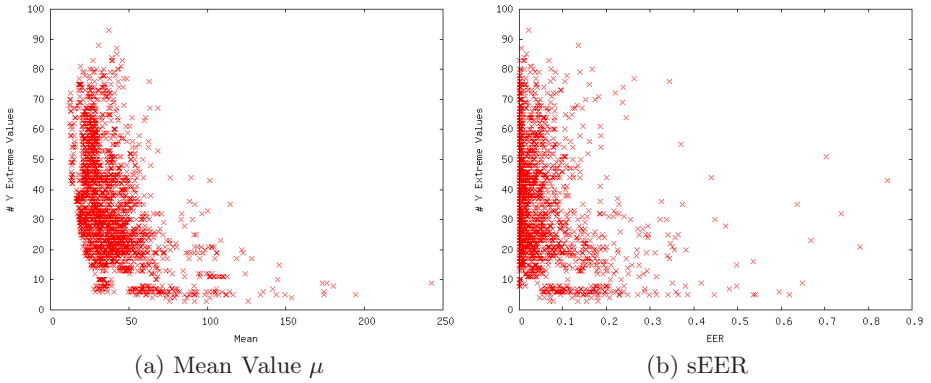
**Fig. 3.** Correlation between number of pen-down strokes and a-posteriori quality



**Fig. 4.** Correlation between average writing speed and a-posteriori quality



**Fig. 5.** Correlation between average absolute value of curvature and a-posteriori quality



**Fig. 6.** Correlation between number of Y extrema and a-posteriori quality

**Number of extrema of Y channel.** This is another example of a nonlinear relationship that can be seen in the plot (Fig. 6). The correlation coefficient when relating the number of extrema to the mean  $\mu$  is  $-0.453$ , and – as with all examined features – the correlation coefficient with the sEER is quite low ( $-0.249$ ).

## 4 Summary and Outlook

Our research confirmed the widespread view that the evaluation of biometric sample quality is a difficult problem. Among the a-priori signature features that have been tested for correlation with the chosen a-posteriori quality measures, there are some that allow a faint prediction of a signature’s stability, but none that allow a prediction of a signature’s forgeability. Stronger correlations with a-posteriori quality measures may be found by looking at feature vectors instead of individual features as in [4]. This is future work.

This paper investigated the influence of the character (the inherent features) of a biometric sample on its utility. There is also an influence of a sample’s fidelity on its utility that should be investigated. The influence of fidelity can be observed by using samples captured with different sensors and sampling rates. Although the influence of different capture devices has been observed before [12], its implications on quality considerations is an open question and will also be the subject of future research.

## Acknowledgments

The authors are grateful to J. Ortega-Garcia and J. Fierrez-Aguilar for making a subcorpus of the MCYT database available for research purposes.

## References

1. Information technology – Biometric application programming interface – Part 1: BioAPI specification. International Standard ISO/IEC 19784-1 (2006)
2. Information technology – Common biometric exchange formats framework – Part 1: Data element specification. International Standard ISO/IEC 19785-1 (2006)
3. Information technology – Biometric data interchange formats. Multi-Part International Standard ISO/IEC 19794
4. Tabassi, E., Wilson, C.R., Watson, C.I.: Fingerprint image quality. NIST Interagency Report NISTIR 7151, NIST, Gaithersburg, MD, USA (2004)
5. Information technology – Biometric sample quality – Part 1: Framework. Working Draft ISO/IEC WD 29794-1 (2007)
6. Biometric sample quality standard. INCITS Draft M1/05-0306 (Revision 4) (2005)
7. Kholmatov, A., Yanikoglu, B.A.: Identity authentication using improved online signature verification method. *Pattern Recognition Letters* 26(15), 2400–2408 (2005)
8. Yeung, D.-Y., Chang, H., Xiong, Y., George, S., Kashi, R., Matsumoto, T., Rigoll, G.: SVC2004: First international signature verification competition. In: Zhang, D., Jain, A.K. (eds.) ICBA 2004. LNCS, vol. 3072, pp. 16–22. Springer, Heidelberg (2004)
9. Sakoe, H., Chiba, S.: Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on acoustics, speech and signal processing* 26(1), 1–2 (1978)
10. Ortega-Garcia, J., Fierrez-Aguilar, J., Simon, D., Gonzalez, J., Faundez-Zanuy, M., Espinosa, V., Satue, A., Hernaez, I., Igarza, J.-J., Vivaracho, C., Escudero, D., Moro, Q.-I.: MCYT baseline corpus: a bimodal biometric database. *IEE Proceedings Visual Image Processing* 150(6), 395–401 (2003)
11. Fierrez-Aguilar, J., Nanni, L., Lopez-Peñalba, J., Ortega-Garcia, J., Maltoni, D.: An on-line signature verification system based on fusion of local and global information. In: Kanade, T., Jain, A., Ratha, N.K. (eds.) AVBPA 2005. LNCS, vol. 3546, pp. 523–532. Springer, Heidelberg (2005)
12. Elliot, S.J.: A Comparison of On-Line Dynamic Signature Trait Variables Across Different Computing Devices. PhD thesis, Purdue University (2001)