# Baseline Extraction and Evaluation for Off-Line Signature Verification

**Bence Nagy, Bence Kővári**

Budapest University of Technology and Economics,
Department for Automation and Applied Informatics
Goldmann György tér 3, H-1111 Budapest, Hungary
nb606@t-online.hu, beny@aut.bme.hu

*Abstract: Present-day off-line signature verification methods definitely could and should be improved, considering that not even the best systems can achieve lower error rates than 5 percent. In this paper we present an off-line comparison method for differentiating between genuine and forged signatures based on feature matching, specifically baseline matching. Since a highly modularized framework has already been created, we developed different modules that suited that system, and were able to create a module chain that extracted baseline information from the signatures, and using the knowledge gained from a small learning set could decide whether the signature was forged or genuine. Of course verification based on only one feature can not be perfect, but the results imply that involving additional – more or less – independent features of the signature can decrease the error rate of the system below the barrier of 5 percent.*

*Keywords: signature verification; off-line; signature baseline; feature matching*

## 1 Introduction

Signature recognition is probably the oldest biometrical identification method, with a high legal acceptance. Forged signatures not only cause economic problems – in the course of cheque frauds – but legal difficulties as well, for instance in the form of forged signatures on a divorce application, hence it is vital to be able to differentiate between genuine and fake signatures. Over the decades two different types of computer based verification techniques arose: on-line and off-line signature verification.

The on-line methods take advantage of dynamic characteristics like velocity, acceleration or even the position and angle of the pen, however this information is not available in most real-life situations, as it requires capturing the process of signing (by camera, digital tablet, etc.). In contrast the off-line methods do not require any special hardware, only the signature itself which makes them more user-friendly, but unfortunately more limited as well. In the past decade a bunch of solutions has been introduced to overcome the limitations of off-line signature verification and to compensate for the loss of accuracy.

This paper is organized as follows: In Section 2 related works are examined, then in Section 3 our algorithm is introduced and explained in detail. The last two sections of the paper show some of our latest experimental results and conclusions drawn from them.

It is also important to note, that this algorithm is part of a bigger project (see references) therefore this paper will not try to cover the whole process of signature verification in detail, allowing us to concentrate on the baseline extraction and evaluation.

## 2 Related Work

In the field of off-line verification there are attempts to authenticate signatures by using image transformations (e.g. the Radon transformation in [1]), but those methods do not take the semantic information into account, and thus their results can not be explained by the means of graphology, and therefore they are especially hard to improve. Most of these techniques are used to filter random forgeries before actual verification begins (like shape matrices in [2]).

The most important limitation off-line verifiers have to face is the absence of temporal information, which allows us to match the selected features of the signatures. Thus on-line methods only need to concentrate on the comparison of the given features [3]. In the off-line case we have to provide a matching between the features before the comparison. In [4] a method is proposed to match strokes of signatures from the same signer for off-line verification.

In [5] an algorithm is proposed to retrieve the main axis of a signature using the convex hull, while the envelope (contour) characteristic of signatures is used in [6] and it is shown that combining feature-based classifiers can increase the accuracy of the verification process.

A highly modular framework has already been created, which divides the process of signature verification into 5 phases: acquisition, preprocessing, feature extraction, processing, classification [7]. Extracting baseline information obviously needs a feature extraction module and less obviously a processing module, where distance values are computed for signature pairs.

## 3 Proposed Method

The very first step of acquiring the baseline of a signature is to define what a baseline is. Using the knowledge gained in a consultation with a handwriting expert a definition was created which presumably would be helpful at the comparison: the baseline is a set of straight lines, where each line represents an imaginary foundation of a component, which can be regarded as an autonomous element of the signature. This definition allows us to assign baselines to the gaps between signature elements, and according to [8] those spaces are just as peculiar as any other feature of a signature.

Our first approach was a naive algorithm, where the goal was to obtain the lower contour of the signature, as it can be seen in Figure 1. This was done by starting vertical scan lines from the bottom left corner of the image and store the lowest pixel which was part of the signature. To determine whether a pixel is representing paper or ink a function was created, which not only used the pixel itself but its environment as well to give the best result. This was necessary because of the different kind of images with different amount of noise on them.

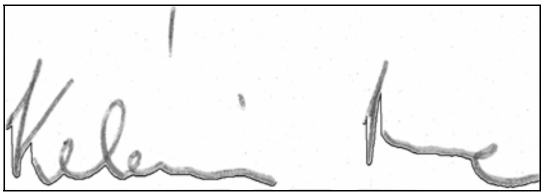

Figure 1
The lower contour of a signature

After obtaining the lower contour a line to each separable segment was fitted using linear regression. Separable segments are parts divided by a horizontal gap. The resulting lines were often convincing enough, but of course this algorithm has its drawbacks: the most important problem was that it frequently had trouble recognizing the separate parts of a signature, in fact it could only distinguish two segments when a significant horizontal gap existed between them, but unfortunately this was not true for most of the signatures in our database. Another remarkable disadvantage was that it used information only from the image itself, while by the time there was additional information available ([4]) that could definitely increase the reliability of the algorithm, like stroke positions.

Using the experience gained so far, we came up with a new approach, whose fundamental element became components. Components are parts of the signature that could and should be treated as an independent part, having their own baseline. Typically a component is a part of the name (first name, last name) or an accent. Experiments showed that accents should have their own baselines, as they are a distinctive feature of signatures.

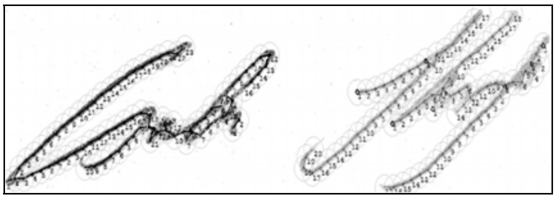

Figure 2
Signatures are segmented into components based on stroke information

The next task was to find these components, using the stroke position information. Two strokes are called neighbors if they are closer than a pre-defined distance value, and two strokes are in the same component if they can be connected by neighboring strokes. Using these simple rules the components of the signatures were separated, and the naive algorithm was applied on the components, with some extensions. Those improvements included some optimization – scanning only within the components bounding box – and the exclusion of strokes from other components when determining the lower contour. The resulting lines seemed to almost perfectly represent our definition of baselines (Figure 3); hence it was time to examine whether they could be used to make a distinction between forged and genuine signatures.

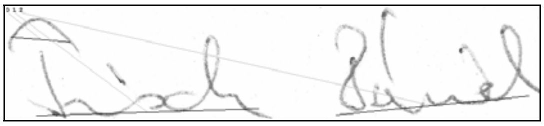

Figure 3
Three baselines obtained by our algorithm

It is important to note, that in some cases more baselines were found than expected (as in Figure 3), but since those extra baselines were found on almost all genuine signatures of the given signer, they should be considered as a feature of the signature and not as an error.

Representing the baselines with numeric values (angle, length, position) and examining these values for both forged and genuine signatures has shown that our baselines were adequate features to verify signatures, however alone they are not sufficient to unquestionably separate valid and forged ones.

# 4 Experimental Results

To test our modules and algorithms a signature database [9] consisting of two different types of samples were used: there were signatures gathered at our university, with forgeries created by students, the other half of signatures were reconstructed from dynamic data publicly available from the First International Signature Verification Competition [10]; henceforth the former set will be referred to as the BUTE database, while the latter one as SVC. The testing environment required 20 genuine signatures and 10 forgeries from every signer, 10 genuine signatures were used to train our system, and the remaining signatures were used as the test set.

It was vital to describe the signatures with simple values to be able to compare them, thus the already known length, angle and position values were used to demonstrate how precisely this algorithm separates forged signatures from genuine ones. It is worth noting, that the count of found baselines were almost constant in case of genuine signatures, while at forged ones it clearly showed a significant fluctuation.

As seen in Figures 4 and 5 almost all measurer can separate the forged signatures from the genuine ones with a confidence of almost 85 percent (when accepting values from the range colored light gray), implying that together they would be even more successful at the separation.

Even though the above results suggest an unexpectedly low error rate (considering only one feature was taken into account), the final output of the system which highly depends on other modules does not reflect that. One cause of this inconsistency lies in the performance of the current classifier modules of the verification system.

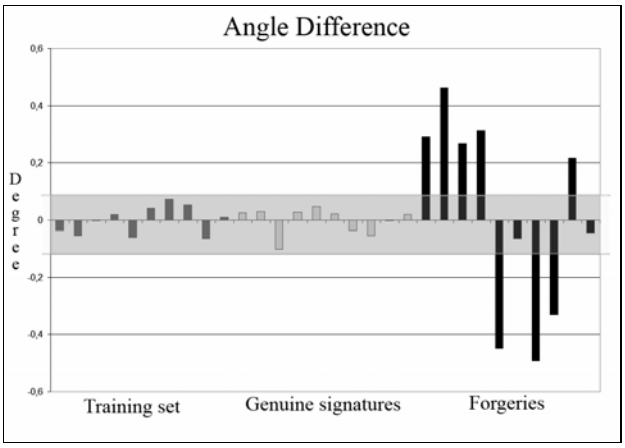

Figure 4
Angle differences

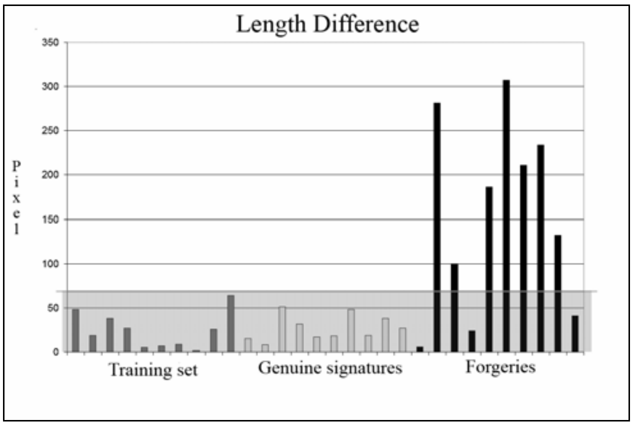

Figure 5
Length differences

The accuracy of our verification system could be represented by the following three variables: false acceptance rate, false rejection rate and average error rate. False acceptance means that a forged signature is evaluated as a genuine one, while false rejection means a genuine one is rejected. The SVC database contained 40 signers, 20 of those had European type signatures, and they were tested separately and as a whole as well. The BUTE database contained only European signatures. As seen in Figure 6 the system clearly prefers European signatures to oriental ones, but maintains an acceptable error rate even with those signatures.

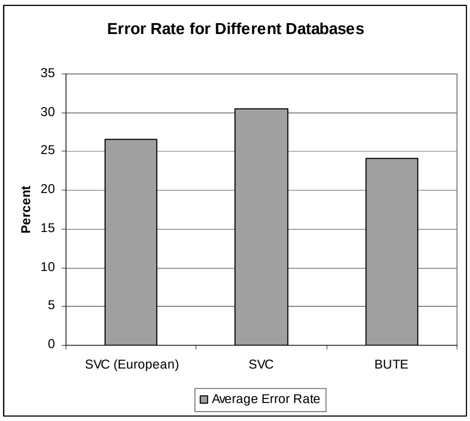

Figure 6
Error rates for different signature sets

## Conclusions

The separation of forged and genuine signatures by their baselines produces an average error rate of 20 to 30 percent, but with better classifiers a better result would be achievable based on the experiments. It is important to note that the FAR (False Acceptance Rate) is much higher than the FRR (False Rejection Rate), which is a reassuring result, considering security reasons. It can be seen that separation based on a single feature is an almost impossible challenge, but by involving other – more or less – independent features in the process the error rate can be decreased dramatically, even below the barrier of 5 percent.

## Acknowledgement

This project was supported in part by the Innovation and Knowledge Centre of Information Technology BME(IT)$^2$, the National Office for Research and Technology (NKTH) and the Agency for Research Fund Management and Research Exploitation (KPI).

## References

[1] Ben Herbst and Hanno Coetzer: On an Automated Signature Verification System, *Proceedings of the 9th annual South African Workshop on Pattern Recognition*, 1998, pp. 39-43

[2] R. Sabourin, J.-P. Drouhard, and E. S. Wah: Shape Matrices as a Mixed Shape Factor for Off-line Signature Verification, *In Proc. Intern. Conference on Document Analysis and Recognition (ICDAR)*, Ulm, Germany, 1997, pp. 661-665

[3] J. Galbally, J. Fierrez, M. R. Freire, J. Ortega-Garcia: Feature Selection Based on Genetic Algorithms for On-Line Signature Verification, *IEEE Workshop on Automatic Identification Advanced Technologies*, Alghero, Spain, 2007, pp. 198-203

[4] Bence Kővári, Gergely Kiss, Hassan Charaf: Stroke Extraction and Stroke Sequence Estimation for Off-line Signature Verification, *The Eighth IASTED International Conference on Visualization, Imaging, and Image Processing*, Palma de Mallorca, Spain, 2008

[5] E. Frias-Martinez, A. Sanchez, J. Veleza: Support Vector Machines versus Multi-Layer Perceptrons for Efficient Off-line Signature Recognition, *Engineering Applications of Artificial Intelligence Volume 19, Issue 6*, September 2006, pp. 693-704

[6] Ramesh, V. E., Murty, M. Narasimha: Off-line Signature Verification using Genetically Optimized Weighted Features, *Pattern Recognition Vol. 32, No. 2*, February 1999, pp. 217-233

[7] Bence Kővári, István Albert, Hassan Charaf: A General Representation for Modeling and Benchmarking Off-line Signature Verifiers, *12th WSEAS Int. Conf. on COMPUTERS*, Heraklion, Greece, 2008

[8] R. A. Huber and A. M. Headrick: Handwriting Identification: Facts and Fundamentals, 1999, *CRC Press, LCC*.

[9] Bence Kővári: Signature Database for Off-line Signature Verification. [Online] 2006, http://www.aut.bme.hu/signature/

[10] Dit-Yan Yeung, et al.: SVC2004: First International Signature Verification Competition, *Lecture Notes in Computer Science, Biometric Authentication, Volume 3072/2004*, pp. 16-22