MODELS AND ALGORITHMS IN OFF-LINE, FEATURE-BASED, HANDWRITTEN SIGNATURE VERIFICATION

MODELEK ÉS ALGORITMUSOK A HAGYOMÁNYOS ALÁÍRÁSOK STATIKUS, JELLEMZŐ ALAPÚ HITELESÍTÉSÉBEN

Ph.D. Thesis Booklet

Bence András Kővári

Advisor:
Dr. Hassan Charaf Ph.D.
Associate Professor

BUDAPEST UNIVERSITY OF TECHNOLOGY AND ECONOMICS
DEPARTMENT OF AUTOMATION AND APPLIED INFORMATICS

Budapest, 2013.
Bence András Kővári

Budapest University of Technology and Economics
Faculty of Electrical Engineering and Informatics
Department of Automation and Applied Informatics

1111 Budapest, Magyar tudósok krt. 2/B.

e-mail: kovari@aut.bme.hu
tel.: +36(1)4632486
fax: +36(1)4632871

Advisor:
Dr. Hassan Charaf Ph.D.
Associate Professor
1 Preliminaries and objectives

The verification of handwritten signatures is one of the oldest biometric identification methods. As it gained a high legal acceptance, both the methods of forgers and verifiers became more elaborate. At the beginning of the 20th century, anyone who was in some way professionally connected to handwriting (teachers, notaries) could be treated as handwriting experts (Kaszab, et al., 2003). Today, however, being a forensic document examiner is an independent profession, which requires special training and a rich technological toolkit. There are even some guides (Bryan & Doug, 1999) to modularize and organize the human workflow of the verification process. Despite these facts, the task of signature verification is still challenging, even for a human expert.

When confronted with professionally forged signatures, the average error rate of a forensic expert lies between 0.5% and 7% (Found & Rogers, 2003) while non-experts’ results may be much worse (Jain, 2010), as error rates between 10% and 26% were also recorded. These numbers show that even the opinions of forensic document examiners are prone to human error and many open questions remain in the field. Questions such as: Why is it that the experts were unable to achieve better results? Are the previous numbers due to the limitations of the verifiers, or the limits of the discriminative power of handwritten signatures? Now, however, rapid advancements in the field of computer science have made it possible to analyze some of these questions more thoroughly.

The aim of computer-based signature verification is to decide, automatically, whether a given signature (question signatures) belongs to a given person. The decision must be based solely on some samples (reference signatures) from the signer. Depending on the format of the samples, the field can be divided into two main categories, the on-line (dynamic) and the off-line (static) approach.

In on-line signature verification, the entire signature process is captured using some kind of an acquisition device (a camera, a digitizing tablet, a stylus-operated PDA etc.), analyzed and then used to make a decision. The captured information usually includes pen position, pen pressure, pen azimuth and pen inclination as a function of time. This set of data gives automatized verification two important advantages. First, because the data is available as a function of time, it is easier to identify corresponding parts. Second, this kind of acquisition records information which is not directly available to the forger (like pen velocity, or pressure). Even when producing an almost perfect visual copy of a signature, these invisible factors will usually differ significantly from the values measured during the original signing process. This is why state of the arte on-line signature verification systems can produce error rates of less than 1% (Impedovo & Pirlo, 2008).

On the other hand, the aim of off-line signature verification is to decide whether a signature originates from a given signer based on the image of the questioned signature and a few sample images of the person’s original signatures. This means the input is limited to two-dimensional images, while important pieces of information like velocity, inclination or pressure are mostly lost. Unlike on-line signature verification, which requires a special acquisition hardware and setup, off-line verification can be performed separately from the normal signing process and is thereby less intrusive and more user-friendly. Moreover, the off-line scenario can have a much wider range of practical applications as it can be seamlessly incorporated into many existing workflows. For example, the majority of financial institutions already digitize their contracts, mail or transfer orders. Therefore, the addition of automatic signature verification to their workflows would require “just” a software upgrade which would not directly affect any clients and only a few employees.
On-line signature verification has a clear advantage over human experts, because of the captured non-visual information. Whereas off-line verification is at a disadvantage, because it works with the same input but lacks the extensive background knowledge and human expertise. Because of this, even the best off-line signature verifiers on the most studied databases cannot break the 9% error barrier (Vargas, et al., 2011) (Batista, et al., 2012). This limits their practical applicability significantly.

Several different approaches exist to overcome these limitations. However, most of the signature verification systems have one thing in common: They behave similar to black boxes (mainly due to the software’s AI-based approach), providing only a limited amount of meaningful information regarding each decision (Kovari, et al., 2009). This makes their improvement a difficult task.

As long as human experts can outperform automatic signature verifiers, there is a need for improvement. Therefore, the objective of my research was to establish a signature model and evaluation-methodologies which are suitable for the classification of signatures. Taking these aspects into account, my goals were the following:

- **Investigate the common characteristics in existing off-line signature verification approaches and identify their strong and weak points**
- **Examine and, when necessary, improve existing feature extraction algorithms**
- **Identify dominant signature features which may be used as discriminating factors.**
- **Establish and investigate classification techniques to differentiate between original and forges signatures**
- **Apply methods to signature databases, and demonstrate the feasibility of their practical application**

### 2 Methodological Summary

The requirements above determined the direction of my research and the individual tasks that should be solved in succession. The logical starting point of my work was to examine the already existing methods, models and algorithms. Firstly, I enumerated the existing approaches then, collected the most important models and algorithms. The methods of the theoretical research were determined by the already existing sub-solutions, and the shortcomings of the existing solutions. The entirety of the research aimed to improve the present signature verification models and can be described by the following tasks:

- **Review of most relevant journal and conference papers from the last 30 years in the field of signature verification.**
- **Realization of several image processing and signature verification algorithms, based on scientific papers and personal ideas.**
- **Design and implementation of a modular software framework that allows for the independent measurement of the efficiency of signature verification phases.**
During my work, I have accessed several international signature databases (GPDS300, SVC2004, SigComp2009) and also created several of my own databases (altogether more than 10,000 signatures) each of them focusing on different signature aspects.

- The creation of a statistical model for the purposes of signature verification.
- Deepening my knowledge of mathematical statistics, with a special emphasis on inferential statistics and the elaboration of the statistical verification model.
- Conducting experiments to confirm the validity of my solutions and demonstrate their applicability.

3 Novel Scientific Results

As the best current systems use artificial intelligence-based approaches, the insight gained, into the reasons behind classification decisions, is very limited. In order to achieve a better understanding of the process as a whole, in my first thesis I propose a simplified probabilistic model for off-line signature verification. In this model, each of the verification steps can be mathematically described and, therefore, individually analyzed and improved. Using this model, it is possible to predict the accuracy of a signature verification system based on just a few a priori known parameters, such as the cardinality and the quality of input samples. Several experiments have been conducted using a statistics-based classifier to confirm the assumptions and results of my model.

In my second thesis, some of the consequences of the previously introduced model are analyzed. It is proven, therein, that increasing the number of reference samples or the number of observed features will improve the achievable classification results. It has also been demonstrated that the only free parameter of my equation – which describes the quality of the forged signatures – can be specifically set to minimize the negative effects of a possible misestimation.

My third thesis deals with feature extraction and matching algorithms used during the process of signature verification. A heuristic stroke extraction, a baseline extraction, a loop extraction and a feature-matching algorithm are introduced and described in detail.

The rest of this work is organized as follows: First, the main contribution of each thesis is described. Second, reference to the appropriate chapter of the dissertation followed by my publications related to the specific thesis is presented. Finally, the detailed description of the thesis is provided.

THESIS I
Statistical Model of Feature-Based Off-Line Signature Verification

I have created a statistical model for the classification phase of off-line feature based signature verification. Using this model I was able to calculate the optimal threshold for the rejection or acceptance of a feature property with minimal error. I have extended this model to incorporate the uncertainty resulting from low sample sizes and I have calculated
the total average error rate of a signature verification system, based on my model. I have provided experimental evidence for the presumptions of the statistical model and I have shown that the calculations of my model are in close proximity to the actual experimental results.


**Definition 1** (Terminology)

*Feature*: a feature is a part of the signature described by a set of properties. We distinguish two types of features:

- **global feature**: a feature which covers the whole signature (e.g. size of signature)
- **local feature**: a feature which only covers a smaller part of the signature (e.g. a loop, a baseline)

*Feature property*: a number describing a given aspect of a feature (e.g. height of a loop)

*Feature type*: describes a group of features that have the same properties (e.g. loops, baselines)

**Assumption 1**

A feature property can be approximated as a normally distributed random variable \( X \), with mean \( m \) and variance \( s^2 \)

\[
X \sim \mathcal{N}(m, s) \tag{3.1}
\]

**Assumption 2**

A forged feature property can be approximated as a normally distributed random variable \( Y \).

\[
Y \sim \mathcal{N}(m_f, s_f) \tag{3.2}
\]

In addition, because the forger is aiming to forge the original signature, the means of these distributions overlap.

\[
M(X) = M(Y) \Rightarrow m_f = m \tag{3.3}
\]

Therefore, the only difference between original and forged feature values is in their deviations. We define the quality of a forgery \( q \) as their quotient.

\[
q = \frac{s_f}{s} \tag{3.4}
\]
Sub-Thesis I.1

I have proven that, provided \( m \) and \( q \) are known, the sum of the probabilities of false rejection and false acceptance is minimized when:

\[
w = \pm \sqrt{2} \sqrt{\ln q \frac{q^2}{q^2 - 1}}
\]  

(3.5)

I have also proven, this decision is will result in the following average error rate:

\[
AER = \frac{1}{2} + \Phi \left( \frac{w}{q} \right) - \Phi(w)
\]  

(3.6)

Sub-Thesis I.2

I have shown that the conclusions of Equation 3.5 may be refined by approximating the distributions with Student’s t-distributions. In this case the error is minimized if:

\[
w = \pm \sqrt{\left( \frac{q^2 - q^2 + 2n}{q^2 - q^2 + 2n} \right) \left( \frac{n^2 - 1}{n} \right)}
\]  

(3.7)

Sub-Thesis I.3

I have proven that by using the decision threshold, introduced in Equation 3.7, the error rates of a verification system working on a single feature property can be calculated as:

False rejection rate:

\[
\alpha = P(m - ws > X) + P(m + ws < X) = 2T_o(\bar{m}_n - w\hat{s}_n)
\]  

(3.8)

False acceptance rate:

\[
\beta = P(m - ws < X < m + w\sigma_o) = 1 - 2T_f(\bar{m}_n - w\hat{s}_n)\]

(3.9)

Average error rate:

\[
AER = \frac{\alpha + \beta}{2} = \frac{1}{2} + T_o(\bar{m}_n - w\hat{s}_n) - T_f(\bar{m}_n - w\hat{s}_n)q
\]

(3.10)

Where \( T_o(x) \) and \( T_f(x) \) denote the cumulative distribution function of the original and forged samples, respectively.

Equation 3.10 describes the accuracy of a decision that is based on a single property of a feature. However, in real world scenarios, we are able to utilize several different feature properties. Therefore, the next sub-thesis extends the previous model for multiple feature properties.
Sub-Thesis I.4

I have proven that considering \( k \) independent feature properties, in which each of the feature properties is either rejected or accepted with known \( \alpha \) and \( \beta \) error rates, the number of rejected features required to reject a signature while minimizing the average error rate is:

\[
l = k \frac{\ln(1-\alpha) - \ln(\beta)}{\ln(1-\alpha) - \ln(\beta) + \ln(1-\beta) - \ln(\alpha)}
\]  

(3.11)

Sub-Thesis I.5

Using the previously introduced equations I have shown, that the lowest achievable average error rate of a signature verification system can be predicted based on the number of reference signatures used to train the system \( n \), the number of independent feature properties \( k \) and the quality of the forged features \( q \), as follows:

\[
AER(q, n, k) = \frac{\sum_{j=0}^{\lfloor l \rfloor} \binom{k}{j} \beta^{k-j} (1-\beta)^j + \sum_{i=\lfloor l \rfloor + 1}^{k-1} \binom{k}{i} \alpha^i (1-\alpha)^{k-i}}{2}
\]  

(3.12)

THESIS II

The Implications of the Statistical Model

I have proven that, in my model, increasing the number of reference samples or the number of observed features will both decrease the average error rate of the system. I have formulated a conjecture that states the error introduced through improper assessment of the forgery quality can be minimized.


Sub-Thesis II.1

Given the constraints

\[
q > 1, k \geq 1, n > 2, n, k \in \mathbb{Z}, q \in \mathbb{R}
\]  

(3.13)

and provided that the number of reference samples \( (n_0) \) and the forgery quality \( (q_0) \) is given, I have proven that the function \( AER(q_0, n_0, k) \) is decreasing.
Sub-Thesis II.2

*Given the constraints*

\[
q > 1, k = 1, n > 2, n, k \in \mathbb{Z}, q \in \mathbb{R} \tag{3.14}
\]

and provided that the number of observed features (\(k_0\)) and forgery quality (\(q_0\)) is given, I have shown that the function \(\text{AER}(q_0, n, k_0)\) is strictly decreasing.

Sub-Thesis II.3

I have formulated the following conjecture and provided simulations to support it:

*Suppose the value of the actual forgery quality (\(q_{act}\)) falls in the interval \((1; q_{max})\). there is an ideal value for estimated forgery quality (\(q_{est}\)) that minimizes the errors resulting from a suboptimal choice for \(w\).*
Figure 3.2 The resulting increase in error depending on the maximal value ($q_{\text{max}}$) and the estimated value ($q_{\text{est}}$) of forgery quality

**THESIS III**

**Signature Processing Algorithms**

*I have created a collection of image processing algorithms with applicability in signature verification, including algorithms for extraction of stroke-, baseline- and loop-information, and a feature matching algorithm.*

Thesis III is contained in Chapter 5 of the dissertation. Related publications: [4] [5] [6] [7] [9] [10] [13] [14] [15] [17] [18] [19] [21] [22] [23] [24] [26] [27].

**Sub-Thesis III.1**

*I have created the Scanning Circle Algorithm for the extraction of stroke information from signature images. I have solved the problem of detecting sudden direction changes in the strokes by using an adaptive scanning circle radius and provided a locally optimal algorithm for estimating the drawing sequence of directly connected strokes. Experimental results show – in terms of thinning – the accuracy of the Scanning Circle Algorithm is comparable to the best thinning algorithms examined.*

**Sub-Thesis III.2**

*I have proposed a heuristic algorithm for the solution of multidimensional assignment problem regarding the off-line signature verification scenario. I have defined a simplified version of the algorithm for matching a sample signature to the reference signatures.*
Sub-Thesis III.3

*I have proposed a way of characterizing baseline information and provided algorithms for the extraction of these features from signature images.*

Sub-Thesis III.4

*I have proposed a way of characterizing loop information and provided algorithms for the extraction of these features from signature images.*

4 Application of the Novel Scientific Results

Related publications: [4] [6] [12] [13] [14] [15] [17] [21] [28].

This thesis is the result of extensive research conducted over seven years. During this period, a rich application system was built around the core of our signature verification system. I appreciate the help of my students who contributed to the implementation of the system and therefore, have been my coauthors in several publications. Moreover, the development of 10 Master’s thesis, 7 bachelor’s thesis and a Scientific Students’ Associations Conference paper, closely related to signature verification, are the fruit of their labor.

The signature verification framework is a collection of 37 C# projects, consisting of approximately 148,000 lines of code. About 10% of this codebase belongs to a single, reusable class library, which contains the most mature modules as well as the framework itself. At the time, 32 different modules were created for the system.

The Signature Manager (SigMan) (Kutasi, 2010) application was created by Bence Kutasi to assist in the visual analysis of signatures and features. The highly modular Windows Presentation Foundation (WPF) based application is able to visualize and edit the previously extracted features beside the signature.

![Figure 4.1 Screenshots from the SigMan application](image)

Mátyás Pálfalvi created a public interactive frontend (Pálfalvi, 2012), available on the homepage of the Department of Automation and Applied Informatics at BME. The tool provides a simple set of original and forged signatures to demonstrate the capabilities of the system swiftly, but one can also upload (through dragging and dropping) his/her own signatures for testing.
The user interfaces used during development and testing are overcrowded and complex. The WPF based user interface created by Daniel Nagy (Nagy, 2010) demonstrates that the complex inner logic of the system can be presented in a clear and user-friendly way.

As a unique feature among off-line signature verifiers, our system is able to create a report about the verification results of a signature to explain the decision. Instead of abstract results, the report provides detailed visual feedback about the main processing results, as well as statistical parameters for each feature. Using these results, a human expert can easily confirm or inquire about a decision made by the system. The current implementation of our reporting component was created by Zoltán Berceli (Berceli, 2012) and Mátyás Pálfalvi (Pálfalvi, 2012).
5 Scientific Publications

International Journals


Conference Proceedings


[9] Mark Vincze, Bence Kovari: Comparative survey of thinning algorithms, 10th International Symposium of Hungarian Researchers on Computational Intelligence and Informatics (CINTI), Budapest, Magyarország, 2009


[12] Daniel David Nagy, Bence Kovari: Improving Designer-Developer Workflow for Better User Experience, 10th International Symposium of Hungarian Researchers on Computational Intelligence and Informatics (CINTI), Budapest, Magyarország, 2009


[16] Adam Horvath, Bence Kovari: Usability of Neural Networks in Off-line Signature Verification, 10th International Symposium of Hungarian Researchers on Computational Intelligence and Informatics (CINTI), Budapest, Magyarország, 2009

[17] Bence Nagy, Bence Kovari: Baseline Extraction and Evaluation for Off-line Signature Verification, 9th International Symposium of Hungarian Researchers on Computational Intelligence and Informatics (CINTI), Budapest, Magyarország, 2008


ACS'07, The 7th WSEAS International Conference on Applied Computer Science, 
Velence, Olaszország 2007

[25] **Bence Kovari**: The development of off-line signature verification methods, 
comparative study microCAD 2007 International Scientific Conference, Miskolc, 
Magyarország 2007

[26] **Bence Kovari**, Zsolt Kertész, Attila Major: Off-Line Signature Verification Based on 
Feature Matching INES, 11th International Conference on Intelligent Engineering 
Systems, Budapest, Magyarország 2007

[27] **Bence Kovari**: Extraction of Dynamic Features for Off-line Signature Analysis 
Automation and Applied Computer Science Workshop (AACS), 2006

International Scientific Conference, Miskolc, Magyarország 2005

6 Citations

[3] is cited by:

Hemanta Saikia, Kanak Chandra Sarma: Approaches and Issues in Offline Signature 
March 2012

Gulzar Ali Khuwaja, Mohammad Shakeel Laghari: Learning Vector Quantisation based 
4, Number 2, pp. 116-129, 2012

[4] is cited by:

Sharil Tumin, Sylvia Encheva: A Multi PurposeWeb-based Contractual Management 
System, *WSEAS Transactions on Information Science and Applications* 7, pp. 819-829, 
2010

Milan Tuba, Nadezda Stanarevic: Relation between Successfulness of Birthday Attack on 
Digital Signature and Hash Function Irregularity, *WSEAS Transactions on 
Information Science and Applications* 7, pp. 186-195, 2010

Milan Tuba: Digital signature and hash function irregularity, *Proceedings of the 8th WSEAS 
International Conference on Telecommunications and Informatics*, pp. 169-174, 2009

[5] is cited by:

Vu Nguyen, Michael Blumenstein: Techniques for static handwriting trajectory recovery: a 
survey, *Proceedings of the 9th IAPR International Workshop on Document Analysis 
Systems*, Boston, Massachusetts, pp. 463-470 , 2010
[8] is cited by:


[9] is cited by:


[11] is cited by:


[16] is cited by:


[18] is cited by


[20] is cited by:


[23] is cited by:


[24] is cited by:


[26] is cited by:


Shashi Kumar D. R., K. B. Raja, R. K. Chhotaray, Sabyasachi Pattanaik: Off-line Signature Verification Based on Fusion of Grid and Global Features Using Neural Networks, /


7 References


Pálfalvi, M., 2012. Bachelor's thesis. *Interaktív aláíráshitelesítési frontend Silverlight platformon (Interactive signature verification frontend on Silverlight platform).*
