On-line Handwritten Signature Identification: The Basics

Tomislav Fotak, Petra Koruga, Miroslav Bača
Faculty of Organization and Informatics, Centre for Biometrics
University of Zagreb
Pavlinska 2, 42000 Varaždin, Croatia
{tomislav.fotak, petra.koruga, miroslav.baca}@foi.hr

Abstract. *Handwritten signature is widely accepted and collectable biometric characteristic. Great entropy makes this characteristic suitable for research and development of new methods for personal authentication and identification. While development of authentication methods based on this biometric characteristic is common in academic and research community, there have been only few attempts of developing personal identification systems based on handwritten signature.*
This paper presents basic differences between authentication and identification methods, followed by previous work in the field of handwritten signature identification and main directions in developing an on-line personal identification system based on handwritten signature. This work can be considered as a theoretic base for further development of an on-line handwritten signature identification system.
Keywords. handwritten signature, signature recognition, personal identification, biometrics

1 Introduction

Use of handwritten signature as a mean of giving our consent for an action or set of actions that need to be done has become the most common thing people do everyday. The problem arises when someone is trying to imitate our signature and steal our identity. That person could easily make some damage to us. Therefore, there is a need to know who actually signed a document. This brings us to the field of handwritten signature identification.
Handwritten signature is a biometric characteristic of a person. It belongs to behavioral characteristics, i.e. it depends on the person's behavior, so it has greater entropy than other biometric characteristics. Looking at our signature and asking ourselves why we sign different every time, we notice that handwritten signature depends on almost everything. There are a few key factors our signature depends on:

- *Physical and psychological state of the person* - includes illness, injuries, fears, heart rate, person's age, calmness, goodwill, etc.
- *Body position* - it is not the same if the person is standing or sitting while signing document, where is person looking at the moment, what is the burden on the signing hand, etc.
- *Writing surface and writing material (pen)* - Signature will look different on the various types of paper. It will look different if taken with a digitizing tablet or a specialized pen. Writing with pen, pencil, feather or stylus also impacts person's signature.
- *Purpose of signing* - Signature is usually significantly different if taken in formal environment than in informal.
- *Environmental factors* - Environment and people that surround the signatory. This includes noise, luminance, temperature, humidity, etc.

How person's signature varies is shown in the Fig. 1. It shows two handwritten signatures of the same person taken in the time interval of only a few seconds.

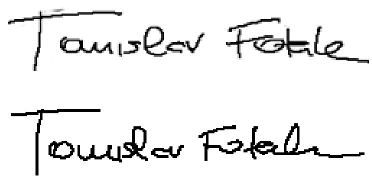

Figure 1: Two signatures of the same person taken within few seconds

According to [1] signature is widely accepted and collectable biometric characteristic. This makes it suitable for further research and development of new personal authentication and identification methods. While development of authentication methods based on this biometric characteristic is a common thing in academic and research community, there are only a few attempts of developing personal identification systems based on handwritten signature.

1.1 Identification vs. authentication

Both identification and authentication are important terms in biometrics. Understanding these gives us basics for the work with the biometric systems.

Biometric authentication is probably simpler and more often used procedure. It answers the question: "Is the person really who he/she tells it is?" User has to provide his/her username and password, but in biometrics instead of using password user provides a biometric characteristic, depending on the biometric system. In this case that characteristic would be handwritten signature. Based on the given signature the system will verify the user authenticity by comparing the signature with a template stored in database. Authentication is often referred as "one-to-one" comparison because it compares one biometric feature to exactly one known template.

On the other hand, biometric identification is known as "one-to-many" comparison. It compares given biometric feature against all templates in the database, thus finding the best match. In the context of handwritten signatures this means that user has to provide only his signature to the biometric system. System will compare his signature with all signatures in the database and calculate match result. Therefore, biometric identification answers the question: "Who is the person?"

The main difference between authentication and identification is shown in Fig. 2.

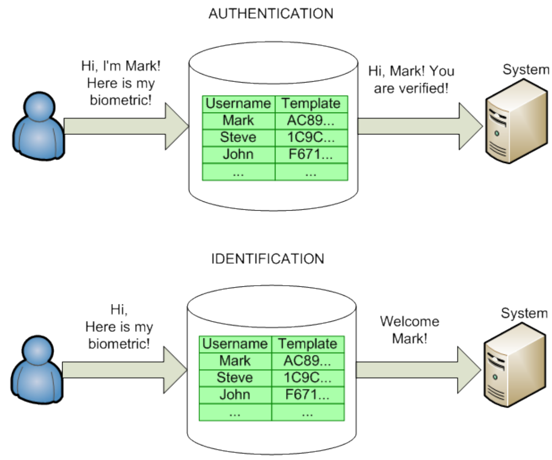

Figure 2: Difference between biometric authentication and biometric identification

2 Previous work

As mentioned earlier, a lot of work has been done in the field of handwritten signature verification, i.e. signature authentication. Some of these works also tried dealing with handwritten signature identification. But, because of the handwritten's signature great entropy it is hard to make a good identification system.

Signature is often equated with a person's handwriting. In the context of this work that is possible because everything that is done on person's handwriting can be applied on the signature. This is probably the reason why most authors write about handwriting recognition and writer identification instead of focusing just on signature. Although they are not synonyms, for the purpose of this work, terms handwriting recognition, writer identification and signature identification will be treated equally.

When dealing with signature identification we can talk about off-line and on-line signature identification. The first one requires only signature image which has to be analyzed in some way and person does not have to be physically present in the moment of the identification. On-line signature identification requires physical presence of the person. It is usually done with a digitizing tablet or a specialized pen which send 'live data' to the biometric system. We will cover both off-line and on-line works.

2.1 Off-line signature identification

Most of the work in the field of signature identification deals with the off-line signature identification, i.e. off-line writer identification.

Said, Tan and Baker [2] presented an algorithm for automatic text-independent writer identification. They took a global approach based on texture analysis, where each writer's handwriting is regarded as a different texture. They applied the multi-channel Gabor filtering technique followed by the weighted Euclidean distance for the recognition task and got result of 96% identification accuracy. The same principle was applied in [3]. It was proven that presented algorithm actually works with the approximately 95.7% identification accuracy. 2-D Gabor filter method has to convolute the whole image for the each orientation and each frequency. This is computational very costly. Using wavelet-based Generalized Gaussian Distribution (GGD) instead is presented in [4]. Authors summarize that compared with Gabor method, GGD method achieves a higher accuracy and significantly reduces the computational time.

Wavelets were also used in [5]. Authors proposed to use the rotated complex wavelet filters (RCWF) and dual tree complex wavelet transform (DTCWT) together to derive signature feature extraction, which captures information in twelve different directions. In identification phase, Canberra distance measure was used.

Different approach, not based on textures is presented in [6] where the distribution of the pixel gray levels within the line was considered. The curve associated with the gray levels in a stroke section was characterized by use of 4 shape parameters. Altogether, 22 parameters were extracted. Three different classifiers were used with and without genetic selection of the most significant parameters for the classifier. Then the classifiers were combined and the results show the gray level distribution within the writing.

Another direction in off-line writer identification process is using mathematical morphology. In [7] the feature vector is derived by means of morphologically processing the horizontal profiles (projection functions) of the words. The projections are derived and processed in segments in order to increase the discrimination efficiency of the feature vector. Both Bayesian classifiers and neural networks are employed to test the efficiency of the proposed feature. The achieved identification success, using a long word, exceeded 95%.

Use of neural networks is very popular in the handwritten signature identification process. Paper [8] combines image processing which consists in extracting significant parameters from the signature image and classification by a multi-layer perceptron which uses the previous parameters as input. The image processing step was described according to the intrinsic features of handwriting. Then, the proposed neural networks were compared with others classifiers as pseudo-inverse, k-nearest-neighbors and k-means and the influence of preprocessing and bad segmentation was measured. For the identification task, they obtained an error rate of 2.8% when there is no rejection, and an error rate of 0.2% when 10% of the signatures were rejected. Another use of neural networks is presented in [9]. They started with breaking the pixels into their RGB values and calculating their corresponding gray scale value which are used to train neural network. They implemented the basic algorithm of artificial neural network through back propagation algorithm and used three (Input, output and hidden) layers, six nodes (three in input layer, two in hidden layer and one in output layer). Artificial neural networks (ANN) were also used in [10] where authors presented an off-line signature recognition and verification system which is based on moment invariant method and ANN with back propagation algorithm used for network training. Two separate neural networks are designed; one for signature recognition, and another for verification (i.e. for detecting forgery). Both networks used a four-step process. Moment invariant vectors were obtained in the third step. They reported 100% signature identification accuracy on the small set of 30 signatures. Back propagation neural network and Radial Basis Function Network were used in [11]. The recognition rate of Radial Basis Function was found to be better compared to that of Back Propagation Network. The recognition rate in the proposed system lied between 90% and 100%.

Other approaches to off-line signature identification include use of Support Vector Machine. In [12] a new method for signature identification based on wavelet transform was proposed. This method uses Gabor Wavelet Transform (GWT) as feature extractor and Support Vector Machine (SVM) as classifier. Two experiments on two signature sets were done. The first is on a Persian signature set and other is on a Turkish signature set. Based on these experiments, identification rate has achieved 96% and more than 93% on Persian and Turkish signature set respectively. SVM has also been used in [13]. This work used Support Vector Machines to fuse multiple classifiers for an offline signature system. From the signature images, global and local features were extracted and the signatures were verified with the help of Gaussian empirical rule, Euclidean and Mahalanobis distance based classifiers. SVM was used to fuse matching scores of these matchers. Finally, recognition of query signatures was done by comparing it with all signatures of the database.

There are other identification methods, but there are only one or two papers that deal with those methods. These include use of Contourlet transform as mentioned in [14]. After preprocessing stage, by applying a special type of Contourlet transform on signature image, related Contourlet coefficients were computed and feature vector was created. Euclidean distance was used as a classifier.

Besides that, use of fractals is mentioned in [15]. Advantage was taken from the autosimilarity properties that are present in one's handwriting. In order to do that, some invariant patterns characterizing the writing were extracted. During the training step these invariant patterns appeared along a fractal compression process and then they were organized in a reference base that can be associated with the writer. A Pattern Matching process was performed using all the reference bases successively. The results of this analysis were estimated through the signal to noise ratio.

One could notice that neural networks are the main approach in the off-line signature identification. This is possible because signature identification can be considered as a pattern recognition problem, where neural networks play an important role. Their implementation has always been of great interest to the researchers.

2.2 On-line signature identification

While on-line signature verification is a common subject among biometric community, there are only a few papers on on-line handwritten signature identification.

Hidden Markov Models are frequently used during authentication process. Therefore, it would be reasonable to apply this approach to handwritten signature identification. Paper [16] describes a Hidden Markov Model (HMM) based writer independent handwriting recognition system. A combination of point oriented and stroke oriented features yields improved accuracy. The general recognition framework is composed of Hidden Markov Models (HMMs), representing strokes and characters, embedded in a grammar network representing the vocabulary. The main characteristic of the system is that segmentation and recognition of handwritten words are carried out simultaneously in an integrated process.

This is only one example of using HMMs. They are usually used for handwritten word recognition, thus it can be applied to on-line signature recognition. Hidden Markov Models are part of the statistical word recognition approach.

Another approach in this field is based on Gaussian Mixture Models (GMMs). In [17] the task of writer identification of on-line handwriting captured from a whiteboard is addressed. The system is based on Gaussian mixture models. The training data of all writers are used to train a universal background model (UBM) from which a client specific model is obtained by adaptation. The system is tested using text from 200 different writers. A writer identification rate of 98.56% on the paragraph and of 88.96% on the text line level is achieved.

A discriminant-based framework for automatic recognition of online handwriting data was presented in [18]. They identified the substrokes that are more useful in discriminating between two online strokes. A similarity/dissimilarity score is computed based on the discriminatory potential of various parts of the stroke for the classification task. The discriminatory potential is then converted to the relative importance of the substroke. An average reduction of 41% in the classification error rate on many test sets of similar character pairs has been achieved.

Many papers from this field actually talk about word recognition. This can be applied to signature identification but is not directly connected with it. This is why we provided small number of references for this part. We can summarize that Hidden Markov Models are so far the most used method in on-line handwritten signature identification.

3 On-line handwritten signature identification system

If one wants to implement signature identification system to gain more security in the company, one would probably use on-line identification system. System architecture of an ordinary on-line handwritten identification system consists of one main module. It is called identification module and it is responsible for all the identification logic. This module contains some of the previously described approaches in signature identification or a completely new approach.

System interacts with user by user interface. User is asked to place his/her signature on some kind of specialized gadget. System records signature main data and derives some new data. Data is then passed to the identification module which also requires data from data template storage. Identification module compares signatory data against all templates in the database, thus finding the best match. Person is identified if best match template satisfies certain predefined rules of identification. Simplified signature identification system architecture is given in Fig. 3.

For the purpose of this work we define twelve initial features to be extracted:

- Number of strokes
- Number of pen-ups
- Signature aspect ratio
- Signature length
- Signing time
- Time-down ratio
- Time-up ratio
- Signature speed

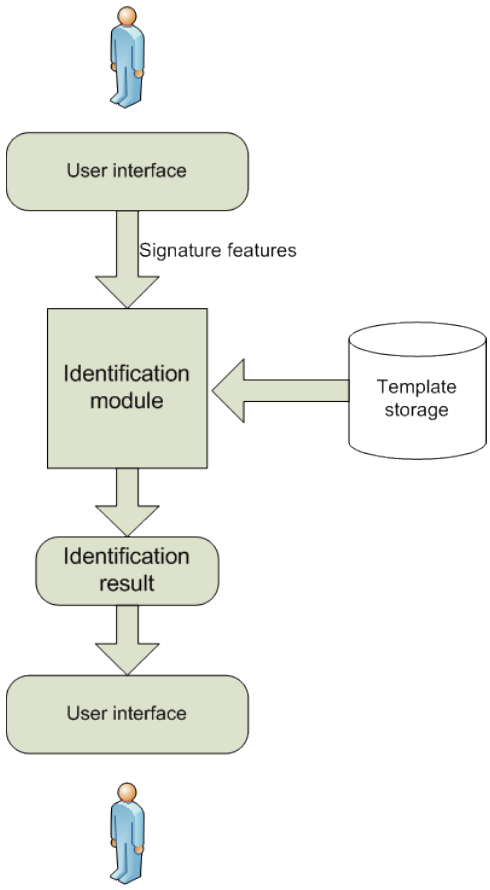

Figure 3: Simplified arhitecture of the on-line handwritten signature identification system

- Velocity along the x-axis
- Velocity along the y-axis
- Average pressure
- Strongest pressure moment

Features are described in detail in [19].

During the user registration process signatories will have to sign no less than 10 times. Statistical measures such as mean, standard deviation, median, minimum value and maximum value will be computed for each signature feature. This data will further be used to create signature template which will be stored in the database. In the initial stage, we will try to identify the person using these features only.

All of these features are extracted using digitizing tablet and are only the beginning in the process of determining the ideal feature subset to be used for personal identification. One could notice that these features are mainly global handwritten signature features. It does not mean that we disregard local features; we rather give the basic set of features that can be used to compute some others and can also be used on local level, e.g. we can determine all these features for each stroke.

3.1 Future work

Successful identification of the person is the main goal of our future work. Since we defined only basic feature set it is obvious that it would be expanded with some derived features, thus finding the ideal feature set. Besides that, we continue developing new identification methods that will combine dynamic and static features of a handwritten signature.

4 Conclusion

Handwritten signature identification is not as common as handwritten signature authentication. Signature has great entropy and it is often hard to distinguish if two signatures were made by the same person. Simply put, almost everything affects our signature.

In this paper we presented basic differences between off-line and on-line handwritten signature identification. It is clear that off-line methods are more often found in the literature. On-line handwritten signature identification is important if one wants to achieve more secure system. It captures live signature data from a sensor and compares it against all templates stored in the database to find the best match. We give a good theoretical base for the further development of an on-line handwritten signature identification system.

5 Acknowledgments

Shown results come out from the scientific project Methodology of biometrics characteristics evaluation (016-01611992-1721) and technological project Multiple biometric authentication using smart card (2008-043) financed by the Ministry of Science, Education and Sport, Republic of Croatia.

References

[1] Jain A.K., Ross A., Prabhakar, S.: An introduction to biometric recognition, IEEE Transactions on Circuits and Systems for Video Technology, 2004, pp. 4-20.

[2] Said H.E.S., Tan, T.N., Baker K.D.: Personal identification based on handwriting, Pattern Recognition 33, 2000, pp. 149-160.

[3] Zhu Y., Tan T.: Biometric personal identification based on handwriting, Pattern recognition, Proc 2., 2000, pp. 797-800.

[4] He Z., You X. Fang B., Tang Y.Y., Du J.: Handwriting-based Personal Identifcation, International Conference on Intelligent Computing, 23th - 26th August, HeFei, China, 2005, pp. 756-765.

[5] Shirdhonkar M.S., Kokare M.: Off-Line Handwritten Signature Identification Using Rotated Complex Wavelet Filters, International Journal of Computer Science Issues, 2011, pp. 478-482.

[6] Wirotius M., Seropian A., Vincent, N.: Writer Identification from Gray Level Distribution, Proceedings of the Seventh International Conference on Document Analysis and Recognition (ICDAR 2003), 3rd - 6th August, Edinburgh, Scotland, 2003.

[7] Zois E.N., Anastassopoulos V.: Morphological waveform coding for writer identification, Pattern Recognition 33, 2000, pp. 385-398.

[8] Pottier I., Burel G.: Identification and Authentification of Handwritten Signatures with a Connectionnist Approach, IEEE-ICNN, 26th June - 2nd July, Orlando, USA, 1994.

[9] Bhattacharyya D., Kim T.: Design of Artificial Neural Network for Handwritten Signature Recognition, International Journal of Computers and Communications, 2010, pp. 59-66.

[10] OZ C., Ercal F., Demir Z.: Signature Recognition and Verification with ANN,

available at `http://www.emo.org.tr/ekler/8b7dc6e8b36bcaa_ek.pdf`, Accessed: 27^{th} March 2011.

[11] Ashok J., Rajan E.G.: Writer Identification and Recognition Using Radial Basis Function, International Journal of Computer Science and Information Technologies, 2010, pp. 51-57.

[12] Sigari M.H., Pourshahabi M.R., Pourreza H.R.: Offline Handwritten Signature Identification using Grid Gabor Features and Support Vector Machine, 16th Iranian Conference on Electrical Engineering, 13th - 15th May, Tehran, Iran, 2008, pp. 281-286.

[13] Kisku D.R., Gupta P., Sing J.K.: Offline Signature Identification by Fusion of Multiple Classifiers using Statistical Learning Theory, International Journal of Security and Its Applications, 2010, pp. 35-45.

[14] Pourshahabi M.R., Sigari M.H., Pourreza H.R.: Offline Handwritten Signature Identification and Verification Using Contourlet Transform, 2009 International Conference of Soft Computing and Pattern Recognition, 4th - 7th December, Melacca, Malaysia, 2009, pp. 670-673.

[15] Seropian A., Grimaldi M., Vincent N.: Writer Identification based on the fractal construction of a reference base, Proceedings of the Seventh International Conference on Document Analysis and Recognition, 3rd - 6th August, Edinburgh, Scotland, 2003.

[16] Hu J., Lim S.G., Brown M.K.: Writer independent on-line handwriting recognition using an HMM approach, Pattern Recognition 33, 2000, pp. 133-147.

[17] Schlapbach A., Liwicki M., Bunke H.: A writer identification system for on-line whiteboard data, Pattern Recognition 41, 2008, pp. 2381-2397.

[18] Alahari K, Putrevu S.L., Jawahar C.V.: Discriminant Substrokes for Online Handwriting Recognition, Eighth International Conference on Document Analysis and Recognition (ICDAR 2005), 29th August - 1st September, Seoul, Korea, 2005, pp. 499-505.

[19] Bača M., Koruga P., Fotak T.: Basic on-line handwritten signature features for personal biometric authentication, 2011 Proceedings of the 34th International Convention MIPRO, 23rd - 27th May, Opatija, Croatia, 2011, pp. 116-121.