Evaluation of Local and Global Features for Offline Signature Verification

Muhammad Imran Malik*† , Marcus Liwicki*, Andreas Dengel*†
* *German Research Center for AI (DFKI GmbH)*
Knowledge Management Department,
Kaiserslautern, Germany
{firstname.lastname}@dfki.de
† *Knowledge-Based Systems Group,*
Department of Computer Science, University of Kaiserslautern,
P.O. Box 3049, 67653 Kaiserslautern, Germany

***Abstract*—In this paper we evaluate the impact of two state-of-the-art offline signature verification systems which are based on local and global features, respectively. It is important to take into account the real world needs of Forensic Handwriting Examiners (FHEs). In forensic scenarios, the FHEs have to make decisions not only about forged and genuine signatures but also about disguised signatures, i.e., signatures where the authentic author deliberately tries to hide his/her identity with the purpose of denial at a later stage. The disguised signatures play an important role in real forensic cases but are usually neglected in recent literaure. This is the novelty of our study and the topic of this paper, i.e., investigating the performance of automated systems on disguised signatures. Two robust offline signature verification systems are slightly improved and evaluated on publicly available data sets from previous signature verification competitions. The ICDAR 2009 offline signature verification competition dataset and the ICFHR 2010 4NSigComp signatures dataset. In our experiments we observed that global features are capable of providing good results if only a detection of genuine and forged signatures is needed. Local features, however, are much better suited to solve the forensic signature verification cases when disguised signatures are also involved. Noteworthy, the system based on local features could outperform all other participants at the ICFHR 4NSigComp 2010.**

***Keywords*-signature verification, mixture models, forgeries, disguised signatures, forensic handwriting analysis**

I. Introduction

Signature verification is in focus of research for decades. Traditionally, automated signature verification is divided into two broad categories, online and offline signature verification, depending on the mode of the handwritten input. If both the spatial as well as temporal information regarding signatures are available to the systems, verification is performed on online data. In the case where temporal information is not available and the systems must utilize only the spatial information gleaned through scanned or even camera captured documents, verification is performed on offline data [1], [2], [3].

The main motivation of this paper is to study the forensic relevance of signature features and their influence on verification. Until now online signature verification is not a common type of criminal casework for a forensic expert because the questioned signatures and the collected reference signatures (known) are commonly supplied offline [4]. Therefore, we focused explicitly on the offline signature verification.

In many recent works signature verification has been considered as a two-class pattern classification problem [1]. Here an automated system has to decide whether or not a given signature belongs to a referenced authentic author. If the system could not find enough evidence of a forgery from the questioned signature feature vector, it simply considers the signature as genuine belonging to the referenced authentic author, otherwise it declares the signature as forged. However, when talk about the forensic aspect, there is another equally important class of signatures that also needs to be identified, i.e., the disguised signatures.

A disguised signature is a signature that is originally written by the authentic reference author. However, it differs from the genuine signatures in the authors intent when it was written. A genuine signature is written by an author with the intention of being positively identified by some automated system or by an FHE. A disguised signature, on the other hand, is written by the genuine author with the intension of denial, that he/she has written that particular signature, later. The purpose of making such disguised signatures can be hundreds, e.g., a person trying to withdraw money from his/her own bank account via offline signatures on bank check and trying to deny the signatures after some time, or even making a false copy of his/her will etc. Potentially whatever the reason is, disguised signatures appear in real world and FHEs have to face them.

The category of disguised signatures has been addressed during the ICFHR 4NsigComp 2010 [5]. This was the first attempt to include disguised signatures into a signature verification competition. The systems had to decide whether the author wrote a signature in a natural way, with an intension of a disguise, or whether it has been forged by another writer.

In this paper we investigate two methods on two benchmark data sets. The first method is based on global features, i.e., a fixed number of features is extracted from signature images. In contrast, the second method uses a local approach, i.e., the number of features might vary - depending on the size of the signature. The two datasets are taken from previous signature verification competitions, i.e., the SigComp09 data set from the ICDAR 2009 [6] and the 4NSigComp10 data set from the ICFHR 2010 [5].

The rest of this paper is organized as follows. Section II summarizes the two datasets used for this study. Section III describes the two robust offline signature verification systems we applied. Section IV reports on the experimental results and provides a comparative analysis of the results. Section V concludes the paper and gives some ideas for our future work.

II. Data sets

A. ICDAR 2009 Signature Verification Competition

The first data set is the training set of the SigComp09 competition [6]. This dataset contains $1,898$ signature samples in all. There are 12 genuine authors – each one of whom wrote 5 of his/her genuine signatures, thereby yielding 60 genuine signatures. 31 forgers were had to forge the genuine signatures. Each forger contributed 5 forgeries for one writer resulting in 155 forged signatures per writer.[1]. Note that this dataset had no disguised signatures.

It is important to note that the said data were collected at a forensic institute where real forensic casework is performed. During dataset generation a special focus was given to the provision of more and more skilled forgeries since automated systems performance could vary significantly with how the forgeries were produced [4].

B. ICFHR 2010 Signature Verification Competition

These signatures were originally collected for evaluating the knowledge of FHEs under supervision of Bryan Found and Doug Rogers in the years 2002 and 2006, respectively. The images were scanned at 600dpi resolution and cropped at the Netherlands Forensic Institute.

The signature collection we used in our evaluation is the original test set of the ICFHR competition. It contains 125 signatures for one reference author. Out of this collection, 25 were the genuine signatures of reference author and remaining 100 were the questioned signatures. These 100 questioned signatures comprised 3 genuine signatures; 90 simulated signatures (written by 34 forgers freehand copying the signature characteristics of the referenced author after training); and 7 disguised signatures written by the reference author himself/herself with the intention of disguise. Note the huge difference between authentic data (3 genuine + 7 disguised signatures) vs. simulations (90 signatures). This did not affect our evaluation since we used the Equal Error Rate (EER) and relied on the Receiver Operating Characteristic curves (ROC-curves).

III. Automated signature Verification Systems

In this section we provide a short description of two state of the art offline signature verification systems we used in this study.

A. Local Features combined with GMM

This system was originally designed by the authors of this paper. A prior version of this system participated already in the ICDAR 2009 signature verification competition and achieved good results. It was not considered for participation during the 4NSigComp 2010 since the authors of this papers were among the organizers of this event. Our system uses Gaussian Mixture Models (GMMs) for the classification of the feature vector sequences. For the purpose of completeness, a short presentation of the system will be given here. For more details refer to [7].

Given a scanned image as an input, first of all binarization is performed. Second, the image is normalized with respect to skew, writing width and baseline location. Normalization of the baseline location means that the body of the text line (the part which is located between the upper and the lower baselines), the ascender part (located above the upper baseline), and the descender part (below the lower baseline) is vertically scaled to a predefined size each. Writing width normalization is performed by a horizontal scaling operation, and its purpose is to scale the characters so that they have a predefined average width.

To extract the feature vectors from the normalized images, a sliding window approach is used. The width of the window is generally one pixel and nine geometrical features are computed at each window position. Thus an input text line is converted into a sequence of feature vectors in a 9-dimensional feature space. The nine features correspond to the following geometric quantities. The first three features are concerned with the overall distribution of the pixels in the sliding window. These are the average gray value of the pixels in the window, the center of gravity, and the second order moment in vertical direction. In addition to these global features, six local features describing specific points in the sliding window are used. These include the locations of the uppermost and lowermost black pixel and their positions and gradients, determined by using the neighboring windows. Feature number seven is the black to white transitions present within the entire window. Feature number eight is the number of black-white transitions between the uppermost and the lowermost pixel in an image column. Finally, the proportion of black pixels to the number of pixels between uppermost and lowermost pixels is used. For a detailed description of the features see [8].

Gaussian Mixture Models [9] have been used to model the handwriting of each person. More specifically, the distribution of feature vectors extracted from a persons handwriting is modeled by a Gaussian mixture density. For a D-dimensional feature vector denoted as x, the mixture

[1] 22 of these forged signatures were not available so they have been ignored (this results in 1,838 forged signatures in all instead of 1860)

density for a given writer (with the corresponding model A) is defined as:

$$p(x\|A) = \sum_{i=1}^{m} w_i p_i(x)$$

In other words, the density is a weighted linear combination of M uni-modal Gaussian densities, $p_i(x)$, each parameterized by a $D \times 1$ mean vector, and D*D covariance matrix. For further details refer to [10].

B. Global Features combined with kNN

Our system is based on the methods introduced in [11]. However, we have modified/optimized it in order to fit in the scenarios presented in the datasets of the two mentioned signature verification competitions. A short summary of the system is given here, for further details consult [11].

First, the signature image is spatially smoothed followed by binarization. In the optimized version of this approach we used various combinations of local and global binarization techniques. After these preprocessing steps following operations were performed.

- Locating the signature image through its bounding box
- Centralizing the signature image to its center of gravity.
- Partitioning the image horizontally and vertically starting at center of gravity until it is divided into 64 cells.
- Finding the size of each cell of the image and normalizing it with the total number of black pixels it has. This constitutes the first feature vector.
- Calculating the angle that is made by the center point of each cell of the image with its lower right corner to obtain the second feature vector.
- Obtaining a third feature vector by calculating the angle of inclination of each black pixel in a cell to the lower right corner of its corresponding part of the image.

Note that the approach divides the signature into 64 small parts, which can be seen as a local feature extraction technique. However, since this division is based on a global analysis and the number of extracted features is fixed, disregarding the length of the signature, this approach is considered as a global approach. Therefore note that a simple disguise attempt would be to add a random character at the end of the signature and the global approach would fail while the local feature extraction would still find many similarities.

After computing these feature vectors, thresholds are computed using means and variances. Following that, nearest neighbor approach is applied to decide on the result of each feature vector and finally a voting based classification is made. In the optimized version different voting strategies have been applied that improved the overall performance.

IV. Evaluation

For reporting the results we primarily use the ROC-curves according to the evaluation procedure of the ICFHR 4NSigComp 2010. ROC-curves are a standard procedure of assessing the performance of signature verification systems. They are especially suited if there are unequal numbers of forged and genuine signatures in the dataset as in the case of both the ICDAR 2009 and ICFHR 2010 datasets. Results depict that, if only accuracy is used to evaluate signature verification systems, a system that votes by chance may show higher accuracy that in fact is false in context of a biometric system.

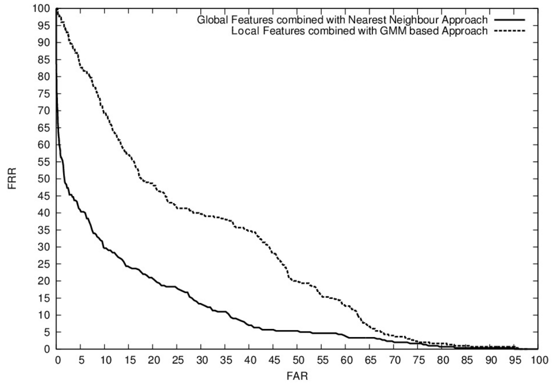

Figure 1: ROC on the ICDAR 2009 data

On the ICDAR 2009 dataset we performed 5-fold cross validation for each of the systems and generated ROC-curves. Furthermore, we evaluated both the systems on the ICFHR 2010 dataset again using the ROC-curves. The details of these evaluations are presented in the following sections.

A. Results on the ICDAR 2009 Dataset

We did 5-fold cross validation in the same way as in [6] and [7], i.e., for each genuine author we used only four of his/her genuine signatures to train and then tested the classifiers. The training set was rotated 5 times.

Figure 1 shows the results of both the systems on the ICDAR 2009 data set. It depicts the average results on all signatures by all writers. As shown in Fig. 1 the global features based system outperforms the local features based system. The Equal Error Rate (EER) for the global features based system is as low as 20 % whereas for the local features based system it is nearly 36 %. Note that the local features based system also participated in the ICDAR SigComp 2009. On the test data it provided an EER of 16 % [6] and was among the best classifiers. Since the test set is not publicly available, therefore we evaluated our system on the training data.

B. Results on the ICFHR 2010 Dataset

We evaluated both of the systems described in Section III according to the scenario posed by the ICFHR 4NSigComp 2010.There, the systems had to present their opinion by

Table I: Interpretation of the output

Decision	Probability		
Value (D)	$P > t$	$P < t$	$P = t$
1	authentic	misleading	inconcl.
2	disguise	simulation	inconcl.
3	inconcl.	inconcl.	inconcl.

Table II: Assessment of the output

True	Probability		
Answer	$P > t$	$P < t$	$P = t$
authentic	correct	incorr.	incorr./ignored
disguise	correct	incorr.	incorr./ignored
simulation	incorr.	correct	incorr./ignored

means of the following two output values for each of the questioned signatures.

- A Probability Value P between 0 and 1.
- A Decision Value D that could be either 1, 2 or 3.

The Probability Value P was compared to a predefined threshold t. A higher value ($P > t$) indicated that the questioned signature was most likely a genuine one. A lower value ($P \leq t$) indicated that the questioned signature was not genuine, meaning that it was not written by the reference author. A probability value of ($P = t$) was considered as inconclusive. The decision value D represents the system's decision about the process by which the questioned signature was most likely generated. A decision value of 1 means that the underlying writing is natural: there is no or not enough evidence of any simulation or disguise attempt and the signature is written by the reference author. The decision value 2 represents that the underlying writing process is unnatural: there is evidence of either a simulation or disguise attempt. Finally, a decision value 3 shows that the system is unable to decide if the underlying writing process is natural or unnatural: no decision could be made whether the signature is genuine, simulated, or disguised.

The output reference showing the various output possibilities is provided as Table I. Here a value of P greater than t with output 1 means correct genuine authorship, with output 2, on the other hand, means that the author has made an attempt to disguise her/his identity. If the Decision Value is 3, then with any value of probability it is simply inconclusive. Any value of P less than t with decision value 2 indicates that the questioned signature is a result of a simulation or disguise process. The final assessment of the output values is given in Table II.

As mentioned already, the novel feature of this dataset is the inclusion of disguised signatures. Various state-of-the-art systems participated in the competition and aimed at correctly classifying these disguised signatures. All of these systems failed to correctly detect the disguised signatures. The EER of the best system was larger than 50 %. More details of these results are provided in [5]. When these systems were evaluated without considering the disguised signatures the results of one participant were nearly perfect. In order to make our systems' performance comparable to those from the ICDAR competition, we present our results in the same manner, i.e., first without considering the disguised signatures and then taking the disguised signatures into account as well.

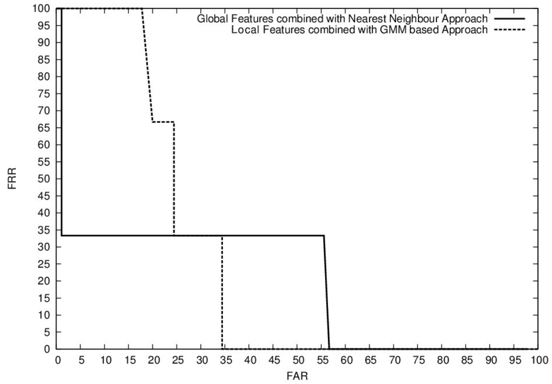

Figure 2: ICFHR 2010 results without disguised signatures

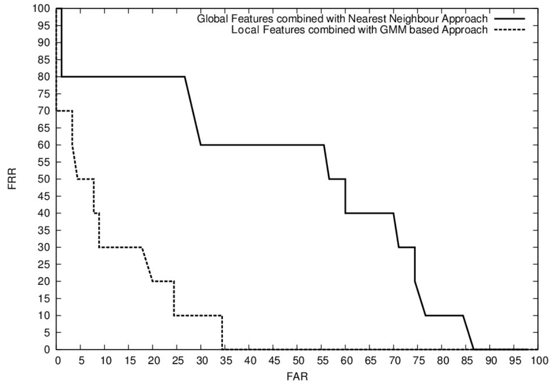

Figure 3: ICFHR 2010 results with disguised signatures

Figure 2 shows the results when we disregard the disguised signatures and consider only the case of forged vs. genuine signatures. The EER of both systems is the same. However, when considering the area under the curve, the local feature based system is slightly better.

The most important aspect of our study is the investigation of the influence of disguised signatures. The results are depicted in Figure 3. As shown, the local features based GMM system performs significantly better than the global features based system. It has an EER of 20% whereas the global feature based system has an EER of nearly 56%. Our point here is that, our GMM classifier performed well because it was relying exclusively on local features. To consolidate our thinking we also performed experimentation with the GMM classifier by feeding it with the global features (the same global features that are used by our NN Classifier). The results were worse in this case. The accuracy went below 50% and the EER was above 70%. Actually the nature of global features is to have a fixed amount of features while local features are not fixed. As such our GMM based system also outperforms all the participants of ICFHR 4NsigComp 2010 in this scenario as well. An important point to mention here is that our GMM based system was not even optimized to work with disguised signatures explicitly. In contrast, it was initially developed as a general-purpose offline writer identification system. We strongly believe that this better performance of our system is attributed to the fact that it relies on the local features.

V. Conclusion and Future Work

In this paper we have reported on the experiments conducted to evaluate the impact of local and global features on automated signature verification for off-line signatures collected by the FHEs. Two state of the art offline signature verification systems were applied on the datasets of the last two signature verification competitions.

Our experimental results show that the global features could produce acceptable results when the traditional paradigm of forged vs. genuine authorship is under consideration. The actual power of local features is revealed when considering the more realistic scenario which involves the presence of disguised signatures among the questioned signatures. This has been shown by using the equal error rates achieved by a GMM based offline signature verification system that heavily relies on the local features of offline signature samples. We strongly believe that the main reason for the good performance of this system is due to the difference that this system is relying on local features.

In future we plan to investigate more local features approaches for signature verification. Using novel image analysis methods like scale-invariant Speeded Up Robust Features (SURF) [12] might be an interesting idea as well. We also plan to combine various offline signature verification systems based on different global and local features through voting strategies to produce even better results.

Furthermore, we plan to perform analyses on data which contains signatures from more reference writers and skilled forgers. Regarding genuine signatures, large and diverse test sets where signatures are produced by different authors under various different psychological and physical conditions may also yield interesting results.

Acknowledgment

The work was supported by the ADIWA project.

References

[1] R. Plamondon and G. Lorette, “Automatic signature verification and writer identification – the state of the art,” *Pattern Recognition*, vol. 22, pp. 107–131, 1989.

[2] R. Plamondon and S. N. Srihari, “On-line and off-line handwriting recognition: A comprehensive survey,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 63–84, 2000.

[3] D. Impedovo and G. Pirlo, “Automatic signature verification: The state of the art,” *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 38, no. 5, pp. 609–635, Sep. 2008.

[4] V. L. Blankers, C. E. v. d. Heuvel, K. Y. Franke, and L. G. Vuurpijl. (2009) Call for participation:signature verification competition, on- and offline skilled forgeries. [Online]. Available: http://sigcomp09.arsforensica.org/

[5] M. Liwicki, C. E. van den Heuvel, B. Found, and M. I. Malik, “Forensic signature verification competition 4NSigComp2010 - detection of simulated and disguised signatures,” in *12th International Conference on Frontiers in Handwriting Recognition*, 2010, pp. 715–720.

[6] V. L. Blankers, C. E. v. d. Heuvel, K. Y. Franke, and L. G. Vuurpijl, “Icdar 2009 signature verification competition,” in *Proceedings of the 2009 10th International Conference on Document Analysis and Recognition*, ser. ICDAR ’09. Washington, DC, USA: IEEE Computer Society, 2009, pp. 1403–1407. [Online]. Available: http://dx.doi.org/10.1109/ICDAR.2009.216

[7] M. Liwicki, “Evaluation of novel features and different models for online signature verification in a real-world scenario,” in *Proc. 14th Conf. of the Int. Graphonomics Society*, 2009, pp. 22–25.

[8] U.-V. Marti and H. Bunke, *Using a statistical language model to improve the performance of an HMM-based cursive handwriting recognition systems*. River Edge, NJ, USA: World Scientific Publishing Co., Inc., 2002, pp. 65–90. [Online]. Available: http://portal.acm.org/citation.cfm?id=505741.505745

[9] J. Marithoz and S. Bengio, “A comparative study of adaptation methods for speaker verification,” 2002.

[10] A. Schlapbach, M. Liwicki, and H. Bunke, “A writer identification system for on-line whiteboard data,” *Pattern Recogn.*, vol. 41, pp. 2381–2397, July 2008.

[11] P. I. S. Dr. Daramola Samuel, “Novel feature extraction technique for off-line signature verification system,” *International Journal of Engineering Science and Technology*, vol. 2, pp. 3137–3143, 2010.

[12] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, “Speeded-up robust features (surf),” *Comput. Vis. Image Underst.*, vol. 110, pp. 346–359, June 2008. [Online]. Available: http://portal.acm.org/citation.cfm?id=1370312.1370556