Abstract

We present an offline signature verification system based on a signature’s local histogram features. The signature is divided into zones using both the Cartesian and polar coordinate systems and two different histogram features are calculated for each zone: histogram of oriented gradients (HOG) and histogram of local binary patterns (LBP).

The classification is performed using Support Vector Machines (SVMs), where two different approaches for training are investigated, namely global and user-dependent SVMs. User-dependent SVMs, trained separately for each user, learn to differentiate a user’s signature from others, whereas a single global SVM trained with difference vectors of query and reference signatures’ features of all users, learns how to weight dissimilarities. The global SVM classifier is trained using genuine and forgery signatures of subjects that are excluded from the test set, while user-dependent SVMs are separately trained for each subject using genuine and random forgeries.

The fusion of all classifiers (global and user-dependent classifiers trained with each feature type), achieves a 15.41% equal error rate in skilled forgery test, in the GPDS-160 signature database without using any skilled forgeries in training.

1. Introduction

Signature is a behavioral biometric that encodes the ballistic movements of the signer for his/her chosen signature. Compared to physical traits such as fingerprint, iris or face, a signature typically shows higher intra-class and time variability. Furthermore, as with passwords, a user may choose a simple signature that is easy to forge. On the other hand, the signature’s widespread acceptance by the public and niche applications (validating paper documents and use in banking applications) make it an interesting biometric.

Depending on the signature acquisition method used, automatic signature verification systems can be classified into two groups: online (dynamic) and offline (static). A static signature image, generally scanned at a high resolution, is the only input to offline systems. Verification of signatures found on bank cheques and vouchers are among important applications for offline systems.

In addition to the signature image, time dimension is also available for dynamically captured signatures that are acquired using pressure sensitive tablets or smart pens. These input devices sample the signature at a high frequency, resulting in a time ordered sequence of signature’s trajectory points. Each point is associated with a corresponding acquisition time stamp and a location coordinate, besides other dynamic features such as pressure and pen inclination angles that can be captured subject to the hardware used. Online signature verification is generally used for access control and electronic document authentication types of application. Due to the differences in the input, preprocessing, feature extraction and classification methods used in online and offline systems show significant variations.

Offline signature verification can be said to be more challenging compared to online signature verification. While variations among a user’s signatures and easy to forge signatures pose a challenge in both cases, dynamic information available in online signatures make the signature more unique and more difficult to forge. In particular, imitating both the shape and dynamic information of an online signature seems to be difficult except for very simple signatures. On the other hand, it is possible in some real life scenarios, for an impostor to trace over a genuine offline signature and obtain a high quality forgery. The availability of the signature’s trajectory also makes it easier for online verification systems to align two signatures and detect differences.

Two general approaches may be considered for the
signature verification problem, though preferred methods vary for online versus offline systems: User-based modeling/discrimination requires one model per user, generally necessitating a large number of references (typically 10+) for which classifiers such as Hidden Markov Models, or Support Vector Machines are often used. In template-based approach, 1 to 5 references of the claimed identity are enough to be used as template. The query signature is accepted or rejected by comparing its distance to the template of the claimed identity, to a global or user-based threshold. Many possible features and matching methods are possible based on the task: Dynamic Time Warping (DTW) is successfully used in online signature verification [9] where signature trajectory facilitates the registration of signatures. In offline signature verification, local features that are more resilient to variations are more commonly used with various types of classifiers, after rigid or elastic registration of two signatures, as summarized in 2.

The system performance is generally reported using the False Rejection Rate (FRR) of genuine signatures and the False Acceptance rate (FAR) of forgery signatures. Other measures such as the Equal Error Rate (EER), the error rate where both FAR and FRR are equal, as well as the ROC curve and false reject rates at fixed false accept rates are also commonly reported. It is also possible to use the Distinguishing Error Rate (DER) which is the average of FAR and FRR. One of the main challenges in assessing the system performance is the skill levels of the collected forgery signatures. A number of forgery types have been defined: a skilled forgery is signed by a person who has had access to a genuine signature for practice, and a random or zero-effort forgery is signed without having any information about the signature, even the name of the person whose signature is forged.

In this paper, we present an offline signature verification system based on local features, intended to be robust to global shape variations that are commonly induced by embellishing strokes that are produced by fast ballistic movements and are floating inside the signature’s overall pattern. The local features we experimented with are based on gradient information (histogram of oriented gradients) and pixel neighborhood patterns (local binary patterns) inside the local zones. We analyzed the discriminative power of the extracted features using support vector machine classifiers and obtained results comparable to the state-of-the-art using classifier combination.

The rest of the paper is organized as follows: in Section 2 the state of the art offline signature verification methods are reviewed and analyzed. In Section 3 proposed method is described, following with Section 4 where performance results are presented. Finally, conclusions and future work are reported in Section 5.

2. Previous work

Offline signature verification is a well-researched topic, where many different features and classifiers have been studied. For instance, in one of the earlier works, local shape descriptors are used as features to train a k-nearest neighbor (KNN) classifier [20]. Later, local correspondence between a model and a query signature is used to compare a set of geometric properties in [7], and Radon transform is used to extract features to feed to a Hidden Markov Model (HMM) in [3]. A series of surveys covering advances in the field are available in [15, 18, 10, 16]. A more up to date overview of proposed methods is detailed in [2]. Here, we review some of the recent research on offline signature verification.

Two approaches are used in [23] to exploit information related to stable parts of signatures (the parts that do not show much variation across the signatures of a user). The first approach is to train a neural network classifier with artificial forgeries generated by removing stable components from genuine signatures, so that the classifier detects changes in these stable components when verifying signatures. The other is to force the neural network classifier to pay special attention to local stable parts of signatures by weighting their corresponding node responses through a feedback mechanism.

A comparison of support vector machine (SVM) and HMM classifiers in the context of the off-line signature verification is reported in [8], where a private database of 100 subjects is utilized to compare the classifiers. Both of the classifiers are trained using signatures of the first 40 subjects, and tested using signatures of the remaining individuals. According to the reported results, SVM was found to be superior to the HMM classifier. However, they just used simple features such as pixel density or gravity center distance extracted from grids.

In [5], interior stroke distributions in polar and Cartesian coordinates are used as features. Three types of classifiers are used to test the performance of the features: HMM, SVM and simply the Euclidean distance of the extracted features. The GPDS-160 database is used to evaluate the method [6]. To find user based thresholds, 3 forgery signatures from each subject are used. This may not be a realistic scenario since it requires knowledge of existing forgeries for each user. Authors report performance results based on DER. When 12 genuine signatures are used as reference, remaining 12 genuine and 30 forgery signatures are used for testing each person; HMM gives 13.35% DER, SVM with radial basis function (RBF) kernel gives 14.27% DER and Euclidean distance metric gives 15.94% DER.

Enhanced modified direction features (MDF) are used to train artificial neural network (ANN) and SVM classifiers in [12]. Location and direction of transitions from background to foreground pixels are used as features. The GPDS-160...
database is utilized to find the performance of the method. For the training part for each writer, 12 genuine signatures are used as positive examples whereas 100 writers are randomly selected to provide 400 random forgeries as negative examples. For testing, they use a mix of random and skilled forgeries where the remaining 12 genuine signatures are used together with 59 random forgeries from the remaining 59 writers and 15 targeted forgery signatures of that specific writer. The SVM is reported to give the best result which is 20.07% DER.

In [14], vertical projection features are used as features fed into a Dynamic time warping (DTW) algorithm with some modifications to incorporate a stability factor to increase the performance of the DTW algorithm. The system gives a DER of 22.5% on skilled forgery test using a private database.

High pressure points are matched in polar coordinates using Probabilistic neural networks (PNN) and KNN as classifiers, in [21]. To evaluate the performance, GPDS-160 database is used. To test the performance of the proposed method, genuine and forgery signatures of each subject is divided into two equal parts; making 12 genuine and 15 skilled forgery training signatures and the same amount of test signatures. Best KNN result is 12.62% DER and best PNN result is 12.33% DER.

Local interest points, which correspond to local maxima in a scale-space representation of a signature, are detected in [19]. The descriptors that characterize local neighborhood around corresponding interest points, are calculated using the scale invariant feature transform (SIFT). The correspondence between descriptors of reference and query signatures is established using wide baseline methodology, while the final decision is performed using a Bayes classifier. The system performance is assessed using the GPDS-160 signature dataset, where 15.3% DER is reported. However, they do not perform a full skilled forgery test, they just use a small subset of all skilled forgeries for testing.

Current state of the art varies between 9.02% and 17.25% EERs for different variations of GPDS database, considering the works where no skilled forgery of a user is utilized in training phase, according to [17, 11, 22]. In Table 2, we give the summary results for the systems utilizing GPDS-160 dataset. Performance results are summarized in the form of average error rate (DER) to be compatible with the previous results.

### 3. Proposed method

#### 3.1. Preprocessing

Offline signature verification may benefit from normalization steps to obtain global rotation, scale and translation invariance, since signing conditions may significantly change size, orientation and location of the signature in a document. In the present system, we don’t do any preprocessing since the features used are inherently invariant to a translation and scale, while rotation normalization is not very straightforward due to the difficulties associated in assessing the required normalization parameter (e.g. rotation angle). Our preliminary studies have shown that comparing two signatures in a few different possible rotation angles (\{-10, 0, +10\} degrees) improves the overall accuracy, but since it considerably slows down the training and verification processes, it is not used in the current system.

#### 3.2. Grids in Cartesian and Polar Coordinates

In order to develop a system robust to global shape variations, we extract features from local zones of the signature image. For this, the image is either divided into zones using a fixed number of rectangular grids in Cartesian coordinates or using a circular tessellation around the origin point in polar coordinates. A sample signature, overlaid with a 10x20 rectangular grid is shown in Figure 1 where the signature is padded with space (shown in gray) to make it fit the grid. An example of the polar grid where the origin is selected as the centroid of the signature is shown in Figure 2. The size of rectangular grid is selected as 10x20, whereas the number of angular bins used in the polar system is selected as 12, and that of distance bins is selected as 15. Grid sizes are found experimentally on a separate validation set, though in future it might be better to use bigger zones for space and time efficiency.

![Figure 1. Sample signature with overlaid with the rectangular grid.](image1)

![Figure 2. Signature tessellation using the polar grid, with origin at the centroid.](image2)
Since signatures do not contain a reference point such as those that can be found in fingerprints or faces, the centroid or center of mass can be used as a lesser alternative in registering two signatures. Unfortunately, the location of both of these points may show large variations due especially to embellishing strokes that often show large variations. If the registration point is selected as the top-left point of the bounding box and the embellishing strokes are on the right, than the left parts of the two signatures align better than the right. With this observation and at the cost of having some redundant features, we decided to use multiple registration points (centroid, top-left, top-right etc.) in the polar grid, to reduce the effect of registration mismatches. The tessellation obtained by using the top-left corner of the signature bounding box is shown in Figure 3.

Note that the zones away from the center point are bigger than those near the center; thus the features computed from them are less specific. This also helps to some extent with the registration problem since coarser features are used where the variation is large and further justifies the use of the multiple tessellations. For this reason, we will switch to the log-polar grid in the future.

3.3. Feature extraction

We have experimented with two separate features: histogram of oriented gradients (HOG, [4]) relative to the dominant orientation and local binary patterns (LBP, [13]).

3.3.1 Histogram of Oriented Gradients

HOG features are proposed by Dalal and Triggs [4]. They involve first computing the gradient information at each pixel inside a particular grid zone (either Cartesian or Polar). Next, histogram of gradient orientations in that zone is computed.

While computing the gradient orientation histogram, we apply a normalization to allow for rotational differences of the strokes within the grid zone. Specifically, after finding the gradient orientation at each point, we find the dominant gradient orientation and represent it at the first bin of the histogram. Without this normalization, a rotation of the strokes in a zone would correspond to a circular shift in the HOG histogram; lowering the match between the original and matched histograms.

3.3.2 LBP features

Local binary pattern (LBP) is a powerful feature proposed to capture the texture in objects [13]. In the basic LBP method, a gray scale image is processed such that a binary code is generated for each pixel in the image. This code encodes whether the intensities of the neighboring pixels are greater or less than the current pixel’s intensity. So, for instance in a 3x3 neighborhood with the current pixel being the center, a binary code of length 8 is generated consisting of 0s and 1s, according to the relative intensities of the neighbors. A histogram is then computed to count the number of occurrences of each binary code, describing the proportion of common textural patterns.

The LBP method is commonly used in object recognition with good success and we expected it to also be useful in offline signature verification. Indeed, it was used previously in signature verification [22] as well. Furthermore, since LBP is a texture feature, we expected it to be complementary to the HOG features that are also used in this work.

3.4. Classification

The classification is performed using Support Vector Machine classifiers [1], where two different approaches to train the classifier are investigated, namely global and user-dependent SVMs.

The number of genuine signatures used as reference is kept variable (5 or 12). Both classifiers are trained with RBF kernels and parameters are optimized with grid search on a separate validation set (users 161-300 from the GPDS-300 dataset, who are not in the test set).

3.4.1 Global SVMs (GSVM)

In the first approach, we train a global SVM which is a user-independent classifier trained to learn to separate difference vectors obtained from genuine signatures of a user, from those obtained from (skilled) forgery signatures of the same user.

To obtain the difference vectors, features obtained from a query signature (genuine or forgery) are compared to the features obtained from each of the reference signatures of the claimed identity. The resulting difference vectors are then normalized so that each element of this vector represents how many standard deviation away the query feature is from the reference feature.

More precisely, let $\{R^1, R^2, \ldots, R^N\}$ be the feature vectors extracted from the reference signatures of a particular user and let $Q = [q_1 \ldots q_M]$ be the feature vector extracted...
from a test signature, where \( N \) is the number of reference signatures and \( M \) is the number of features. Then, we compute \( N \) difference vector for each query, where the \( i^{th} \) difference vector is computed as:

\[
D^i = Q - R^i = \begin{bmatrix}
q_1 - R^i_1/\left(\sigma_1 + \tau\right)
q_2 - R^i_2/\left(\sigma_2 + \tau\right)
\vdots
q_M - R^i_M/\left(\sigma_M + \tau\right)
\end{bmatrix}
\]

where \( \sigma_i \) is the standard deviation of the features among the \( i^{th} \) feature of the claimed user’s reference signatures; and \( \tau \) is a small constant to eliminate division by zero.

We devote some of the users who are not in the test set (users 161-300 from the GPDS-300 dataset), and use all of their signatures (genuine and skilled forgery) to train the system. It is important to underline that, in this way, no skilled forgeries belonging to users in the test set, are used during training.

Note here that the SVM is learning which changes in the feature vector may be within the normal variations of a signer and which changes indicate forgeries. This can be better explained considering the case of a system using global features where the SVM learns how much variation in a particular feature (e.g. size, pixel density, width-to-height ratio) matters. In the case of local features, the SVM can learn how to weight differences in the center versus periphery of the signature for instance.

### 3.4.2 User-dependent SVMs (USVM)

In the second approach, we train user-dependent SVMs, one for each user, with the expectation that the user-dependent SVM can learn to differentiate genuine signatures of a person from forgeries.

For this, each SVM is trained with the raw feature vectors obtained from the reference signatures of the corresponding user and those obtained by random forgeries – other users’ reference signatures reserved for training. Note that in this case, we do not need a separate group of users for training as opposed to GSVM, since we only use genuine signatures of others.

### 4. Experimental results

#### 4.1. Dataset

A publicly available subset of the GPDS-960 dataset [6], namely GPDS-300, was used in training and testing the system, such that training and testing data were completely separate. Specifically, the first 160 users of the GPDS-300 dataset (GPDS-160) was used in testing, while users 161 to 300 from the GPDS-300 dataset were used in training.

The amount of data which is publicly available has recently risen from 160 to 300 individuals; hence, most of the work using the GPDS database reports on the GPDS-160 subset. In order to make our results comparable to those reported in the literature, we decided to also use the GPDS-160 dataset to test our system.

The GPDS-960 dataset contains signatures provided by 960 individuals, where each individual provided 24 genuine signature samples. Genuine signatures were collected in a single session, where each subject was asked to sign his/her signature into a form with a preprinted grid containing two types of cells 5x3.5cm and 5.5x2.5cm, respectively. Additionally, a total of 30 practiced forgery signatures, provided by 10 forgers, were collected for each individual. Prior to collecting forgery signatures of a corresponding individual, a number of high resolution signature images were made available to forgers for practice. Likewise, forgers submitted corresponding signatures using forms with the similar grid size. Finally, both reference and forgery signatures were scanned at 300dpi resolution and preprocessed to a black and white format. Figure 4 depicts sample genuine (odd columns) and their corresponding forgery (even columns) signatures from the dataset.

Figure 4. Sample genuine (odd columns) and their corresponding forgery (even columns) signatures from GPDS-160 database.

#### 4.2. Test Protocol

In order to obtain results that are comparable to those reported in the literature, we trained classifiers using 12 reference signatures. However this many reference signatures are not common in real life applications. So, in the second part of our tests, we used only 5 references to obtain results that better reflect applications where users are willing to provide only a few reference signature for enrollment.

In skilled forgery tests, we used all genuine signatures of a user except those that are used as reference; thus resulting in 12 and 19 genuine tests per user, for the cases of 12 and 5 reference signatures, respectively. Since we do not use any skilled forgeries of test users in training, all skilled forgeries of a user (30) are used in testing.
4.3. Results

The results of our experiments using two different feature types, two different grids and two different classifier training approaches described earlier, are given in Table 1. The entries marked with a star were runs that could not be fully completed due to time or space requirements.

Analysis of these results shows that the USVM significantly outperforms GSVM; this is not very surprising as the USVMs are specifically trained for each user, while GSVMs only now about variations in each dimension. On the other hand, the global SVM improves the performance when used in conjunction with user SVMs.

Classifier combination was applied at score level to combine the decisions of the six classifiers. As found in many studies in different fields, we also found that classifier combination using a weighted sum rule improves overall accuracy (15.41% EER using 12 references). The weights are found in a separate validation set using grid search. For the 5-reference case, they are \{0.10,0.15,0.25,0.20,0.15,0.15\} for classifiers shown in Table 1, from top-to-bottom.

In combining classifiers, we handled the two partially finished tests as follows: we took the results of the corresponding tests with 5 references and removed the references 6-12 from the genuine tests. In other words, we could not use those as references, but we did not use them as genuine either. Note that with the GSVM, using more references corresponds to having more training data (more difference vectors) to learn the boundary between forgery and genuine classes. Hence, we would expect similar or better results, had we fully completed these two tests.

As for the features, HOG features obtained using Polar coordinates (HOG-Polar) outperform all other types of features with 19.58% EER using USVMs. The LBP-Grid feature closely follows, with a 19.84% EER.

Finally, we observe that using a greater number of reference signatures significantly improves performance, as expected and observed in previous work also.

Table 1. Summary of the EER performance results of skilled forgery tests.

<table>
<thead>
<tr>
<th>Features</th>
<th>Classification</th>
<th>12 ref.</th>
<th>5 ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>HOG-Polar</td>
<td>USVM</td>
<td>19.58%</td>
<td>21.73%</td>
</tr>
<tr>
<td>HOG-Grid</td>
<td>USVM</td>
<td>21.13%</td>
<td>22.65%</td>
</tr>
<tr>
<td>LBP-Grid</td>
<td>USVM</td>
<td>19.84%</td>
<td>22.90%</td>
</tr>
<tr>
<td>HOG-Polar</td>
<td>GSVM</td>
<td>23.57%*</td>
<td>22.79%</td>
</tr>
<tr>
<td>HOG-Grid</td>
<td>GSVM</td>
<td>24.13%</td>
<td>29.61%</td>
</tr>
<tr>
<td>LBP-Grid</td>
<td>GSVM</td>
<td>35.82%*</td>
<td>34.11%</td>
</tr>
<tr>
<td>All</td>
<td>Combined</td>
<td>15.41%</td>
<td>17.65%</td>
</tr>
</tbody>
</table>

For comparison, we give state-of-the-art results on the GPDS database in Table 2; due to space shortage, we could not include results of other work using a smaller portion of the database. Compared to the results given in this table, our classifier combination result when using 12 reference signatures (15.41% EER, equal to 15.35% DER for our system) is better than the systems that do not use any skilled forgeries of a tested user during training ([12, 11]), as is also the case for our system. Furthermore, our results are only slightly lower compared to those that do use skilled forgeries in testing ([21, 5]).

5. Conclusion

We presented an automatic offline signature verification system based on signature’s local histogram representations. The signature is divided into zones using both fixed size rectangular or polar grids, where HOG and LBP features are calculated. For either of the representations, features obtained from grid zones are concatenated to form the final feature vector. Two different types of SVM classifiers are trained, namely global and user dependent SVMs, to perform verification. We also experimented with the fusion of classifiers, and showed that their combination improves overall verification performance. Feature-level fusion is also possible but we preferred training classifiers to be experts for each feature type.

The system performance is measured using the skilled forgery tests of the GPDS-160 signature dataset. Additionally, a classifier fusion is performed, where global and user dependent SVM classifiers are combined giving the best result of 15.41% and 17.65% equal error rate on skilled forgery test with 12 and 5 references, respectively.

In summary, obtained results are comparable or better compared to those reported in the literature for the GPDS database, as can be seen in the Table 2. Considering that using skilled forgeries brings a potentially significant advantage in accuracy, the results should be deemed comparable and possibly better than state-of-the-art results. On the other hand, the fact that the proposed system does not require skilled forgeries of the enrolling user, is attractive for real life applications.

In future work, we will incorporate some ideas such as improving alignment of two signatures; using the log-polar coordinates for better use of the grids; considering different scales and orientations of the query; and adding complementary features such as gradient magnitudes in addition to gradient directions. We will also train our classifiers with skilled forgeries in order to compare results to those in literature.

References

Table 2. Summary of recent results (Distinguishing Error Rates) on GPDS dataset.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Method</th>
<th>GPDS Set</th>
<th>Training</th>
<th>Testing</th>
<th>DER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vargas et al. [21]</td>
<td>PNN</td>
<td>GPDS-160</td>
<td>12 genuine + 12 skilled forg.</td>
<td>12 gen. + 12 skl. forg.</td>
<td>12.33%</td>
</tr>
<tr>
<td>Ferrer et al. [5]</td>
<td>HMM</td>
<td>GPDS-160</td>
<td>12 genuine + 3 skilled forg.</td>
<td>12 gen. + 27 skl. forg.</td>
<td>13.35%</td>
</tr>
<tr>
<td>Nguyen et al. [12]</td>
<td>SVM</td>
<td>GPDS-160</td>
<td>12 genuine + random forg.</td>
<td>12 gen. + 30 skl. forg.</td>
<td>20.07%</td>
</tr>
<tr>
<td>Proposed</td>
<td>SVM</td>
<td>GPDS-160</td>
<td>12 genuine + random forg.</td>
<td>12 gen. + 30 skl. forg.</td>
<td>15.35%</td>
</tr>
<tr>
<td>Proposed</td>
<td>SVM</td>
<td>GPDS-160</td>
<td>5 genuine + random forg.</td>
<td>19 gen. + 30 skl. forg.</td>
<td>17.63%</td>
</tr>
</tbody>
</table>


