Information Fusion 32 (2016) 109–119

Contents lists available at ScienceDirect

Information Fusion

journal homepage: www.elsevier.com/locate/inffus

Score level fusion of classifiers in off-line signature verification

Mustafa Berkay Yılmaz*, Berrin Yanıkoğlu

Faculty of Engineering and Natural Sciences, Sabancı University, İstanbul, Turkey

ARTICLE INFO

Article history:
Available online 12 February 2016

Keywords:
Offline signature
Score-level fusion
Histogram of oriented gradients
Local binary patterns
Scale invariant feature transform

ABSTRACT

Offline signature verification is a task that benefits from matching both the global shape and local details; as such, it is particularly suitable to a fusion approach. We present a system that uses a score-level fusion of complementary classifiers that use different local features (histogram of oriented gradients, local binary patterns and scale invariant feature transform descriptors), where each classifier uses a feature-level fusion to represent local features at coarse-to-fine levels. For classifiers, two different approaches are investigated, namely global and user-dependent classifiers. User-dependent classifiers are trained separately for each user, to learn to differentiate that user's genuine signatures from other signatures; while a single global classifier is trained with difference vectors of query and reference signatures of all users in the training set, to learn the importance of different types of dissimilarities.

The fusion of all classifiers achieves a state-of-the-art performance with 6.97% equal error rate in skilled forgery tests using the public GPDS-160 signature database. The proposed system does not require skilled forgeries of the enrolling user, which is essential for real life applications.

1. Introduction

Signature verification is used in verifying the claimed identity of a person through his/her chosen and previously registered signature. The signature's widespread acceptance by the public and niche applications (validating paper documents and use in banking applications) makes it a desirable biometric.

Signature is considered to be a behavioral biometric that encodes the ballistic movements of the signer and as such is difficult to imitate. On the other hand, 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.

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 is the only input to offline systems, while signature trajectory as a function of time is also available in online signatures. Main difficulties in both tasks are simple (easy to forge) signatures and variations among a user's signatures, but the dynamic information available in online signatures make the signature more unique and more difficult to forge.

Research databases define two types of forgeries: a *skilled forgery* refers to a forgery which is signed by a person who has had access to some number of genuine signatures and practiced them for some time. In contrast, a *random forgery* is typically collected from other people's real signatures, simulating the case where the impostor does not even know the name or shape of the target signature and hence uses his/her own for forgery. Random forgery detection is a much easier task compared to skilled forgery detection. In this work, as in the literature, when the term "forgery" is used without further qualifications, it may refer to either a skilled forgery or random forgery.

Systems' evaluation is often done in terms of the *Equal Error Rate (EER)* which is the point where the False Accept Rate (FAR) and False Reject Rate (FRR) are equal and occasionally in terms of the Distinguishing Error Rate (DER), which is the average of FAR and FRR.

While the use of the public signature databases has become the norm in the last years, the databases do not always have strictly specified protocols. As a result, many reported accuracies cannot be directly compared with if they use a different (sometimes random) subset of the users; or a different number of reference signatures (using more helps the system as it provides more information); or a different number of skilled forgeries.

In this work, we present a state-of-the-art offline signature verification system that uses a fusion of complementary features, classifiers and preprocessing techniques, with the aim to explore the limits in signature verification accuracy.

* Corresponding author. Tel.: +90 216 4839528; fax:+90 216 4839550.
E-mail addresses: berkayyilmaz@sabanciuniv.edu (M.B. Yılmaz), berrin@sabanciuniv.edu (B. Yanıkoğlu).

http://dx.doi.org/10.1016/j.inffus.2016.02.003
1566-2535/

Our main contribution is the comprehensive study and treatment of different aspects of offline signature verification, which are fused at the end to form a state-of-the-art verification system, with novel aspects including the following:

- We propose an alignment algorithm that improves overall accuracy by more than 2% on average. While alignment of test images degrades overall performance, we have found that automatic alignment of references is when used with the global classifiers.
- We improve on the use of the well-known features and approaches by novel adaptations. (i) We use coarse-to-fine grids for capturing a spectrum of global to local features when using the histogram of oriented gradients (HOG) and local binary patterns (LBP). (ii) We select the best LBP templates according to term frequencies and combine similar LBP template histogram bins to obtain a dense histogram. (iii) We use a novel scale invariant feature transform (SIFT) descriptor matching algorithm that seeks more than one global transformation in order to allow different transformations in different parts of a signature.
- We incorporate user-dependent and user-independent verification concurrently. We apply a score level fusion to combine classifiers with complementary feature types, where the weights are learnt from a separate validation set.

2. Literature review

Offline signature verification is a well-researched topic where many different approaches have been studied. A series of surveys covering advances in the field are available [1–10]. Here, we review some of the recent works, grouped according to focus areas.

Note that while we give some performance figures for completeness, many of the reported numbers are not directly comparable as they are obtained under different conditions (number of reference signatures, use of skilled signatures etc.). We discuss this issue in Section 6.6.

Feature extraction

Several different features are used in offline signature verification, especially local features such as SIFT descriptors, wavelet features and LBP, among others. Solar et al. use SIFT descriptors in conjunction with the Bayes classifier [11]. The performance is assessed using the GPDS-160 signature dataset, with a 15.3% DER. However, only a small subset of all skilled forgeries, and not the full test set, is used for testing.

Vargas et al. use complex features based on LBP to perform statistical texture analysis [12]. To extract second order statistical texture features from the image, another feature called the gray level co-occurrence matrix method is utilized. The best combination of features is reported to achieve an EER of 9.02% on the gray-level GPDS-100 database, using 10 reference signatures.

Different base classifiers

Ferrer et al. [13] have evaluated the effectiveness of hidden Markov models (HMMs), support vector machines (SVMs) and the Euclidean distance classifier on the publicly available GPDS-160 database. When 12 genuine signatures and 3 skilled forgeries are used in training the classifiers, the DER rates are found as 13.35%, 14.27% and 15.94% for the HMMs, SVM (radial basis function kernel) and the Euclidean distance classifier, respectively.

A comparison of probabilistic neural networks (PNN) and K-nearest neighbor (KNN) is done by Vargas et al. [14]. Genuine and skilled forgery signatures of each subject are divided into two equal parts, resulting in 12 genuine and 12 skilled forgeries in train set and the same amount in the test set. The results on the gray-level GPDS-160 database are found to be close: the best results are found to be 12.62% DER with the KNN ($k = 3$) and 12.33% DER with the PNN.

Use of classifier combination

There are quite a lot of studies on the effect of classifier combination in offline signature verification. In one of the earlier works, Fierrez-Aguilar et al. consider the sum rule for combining global and local image features [15]. One of the experts in this work is based on a global image analysis and a statistical distance measure, while the second one is based on local image analysis with HMMs. It is shown that local information outperforms the global analysis in all reported cases. The two proposed systems are also shown to give complementary recognition information, which is desired in fusion schemes.

Receiver operating characteristic (ROC) curves are used for classifier combination by Oliveira et al. [16]. Different fusion strategies to combine the partial decisions yielded by SVM classifiers are analyzed and the ROC curves produced by different classifiers are combined using the maximum likelihood analysis. Authors demonstrate that the combined classifier based on the writer-independent approach reduces the FRR, while keeping FAR at acceptable levels.

An ensemble of classifiers based on graphometric features is used to improve the reliability of the classification by Bertolini et al. [17]. A pool of base classifiers is first trained using only genuine signatures and random forgeries; then an ensemble is built using genetic algorithms with two different scenarios. In one, it is assumed that only genuine signatures and random forgeries are available to guide the search; while simple and simulated forgeries also are assumed to be available in the second one. Different objective functions are derived from the ROC curves, for ensemble tuning. A private database of 100 writers is utilized for evaluation, considering 5 genuine references for training and only skilled forgeries for testing. The best result is found as 11.14% DER using the area under curve optimization.

Score level combination is examined for offline signature verification by Prakash and Guru [18]. Classifiers of distance and orientation features are used individually and in combination. Distance features and orientation features individually provide 21.61% and 19.88% DER on the MCYT-75 corpus. The max fusion rule decreases the DER to 18.26%, while the average rule decreases the DER to 17.33% when the weights are fixed empirically.

Hybrid generative discriminative ensemble of classifiers is proposed by Batista et al. to design an offline signature verification system from few references, where the classifier selection process is performed dynamically [19]. To design the generative stage, multiple discrete left-to-right HMMs are trained using a different number of states and codebook sizes, allowing the system to learn signatures at different levels of perception. To design the discriminative stage, HMM likelihoods are measured for each training signature and assembled into feature vectors that are used to train a diversified pool of two-class classifiers through a specialized random subspace method. The most accurate ensembles are selected based on the K-nearest-oracles algorithm. The GPDS-160 database is used to evaluate the system and 16.81% EER is reported using 12 references per user.

An offline signature verification system using two different classifier training approaches is proposed by Hu and Chen [20]. In the first mode, each SVM is trained with feature vectors obtained from the reference signatures of the corresponding user and random forgeries, while the global Adaboost classifier is trained using genuine and random forgery signatures of signers that are excluded from the test set. Global and user-dependent classifiers are used separately. Combination of all features for writer-dependent SVMs results in 7.66% EER for 150 randomly selected signers from the gray-level GPDS-300 dataset, using 10 references. The