УДК 519.254

OFF LINE SIGNATURE VERIFICATION USING RADON TRANSFORM AND SVM/KNN CLASSIFIERS

A.A. Abdalla Ali, V.F. Zhirkov

Department of Information Technology, Vladimir State University;
mogwari@mail.ru

Represented by a Member of the Editorial Board Professor V.I. Konovalov

Key words and phrases: discrete Radon transform; extreme points warping; k-nearest neighbor (**KNN**); signature verification; support vector machine (**SVM**).

Abstract: We present a system for off-line signature verification, approaching the problem as a two-class pattern recognition problem. We use Discrete Radon Transform to extract global features from the signatures. During enrollment, a number of reference signatures are used for each registered user and cross aligned to extract statistics about that user's signature. A test signature's verification is established by first aligning it with each reference signature for the claimed user. The signature is then classified as genuine or forgery, according to the alignment scores which are normalized by reference statistics, using standard pattern classification techniques. We experimented with SVM classifier and KNN classifier. Using a database of 2250 signatures (genuine signatures and skilled forgeries) from 75 writers our present system achieves a performance of approximately 80 % when used SVM classifier and a performance of approximately 70 % in the case of KNN classifier.

1. Introduction

Signature verification is split into two types according to the available data in the input. Offline (static) signature verification takes as input the image of a signature and is useful in automatic verification of signatures found on bank checks and documents. Online (dynamic) signature verification uses signatures that are captured by pressure-sensitive tablets that extract dynamic properties of a signature in addition to its shape.

As a behavioral biometric, signature is not as unique or difficult to forge as iris patterns or fingerprints, however signature's widespread acceptance by the public, make it more suitable for certain lower-security authentication applications, as well as certain applications where online signatures can be the most suitable biometric (e.g. online banking and in credit card purchases). Furthermore, one's signature may change over time; yet, a person signs his/her signature rather uniquely at any given time period and forgeries can be identified by human experts quite well.

In an online or offline signature verification system, users are first enrolled by providing signature samples (reference signatures). Then, when a user presents a signature (test signature) claiming to be a particular individual, this test signature is compared with the reference signatures for that individual. If the dissimilarity value is above a certain threshold the user is rejected, otherwise authenticated. Since obtaining actual forgeries is difficult, two forgery types have been defined in signature verification papers: A skilled forgery is signed by a person who has had access to a genuine signature for practice. A random or zero-effort forgery is signed without having any information about the signature, or even the name, of the person whose signature is forged.

In the verification process, the test signature is compared to all the signatures in the reference set, resulting in several dissimilarity/distance values. One then has to choose a method to combine these distance values so as to represent the dissimilarity of the test signature to the reference set in a single number, and compares it to a threshold to make a decision. The single dissimilarity value can be obtained from the minimum, maximum or the average of all the distance values. Typically, a verification system chooses one of these approaches and discards the other ones. For instance, Jain et al. report the lowest error rates with the minimum distance criterion, among the other three [1]. We use the minimum and maximum in deciding whether the signature is genuine or not, instead of choosing which distance is most useful for the task. These distance values, normalized by the corresponding average values of the reference set, are used as the features of a signature in its classification as genuine or forgery, as explained in Section 2.

2. Proposed Method

During the enrollment phase, a set of reference signatures are used to determine user dependent parameters characterizing the variance within the reference signatures. The reference set of signatures, together with these parameters, are stored with a unique user identifier in the system's database.

In the training phase we choose a number of genuine and forged signatures for training each classifier.

In the verification phase when a test signature is input to the system, it is compared to each of the reference signatures of the claimed person. The person is authenticated if the resulting dissimilarity measure is low a threshold of the classifier, rejected otherwise. The details of the system are described in the following sections.

2.1. Discrete Radon Transform and feature extraction

The Discrete Radon Transform (**DRT**) is a matrix, where each column represents a projection or shadow of the original image at a certain angle. DRT can be expressed as follows [2], [4]:

$$R_j = \sum_{i=1}^{\Psi} w_{ij} I_i ; j = 1,2,..., N_\varphi N_\theta , \qquad (1)$$

where R_j – the cumulative intensity of the pixels that lie within the j^{th} beam; Ψ – total pixels in an image; w_{ij} – the contribution of the i^{th} pixel to the j^{th} beam-sum; I_i – the intensity of the i^{th} pixel; N_φ – nonoverlapping beams per angle; N_θ – number of total angles.

For extracting the global features firstly the background of the signature image is mapped to zero and the pen strokes to one. After that, median filtering is applied to remove speckle noise. Subsequently the DRT of the signature image is calculated, using the algorithm discussed in section 2.1 (Fig.1). This algorithm calculates the DRT at N_θ angles. These angles are equally distributed between 0 and 180°.

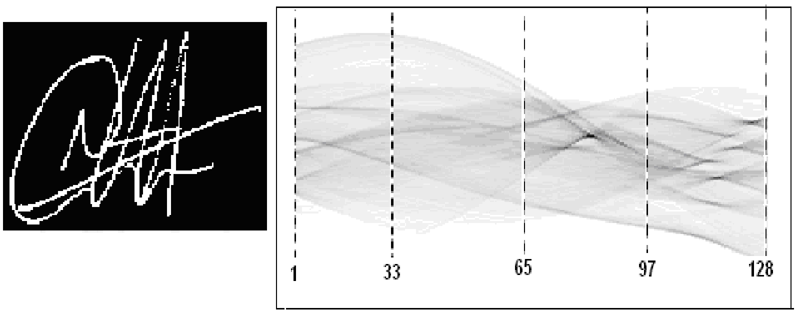


Fig. 1. A signature and its DRT, which is displayed as a gray-scale image. This image has 128 columns, where each column represents a projection

Although the DRT is not a shift invariant representation of a signature image, shift invariance is ensured by the subsequent image processing. This is done by removing (decimation) all the zero-valued components from each projection. These decimated vectors are then shrunk or expanded to the required dimension *d* through linear interpolation. Each vector is subsequently normalized by the variance of the intensity of the entire set of feature vectors. In order to ensure rotation invariance, the projections at angles that range from 180 to 360° are also included in the observation sequence.

An observation sequence therefore consists of $T = 2N_\theta$ feature vectors, that is

$$X_1^T = \{x_1, x_2, ..., x_T\}. \tag{2}$$

2.2. Signature Alignment

In order to compare two signatures of differing lengths, we use the extreme points warping (**EPW**) algorithm [5]. EPW algorithm warps a set of selected important points (peaks and valleys) instead of warping whole signal (Fig. 2). EPW algorithm finds the best linear alignment of two vectors such that the overall distance between them is minimized.

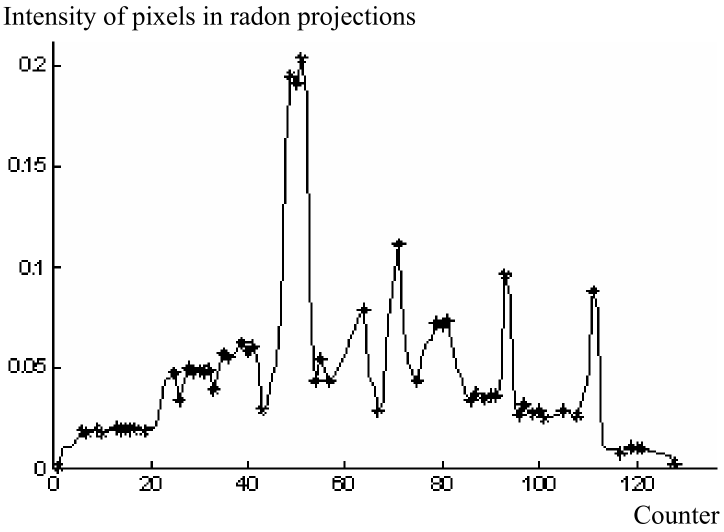


Fig. 2. Extreme points for radon projection:
—— – whole signal; ∗ – extreme points

In order to ensure that each observation sequence is a rotation invariant representation of the corresponding signature image, observation sequence alignment is necessary. The optimal alignment of two observation sequences can be achieved in a linear way. Iteratively shifts the observation sequences with respect to each other. During any iteration the distances between the corresponding observations (feature vectors) are calculated. The alignment is optimal when the average distance between the corresponding observations is a minimum. The distance between two signatures is simply the average of the distances between the optimally aligned feature vectors.

2.3. Enrollment

During enrollment to the system, we use a number of signatures (five in our system) for each user. These signatures are pair wise aligned to find the distance between each pair, as described in section 2.2.

From these alignment scores, the following reference set statistics are calculated:

1) average distance to farthest signature ($d_{\max}$);

2) average distance to nearest signature ($d_{\min}$).

A training data set consisting of five genuine signatures and five forgery signatures are used in order to learn the threshold parameter separating the forgery and genuine classes. These signatures are separate from the signatures used as reference signatures.

2.4. Training

First, each training signature is compared to the reference set of signatures it claimed to belong, using the EPW algorithm described in Section 2.3, giving a 2-dimensional feature vector ($p_{\min}$, $p_{\max}$). The feature values are then normalized by the corresponding averages of the reference set ($d_{\min}$, $d_{\max}$) this is calculated as in equations (3) and (4) to give the distribution of the feature set.

$$N_{\max} = d_{\max} / p_{\max} ; \quad (3)$$

$$N_{\min} = d_{\max} / p_{\max} ; \quad (4)$$

The distribution of this normalized data supports that genuine and forgery samples in the training set are well separated with these normalized features. Note that by normalizing the measured distance vectors by the corresponding reference set averages, we eliminate the need for user-dependent thresholds commonly used in deciding whether a signature is similar enough to the reference set.

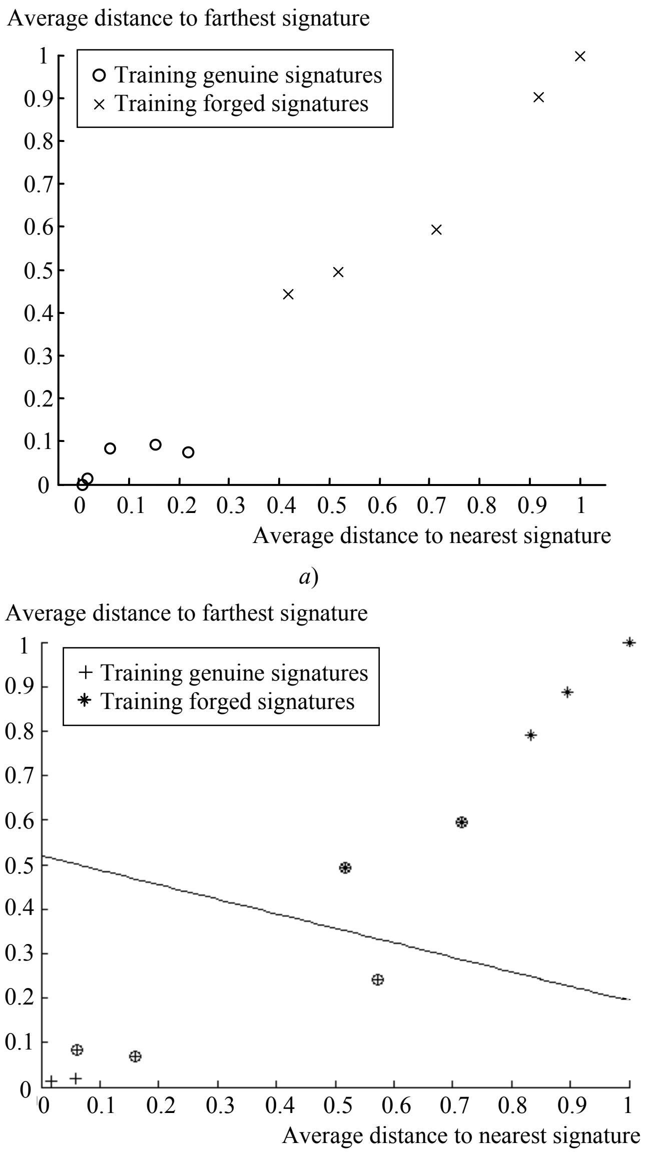
Finally, we train a classifier to separate the genuine and forgery samples in this normalized feature space shown in Fig. 3. For this work, we trained two classifiers: the SVM classifier and a KNN classifier using the 2-dimensional feature vectors. Then, a linear classification is made by picking a threshold value separating the two classes within the training set.

This threshold is fixed and later used in the verification process. The results are summarized in Section 3.

2.5. Verification

A verification data set consisting of five genuine signatures and ten forgery signatures are used in order to test the trained classifiers. These signatures are separate from the signatures used in the enrollment and in the training phases.

In order to verify a test signature as genuine or forgery, we first proceed as in the training stage: the signature is compared to all the reference signatures belonging to the



a)

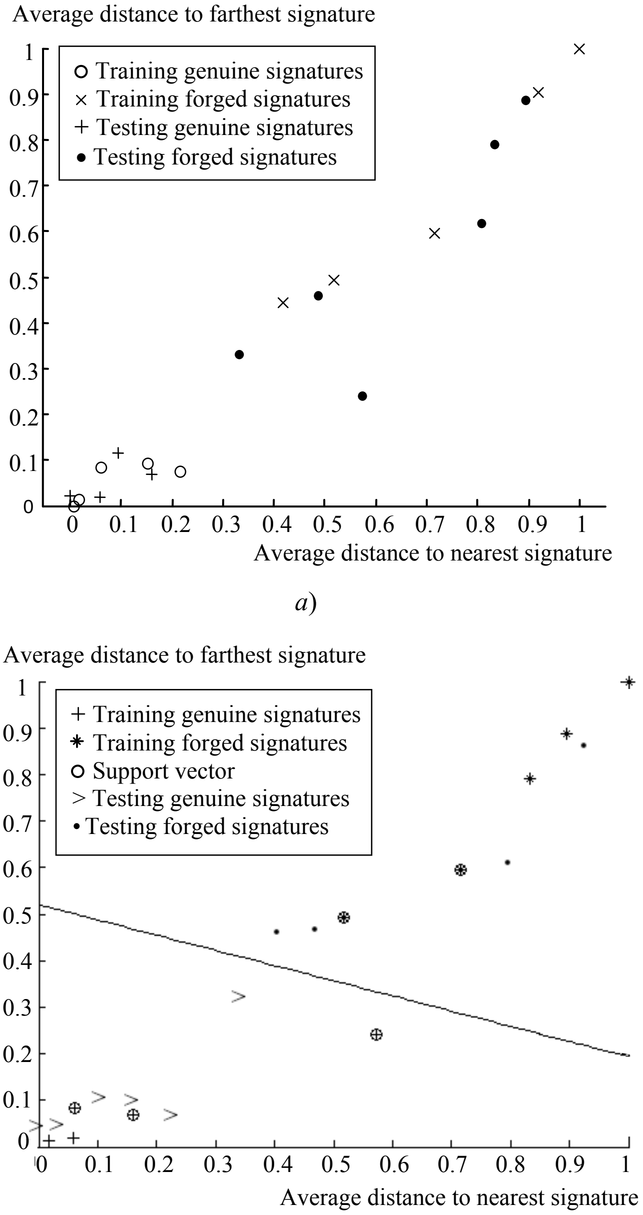
b)

Fig. 3. Training of KNN (*a*) and SVM (*b*) classifiers with respect to the 2-dimensional normalized distance vector

claimed ID using the EPW algorithm described in Section 2.2. Then, the resulting distance values ($p_{\min}$, $p_{\max}$), normalized by the averages of the claimed reference set ($d_{\min}$, $d_{\max}$), then these normalized values are used in classifying the signature as genuine or forgery, by the trained classifier (Fig. 4).

3. Experiments and Results

Experiments are conducted on a dataset; which called «MCYT-100 signature CORPUS» [3]; that contain static signature images. The dataset contains 2250 signatures from 75 writers. Each writer has 15 genuine signature and 15 skilled forgeries.



a)

b)

Fig. 4. Verification results obtained by KNN (*a*) and SVM (*b*) classifiers using the 2-dimensional normalized data

For each writer setup 5 genuine signatures are used for profile creation (reference set) and the rest of signatures are used for training and testing. Note that training data is separate from both the reference set of genuine signatures and the test data used in experiments.

Best results were obtained using SVM Classifier with performance of approximately (80 %) and the results with the KNN classifier is (70 %).

4. Summary and Conclusion

We have presented an offline signature verification system that approaches the problem as a two-class pattern recognition problem.

We use DRT for global feature extraction from the signatures and it shows us that it is a stable and robust method. The DRT creates simulated time evolution from one feature vector to the next and enables us to model a signature with EPW.

We experimented with two different classifiers and obtained 80 % overall performance for a data set of 75 people and 2250 signatures (genuine signatures and skilled forgeries). These results are quite good, given the fact that the forgeries used on the experiments were skilled forgeries.

We thank J. Ortega-Garcia and J. Fierrez-Aguilar for the provision of the MCYT signature sub-corpus.

References

1. Jain, A., Griess, F., Connell, S. On-line signature verification. Pattern Recognition 35 (2002) 2963–2972.

2. Bracewell, R.N. Two-Dimensional Imaging, Prentice-Hall, Englewood Cliffs, NJ, USA, 1995.

3. Ortega-Garcia, J. Fierrez-Aguilar, J. Simon, D. Gonzalez, J. Faundez-Zanuy, M. Espinosa, Satue, V.A. Hernaez, I. Igarza, J.-J. Vivaracho, C. Escudero D. and Moro, Q.-I. «MCYT baseline corpus: a bimodal biometric database», IEE Proc.-Vis. Image Signal Process., Vol. 150, No. 6, December 2003.

4. Peter Toft, «The Radon Transform – Theory and Implementation», Ph.D. thesis, Technical University of Denmark, June 1996, 326 p.

5. F. Hao, C.W. Chan, «Online signature verification using a new extreme points warping technique,» Pattern Recognition Letters, Vol 24, Issue 16, 2943-2951 (2003).

Верификация статических подписей с использованием преобразования Радона и SVM/KNN классификаторов

А.А. Абдалла Али, В.Ф. Жирков

Факультет информационных технологий,
ГОУ ВПО «Владимирский государственный университет»;
mogwari@mail.ru

Ключевые слова и фразы: дискретное преобразование Радона; k-ближайших соседей (**KNN**); поиск соответствия экстремальных точек; методы опорных векторов (**SVM**); верификация подписи.

Аннотация: Рассматривается метод верификации статических подписей с использованием дискретного преобразования Радона. Приведены процедуры обработки и сравнения подписей с учетом особенностей их представления. В экспериментах используется база данных, содержащая 2250 подписей 75 авторов, для каждого из которых имелись контрольные подписи и квалифицированные подделки. Верификация проводится путем сравнения с каждой контрольной подписью автора. Классификация на подлинник/подделку выполняется по результатам сравнения, которые нормализуются с использованием контрольных статистических данных. Используются стандартные классификаторы SVM/KNN. Доля правильно распознанных подписей составила 80 % для SVM и 70 % для KNN.

Verifikation der statischen Signaturen mit der Benutzung der Radon-Transformation und der SVM/KNN Klassifikatoren

Zusammenfassung: Es wird die Methode der Verifikation der statischen Signaturen mit der Benutzung der diskreten Radon-Transformation betrachtet. Es sind die Prozeduren der Bearbeitung und der Vergleichung der Signaturen mit der Berücksichtigung der Besonderheiten ihrer Vorlegung angeführt. In den Experimenten wird die Datenbank, die 2250 Signaturen der 75 Autoren enthält, benutzt. Es wurden die Kontrollsignaturen und qualifizierte Fälschungen. Die Klassifikation auf Original und Fälschung wird nach den Vergleichungsresultaten erfüllt. Es werden die Standartklassifikatoren SVM/KNN benutzt. Der Anteil der richtigerkannten Signaturen bildete 80 % für SVM und 70 % für KNN.

Vérification des signatures statiques avec l'emploi de la transformation Radon et classificateurs SVM/KNN

Résumé: Est examinée la méthode de la vérification des signatures statiques avec l'emploi de la transformation discrète Radon. Sont cités les procédés du traitement et de la comparaison des signatures compte tenu de leur présentation. Dans les expériments est utilisée une base de données contenant 2250 signatures de 75 auteurs pour chacun desquels il y avaient les signatures de contrôle et les imitations qualifées. La vérification était réalisée par la voie de la comparaison avec chaque signature de contrôle de l'auteur. La classification identité/imitation était exécutée par les résultats de la comparaison normalisée avec l'emploi des données statistiques de contrôle. Sont utilisés les classificateurs standartisés SVM/KNN. La part des signatures identifiées correctement est 80 % pour les méthodes des vecteurs de contrôle et 70 % pour les k-voisins les plus proches.

Авторы: ***Абдалла Али Ахмед Абдельрахман*** – аспирант кафедры вычислительной техники факультета информационных технологий; ***Жирков Владислав Федорович*** – кандидат технических наук, профессор кафедры вычислительной техники факультета информационных технологий, ГОУ ВПО «ВГУ».

Рецензент ***Литовка Юрий Владимирович*** – доктор технических наук, профессор кафедры «Системы автоматизированного проектирования» ГОУ ВПО «ТГТУ».