A new approach for High Pressure Pixel Polar Distribution on Off-line Signature Verification

Jesús F. Vargas[1],
Miguel A. Ferrer[2],
Carlos M. Travieso[3],
Jesús B. Alonso[4]

Abstract:

Features representing information of High Pressure Points from a static image of a handwritten signature are analyzed for an offline verification system. From grayscale images, a new approach for High Pressure threshold estimation is proposed. Two images, one containing the High Pressure Points extracted and other with a binary version of the original signature, are transformed to polar coordinates where a pixel density ratio between them is calculated. Polar space had been divided into angular and radial segments, which permit a local analysis of the high pressure distribution. Finally two vectors containing the density distribution ratio are calculated for nearest and farthest points from geometric center of the original signature image. Experiments were carried out using a database containing signature from 160

1   Departamento de Ingeniería Electrónica, GEPAR, Universidad de Antioquia, Colombia, jfvargas@udea.edu.co

2   Centro Tecnológico para la Innovación en Comunicaciones, Departamento de señales y Comunicaciones, Universidad de las Palmas de Gran Canaria, España, mferrer@dsc.ulpgc.es

3   Centro Tecnológico para la Innovación en Comunicaciones, Departamento de señales y Comunicaciones, Universidad de las Palmas de Gran Canaria, España, ctravieso@dsc.ulpgc.es

4   Centro Tecnológico para la Innovación en Comunicaciones, Departamento de señales y Comunicaciones, Universidad de las Palmas de Gran Canaria, España, jalonso@dsc.ulpgc.es

individual. The robustness of the analyzed system for simple forgeries is tested out with Support Vector Machines models. For the sake of completeness, a comparison of the results obtained by the proposed approach with similar works published is presented.

Keywords:

Offline signature verification system, High pressure points, pixel density distribution, LSSVM.

Resumen:

Se analiza el uso de características que representan la información de los Puntos de Alta Presión en la imagen estática de una firma manuscrita, en un sistema orientado a la verificación tipo offline de firmas. A partir de imágenes en niveles de gris, se propone un nuevo método para calcular el umbral de alta presión. Realizando la transformación de dos imágenes a coordenadas polares, una con la información de los Puntos de Alta Presión y otra con la versión binarizada de la imagen original, se calcula la densidad de pixeles entre ellas. El espacio polar se divide en secciones angulares y radiales, lo que permite realizar un análisis localizado de la distribución de los puntos de alta presión. Se calculan los vectores con la distribución de la densidad para las dos regiones radiales definidas (cercada y lejana) a partir del centro geométrico de la imagen original. Para las pruebas se usó una base de datos que contiene muestras de las firmas de 160 personas. Para validar la metodología planteada se construyeron modelos basados en Maquinas de Soporte Vectorial. Con el ánimo valorar los resultados obtenidos, se presentan algunos resultados obtenidos por otros autores.

Palabras clave:

Sistema de verificación de firma offline, puntos de alta presión, distribución de densidad de píxeles, LSSVM.

1. INTRODUCTION

Biometric verification is an important research area where applications have as target the automatic identity verification. There are two types of biometrics: physiological, e.g. iris pattern and finger print; and behavioral e.g. speech and handwriting. Handwritten signature verification is a behavioral biometric verification. Signing is the most traditional and social accepted method of identification and most of us are familiar with the process of verifying some handwriting against a signature record for identification, especially in legal, banking, and other high security environments (Huang, 1997).

There are two major methods of signature verification (Ortega-Garcia, 2003). One is an on-line method to measure the sequential data such as handwriting speed and pen pressure with a special device. The other is an off-line method that uses an optical scanner to obtain handwriting data written on paper. There are two main approaches for off-line signature verification: static approaches and pseudodynamic approaches (Fierrez-Aguilar, 2004). The static one involves geometric measures of the signature while pseudodynamic one tries to estimate dynamic information from the image.

Offline approaches are more difficult to apply than online systems. Online systems use special input devices such as tablets, whereas the only available information in an offline system is a static two-dimensional image obtained from scanned prewritten signatures on a paper. The dynamic information of the pen-tip (stylus), movement such as pen-tip coordinates, pressure, velocity, acceleration, and pen-up and pen-down can be captured by a tablet in real time but not by an image scanner. The offline method, therefore, needs to apply complex image processing techniques to segment and analyze signature shape for feature extraction (Kato, 2000).

Signature Verification in off-line systems is more difficult than in on-line systems as a lot of dynamic information is lost. Hence, online signature verification is generally more successful.

Nevertheless, off-line systems have a significant advantage in that they do not require access to special processing devices when the signatures are produced. In fact, if the accuracy of the verification promoted greatly, the off-line method has much more practical application areas than that of the on-line one. Consequently more and more researches have looked into the feature-extraction methodology of offline signature recognition and verification (Lee, 1992).

Dynamic information cannot be derived directly from static signature images. Instead, some features can be derived that partly represent dynamic information. These special characteristics are referred to as pseudo-dynamic information. The term "pseudo-dynamic" is used to distinguish real dynamic data, recorded during the writing process, from information, which can be reconstructed from the static image (Cardot, 1994). A grey level image of a scanned handwritten signature indicates that some pixels may represent shapes written with high pressure which appear as darker zones. High Pressure Points (HPPs) can be defined as those signature pixels which have grey level values greater than a suitable threshold. The study of high pressure features was proposed by Ammar et al. (1986) to indicate regions where more physical effort was made by the signer.

This idea of calculating a threshold to find the HPP was adopted and developed by others researchers (Huang, 1997; Sansone, 2000). Lv et al. (2005) set two thresholds to store only the foreground points and edge points. They analyze only the remaining points whose grey level value is between the two thresholds and divide them into 12 segments. The percentage of the points whose grey level value falls in the corresponding segment is one of the values of the feature vector which reflects the grey level distribution. Lv et al. also consider stroke width distribution. In order to analyze not only HPPs but also Low Pressure Points (LPP) a complementary threshold has been proposed by Mitra et al. (2005). In this work, a new procedure for high pressure threshold calculation and its polar distribution estimation as new approach for signature

parametrisation is presented. Once the HPP's are extracted from original grayscale image, a transformation to polar coordinates is carried out, and then we calculate the density distribution ratio of HPP's respect to a binary version of the original image. Polar space had been divided into angular segments, which permit a local analysis of the high pressure distribution; an analysis about how this division could be carried out is presented.

The paper is organized as follows: Section 2 introduces the features extraction. Section 3 is devoted to the classifiers. Section 4 discuss about the Database. Section 5 presents the evaluation protocol and reports the experimental results, and the paper ends with concluding remarks.

2. MATERIALS AND METHODS

In this section, we describe the HPP's extraction method and the calculation of the density distribution ratio in polar space.

2.1 High Pressure Point´s extraction

When we analyze a grayscale image containing a scanned handwritten signature, it is possible to say that pixels representing shapes written with high pressure appear as darker zones. In this way, High Pressure Points are those signature pixels which have gray level values upper than a suitable threshold.

High pressure feature has been used by Ammar et al. (1986) to indicate regions where more emphasis has been made by the signer. In referenced work, a threshold θ_{hpr} is set to be as follows:

$$\theta_{hpr} = g_{min} + 0.75(g_{max} - g_{min}) \tag{1}$$

where g_{max} and g_{min} represent maximum and minimum gray level intensities respectively (when white:255 and black:0). This approach has been used too by Huang et al. (1997), and Sansone et al. (2000) in their respective works. Recently, Mitra et al. (2005), proposed a different procedure for threshold selection; from the

original gray level density information of the signature, a threshold point is adaptively selected at the gray levels which correspond to 1/sqrt(2) of the peak frequency (maximum number of pixels at a particular gray level in any given signature) of the same normalized histogram of the signature in question. In other way, Lv et al. (2005) noted that nearly all of the pixels gray values in a signature image exceed a given threshold, so they just set two thresholds to hold only the foreground points and edge points. The two thresholds were set to 85 and 205 based on experimental results. This established range is divided into 12 segments, and the percentages of the points whose gray value falls in each segment are store in a 12-D vector which reflects the gray distribution.

2.2 Threshold estimation proposed

From 256-graylevel histogram information, the more frequent element *GmaxHis*, is calculated after background gray levels have been removed, and then *Gmin* value is define as:

$$hisI=histogram(I) \tag{2a}$$

$$GmaxHis=max \ (hisI) \tag{2b}$$

$$Gmin=min \ (hisI>mean(hisI) \ ) \tag{2c}$$

where *hisI* is a vector containing the values from histogram of the grayscale image. Using Eq. 2, the *HP* threshold could be defined as:

$$HPPthres=Gmin+\gamma(GmaxHis-Gmin) \tag{3}$$

where ɣ is an empirically determined factor. Fig. 1 illustrates *GmaxHis, Gmin* and *HPPThresh* estimation. HPP's can be extracted by the relation

$$I_{ij}^{HPP} = \begin{cases} 1 & \forall \ I_{ij} \leq HPPThresh \\ 0 & otherwise \end{cases} \tag{4}$$

where is an image matrix containing only HPP's, and $i=1...M$, $j=1...N$, M and N being the number of rows and columns of the original image I, respectively.
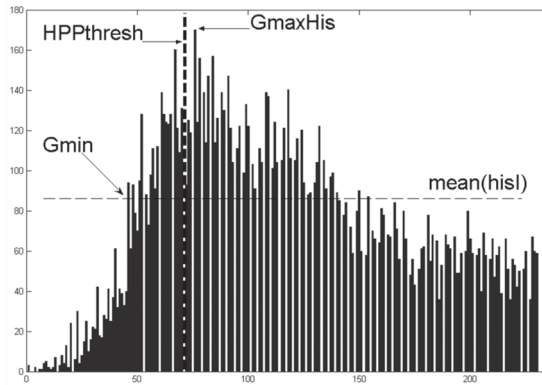

FIG. 1. *HPP* THRESHOLD AND *GMAXHIS* LEVEL ESTIMATION IN GRAY LEVEL HISTOGRAM. BACKGROUND´S GRAY LEVELS HAVE BEEN REMOVED

FIG. 2 SHOWS A SAMPLE SIGNATURE FROM DATASET AND RESULTING HPP IMAGES FOR GENUINE AND FORGERY CASE.
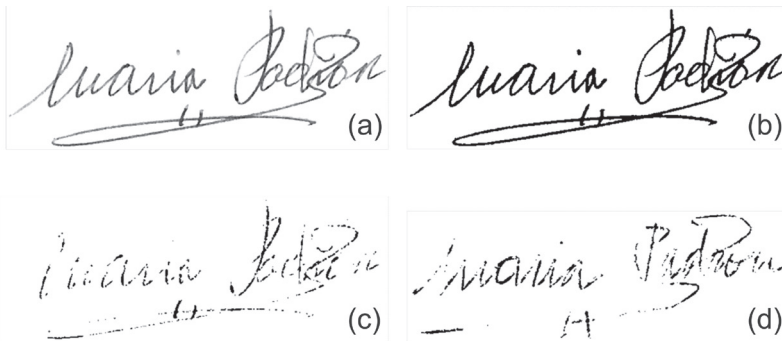

FIG. 2. A) GENUINE SCANNED SIGNATURE IMAGE, B) BINARY IMAGE, C) HPP'S DETECTED FOR GENUINE SIGNATURE, AND D) HPP´S DETECTED FOR FORGERY SIGNATURE.

2.3 Polar Density Distribution

Ferrer et al. (2005) proposed a new approach based on geo-metrical feature which represents the interior stroke distribution

in polar coordinates. The new component contains the number of black pixels of the signature strokes that the radius crosses when sweeping from θ_t to θ_{t+1} normalized to maximum value equal to 1 in order to increase the stroke thickness independence.

In this work, we propose the polar density distribution estimation as a complement to HPP's detection. Both HPP image and a binary version of the original image are transformed to polar coordinates, and there the geometric center is calculated and subtracted for every point of the images. To determine point distribution over polar space, angular division is carried out, and then density ratio between HPP's and Binary points contained in each angular segment θ_k is calculated as:

$$HPPPD_k = \frac{\sum_{i=1}^{M} \sum_{j=1}^{N} I_{ij}^{HPP} \in \theta_k}{\sum_{i=1}^{M} \sum_{j=1}^{N} I_{ij}^{bin} \in \theta_k} \qquad (5)$$

where , $i=1…M$, $j=1…N$, represents the binary version of the original signature image I. Depending how far points are from geometric center, two zones were established. So each angular segment contains two subzones labeled as *near* and *far* respectively. The threshold τ for this separation is calculated as:

$$\tau = \frac{\max(R_{bin})}{2} \qquad (6)$$

where R_{bin} is a vector containing the normalized ratios of transformed binary image. Finally, two vectors contained the HPPPD of the signature will be took into account for the verification stage. Fig. 3 presents signature representation in polar coordinates and the HPP's extracted. Fig. 4, illustrates a comparison between two HPPPD vectors extracted for genuine and forgery samples of signature showed in Fig. 2.
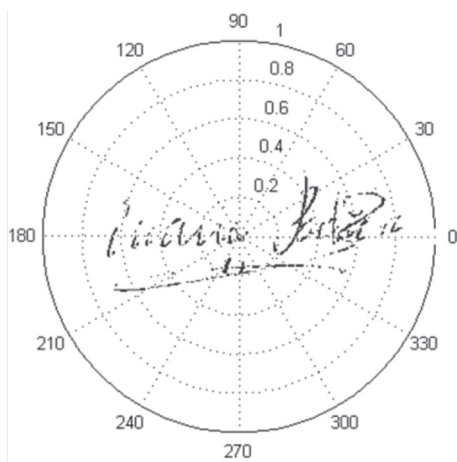
F<small>IG</small>. 3. P<small>OLAR REPRESENTATION OF</small> HPP'<small>S EXTRACTED</small>

2.4 Classifiers

In order to avoid the curse of dimensionality, a Karhunen-Loeve Mapping was used to perform a Principal Component Analysis (PCA) on the mean class covariance matrix of the training set. This mapping finds a rotation of the dataset to an N-dimensional linear subspace such that at least 90% of the total variance is preserved. Once the matrix transformation is estimated, we need to solve a two-class classification (genuine or forgery) problem.
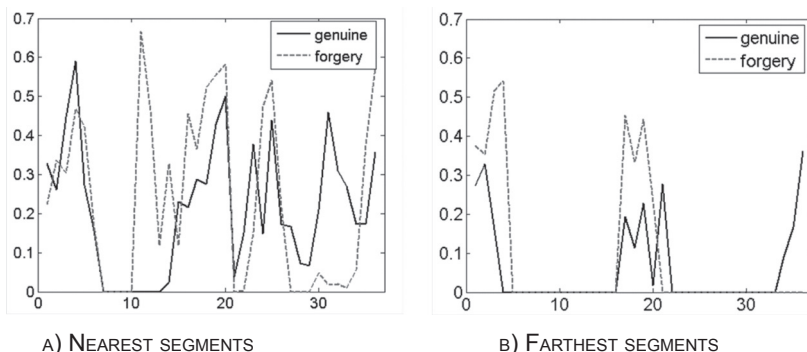

A) N<small>EAREST SEGMENTS</small>   B) F<small>ARTHEST SEGMENTS</small>

F<small>IG</small>. 4. F<small>EATURE VECTORS COMPARISON FOR</small> A) N<small>EAREST AND</small> B) F<small>ARTHEST ANGULAR SEGMENTS OF GENUINE AND FORGERY SAMPLES</small>

As verifier we have used a Least Squares Support Vector Machine (LS-SVM). SVMs have been introduced within the context of statistical learning theory and structural risk minimization. Least Squares Support Vector Machines (LSSVM) are reformulations to standard SVMs which lead to solving linear KKT systems. Robustness, sparseness, and weightings can be imposed to LS-SVMs where needed and a Bayesian framework with three levels of inference is then applied (Suykens, 2002).

The meta-parameters of the LS-SVM model are the width of the Gaussian kernels (taken identical for all kernels) and the regularization factor. For this work, a RBF kernel with $\gamma=20$ and a σ parameter is optimized experimentally in the range [0,…, 1000] for each signature. Keeping in mind the limited number of samples in the training, leave-one-out cross-validation was used for training.

To verify that a questioned signature is an authentic signature, we calculated the distance of the questioned signature features to the separator hyper plane of the LS-SVM that models the authentic signature. If the distance is greater than a threshold, the questioned signature is accepted as authentic.

2.5 GPDS signature corpus

The GPDS-160 signature corpus contains 24 genuine signatures and 24 forgeries of 160 individual (Vargas, 2007). So, there are 160 x 24 = 3840 genuine signatures and forgeries. The genuine signatures were taken in just one session. To organize so many people in different session was a lost fight. The signers filled up a form with 24 boxes of different size. The forgeries were collected by form with 15 boxes. Each forger form contains 5 images of different genuine signatures chosen randomly. The forger imitated 3 times each one of the 5 signs. They taken as long as they like to learn the signature and perform the forgeries. As the forgers are not expert people, these forgeries are simple forgeries. The whole process of signing was accomplished under the supervision of an operator.

Once the signature forms were collected, each form was scanned with a HP4400 device using 256 level gray scale and

300dpi resolution. Once the images have been scanned, they were converted to black and white using a global threshold that was computed using Otsu's method. The resulting binary image is then passed through a median filtering to eliminate anomalies such as minor discontinuities and salt and pepper noises. Gray scale images were kept in order to deal with HPP approach.

2.6 Algorithm benchmarking

In this section, it is studied the ability of the GPDS-160 signature corpus for producing statistically significant results. In (Guyon, 1998) the minimum size of the test data set, N that guarantees statistically significant results in a pattern recognition task is derived. The goal in the above mentioned work is to estimate N so that it is guaranteed, with a risk α of being wrong, that the error rate P does not exceed that estimated from the test set, $\hat{P}$, by an amount larger than $\varepsilon(N,a)$, that is

$$\Pr\{P > \hat{P} + \varepsilon(N,\alpha)\} < \alpha \qquad (9)$$

Letting $\varepsilon(N,a) = \beta P$, and supposing errors as Bernoulli trials (independently and identically distributed errors) and for typical values of $\alpha = 0.05$ and $\beta = 0.2$, the following simplified criterion is obtained: $N \approx 100/P$.

This ensures that with 95 percent confidence the probability of error is not worse than 1.25 $\hat{P}$. For instance, with $N = 10,000$ samples, with confidence 95%, our experiment guarantee statistical significance with an empirical error rate $\hat{P}$ down to 0.01%. If the samples in the test dataset are not independent (due to correlation factors) then N must be further increased. For a detailed analysis of this case, see (Guyon, 1998).

If we suppose that we train with 10 signatures and test with the remainder signatures, the minimum empirical error with statistical significance can be seen in Table 1 for several signature corpuses referenced there. The random forgeries experiment refers to use as forgeries the genuine signatures of the other signers and the

simple forgeries experiment uses as forgeries the simple forgeries of the corpus.

TABLE 1. MINIMUM EMPIRICAL ERROR WITH STATISTICAL SIGNIFICANCE TRAINING WITH 10 GENUINE SIGNATURES

Corpus Name	Random forgeries	Simple forgeries
GPDSsignature (Ferrer, 2005)	0,03%	1,44%
SVC2004 (Yeung, 2004)	0,78%	10,4%
MCYT-100 (Ortega-Garcia, 2003)	0,08%	3,1%
MCYT-75 (Fierrez-Aguilar, 2004)	0,44%	8,3%

3. EVALUATION PROTOCOL

The system was trained respectively with 5 and 10 genuine samples. These samples were chosen randomly. Random forgeries (genuine samples from other signers) were used as negative samples, in a similar way to that outlined by (Bertolini, 2009), in our case taking a genuine sample of each one of the other users of the database (99 for the GPDS corpus). Keeping in mind the limited number of samples in the training, leave-one-out cross-validation (LOOCV) was used to determine the parameters of the SVM classifier with RBF kernel (γ, σ). For testing, skilled forgeries were taken into account. All available forgeries were used, this is, 24 for the GPDS corpus.

Training and testing procedure were repeated 10 times with different training and testing subsets for the purpose of obtaining reliable results. Two classical types of error were considered: Type I error or False Rejection Rate (FRR), which is when an authentic signature is rejected, and Type II error or False Acceptance Rate (FAR), which is when a forgery is accepted. Finally the Equal Error Rate (EER) was calculated keeping in mind that the classes are unbalanced.

4. RESULTS

Table 2 shows system performance when HPPPD is calculated in polar coordinates. Optimal value for θ-factor corresponding angular division is set to 15, that is, every 15 degrees the pixel density is extracted in each angular segment.

TABLE 2. θ-FACTOR TUNING FOR ANGULAR DIVISION IN POLAR COORDINATES. TYPE I, TYPE II AND AVERAGE ERROR ARE PRESENTED

θ	%FAR	%FRR	AvE
5	16,10	9,98	13,04
10	16,26	9,77	13,01
15	17,15	8,14	12,64
20	16,29	8,99	12,64
30	16,56	8,41	12,48
45	16,43	8,81	12,62

Table 3 shows γ-factor tuning results. Suitable value for γ is set to 0,90, that is, HPP threshold is displaced 0.90% of *GmaxHis* and *Gmin* difference where *Gmin* value conforms with Eq. 2, here θ=15 from Table 2.

TABLE 3. γ-FACTOR TUNING. TYPE I, TYPE II AND AVERAGE ERROR ARE PRESENTED

γ	%FAR	%FRR	AvE
0,60	18,10	8,61	13,35
0,65	17,81	8,96	13,39
0,70	17,92	8,55	13,23
0,75	17,96	8,51	13,24
0,80	17,59	8,45	13,02
0,85	17,52	8,39	12,95
0,90	17,15	8,14	12,64
0,95	17,56	8,28	12,92

A comparison of the performance of different signature verification systems is a difficult task since each author constructs his own signature datasets. The lack of a standard international signature database continues to be a major problem for performance comparison. For the sake of completeness, in Table 4 we present some results obtained by similar works published which include features based on pressure and pixel distribution information. Although it is not possible to carry out a direct comparison of the results, since the methodologies of training and testing, and the classification strategies used by each author are different, Table 4 enables one to visualize results from the proposed methodology alongside results published by other authors. Table 5 describes dataset used by other authors.

TABLE 4. COMPARISON OF PROPOSED APPROACH WITH OTHER PUBLISHED METHODS

	%FAR	%FRR	%AvError
Ammar et al. (1986)	6,50	4,00	5,25
Lv et al. (2005)	5,30	4,60	5,00
Huang et al. (1997)	11,80	11,10	11,45
Samsone et al. (2000).	4,29	2,04	3,16
Mitra et al. (2005)	2,50	4,00	3,25
Ferrer et al. (2005)	12,60	14,10	13,35
Proposed app.	17,15	8,14	12,64

TABLE 5. DATASETS USED BY OTHER AUTHORS

	Signers	Genuine	Forgeries
Ammar et al. (1986)	20	10	10
Lv et al. (2005)	20	25	30
Huang et al. (1997)	21	24	24
Samsone et al. (2000)	49	20	10+10
Mitra et al. (2005)	20	10	10
Ferrer et al. (2005)	160	24	24
Proposed app.	160	24	24

5. CONCLUSIONS

A new off-line signature verification method based on geometric feature extraction of High Pressure Points Polar Density Distribution in grayscale images is described. Fig. 5 shows the proposed system flow-diagram. The performance of the system have been presented with reference to a big-size experimental signature database containing samples from 160 individual including simple forgeries. The experimental results show that using HPP´s density distribution in polar coordinates achieves acceptable system performance when compared with similar systems. It is important to remark that in our system only Pressure information was used for verify while other works use a combination of different features.

In Table 2, the proposed system provides the lowest FRR when θ=5, that suggest future research on a multistage system take into account to different resolutions, for angular segmentation of the polar space, as complement to each other.
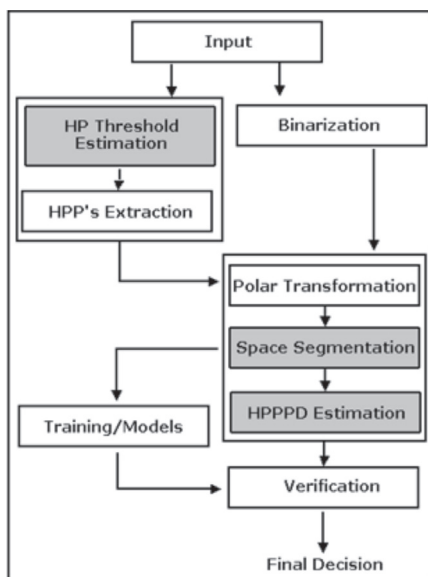
FIG. 5. SIGNATURE VERIFICATION SYSTEM DIAGRAM. IT SHOWS THE FLOW PATH FOR BOTH THE TRAINING AND TESTING PHASE OF THE SYSTEM PROPOSED

6. ACKNOWLEDGMENTS

7. REFERENCES

Ammar, M., Yoshida, Y., Fukumura, T., (1986); A new Effective Approach for Automatic Off-line Verification of Signatures by using Pressure Features, in Proceedings 8th International Conference on Pattern Recognition, Paris, 566-569.

Bertolini, D., Oliveira, L., Justino, E., Sabourin, R., (2009); Reducing forgeries in writer-independent off-line signature verification through ensemble of classifiers, Pattern Recogn. 43(1), 387–396.

doi:http://dx.doi.org/10.1016/j.patcog.2009.05.009.

Cardot, H., Revenu, M., Victorri, B., Revillet, M., (1994); A static signature verification system based on a cooperative neural network architecture, International Journal on Pattern Recognition and Artificial Intelligence, 8(3), 679–692.

Ferrer, M., Alonso, J., Travieso, C., (2005); Offline geometric parameters for automatic signature verification using fixed-point arithmetic, IEEE Transactions on Pattern Analysis and Machine Intelligence, 27(6), 993-997.

Fierrez-Aguilar, J., Alonso-Hermira, N., Moreno-Marquez, G., Ortega-Garcia, J., (2004); An off-line signature verification system based on fusion of local and global information, Workshop on Biometric Authentication, Springer LNCS-3087, 295-306.

Guyon, I., Makhoul, J., Schwartz, R., Vapnik, V., (1998); What size test set gives good error rate estimates?, IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(1), 52–64.

Huang, K., Yan, H., (1997); Off-line signature verification Based On Geometric Feature Extraction and Neural Network classification, Pattern Recognition, Elsevier Science, 30(1), 9-17.

Kato, Y., Yasuhara, M., (2000); Recovery of Drawing Order from single-Stroke Handwriting Images, IEEE Transaction on Pattern Analysis and Machine Intelligence, 22(9).

Lee, S., Pan, J., (1992); Offline tracking and Representation of Signatures, IEEE Transactions on Systems, Man and Cybernetics, 22(4), 755-771.

Lv, H., Wang, W., Wang, C., Zhuo, Q., (2005); Off-line Chinese Signature Verification Based on Support Vector Machines, Pattern Recognition Letters, Elsevier, 26, 2390-2399.

Mitra, A., Kumar, P., Ardil, C., (2005); Automatic Authentification of Hand-written documents via Low Density Pixel Measurements, International Journal of Computational Inteligence, 2(4), 219-223.

Ortega-Garcia, J., Fierrez-Aguilar, J., Simon, D., Gonzalez, J., Faundez-Zanuy, M., Espinosa, V., Satue, A., Hernaez, I., Igarza, J.J., Vivaracho, C., Escudero, D., Moro, Q.I., (2003); MCYT baseline corpus: a bimodal biometric database, IEE Proceedings of Visual Image Signal Processsing, 150(6).

Sansone, C., Vento, M., (2000); Signature Verification: Increasing Performance by a Multi-Stage System, Pattern analysis & Applications, Springer, 3, 169-181.

Suykens, J.A.K., Gestel, T.V., Brabanter, J.D., Moor, D., Vandewalle, J., (2002); Least Squares Support Vector Machines, World Scientific Publishing Co., Pte, Ltd.

Vargas, J.F., Ferrer, M., Travieso, C., Alonso, J., (2007); Off-line Handwritten Signature GPDS-960 Corpus, Ninth ICDAR 2007, 2, 764-768.

Yeung, D., Chang, H., Xiong, Y., George, S., Kashi, R., Matsumoto, T., Rigoll, G., (2004); SVC2004: First International Signature Verification Competition, Proceedings of the International Conference on Biometric Authentication, Hong Kong, 15-17 July.