signature recognition,
DTW method
linear regression

Tomasz PARA*, Piotr PORWIK*, Krzysztof WRÓBEL**

ON-LINE SIGNATURE RECOGNITION METHOD BASED ON LINEAR REGRESSION

Nowadays, automatic signature verification is an active area of researches in numerous applications such as bank check verification, access restriction or special areas such as police investigations. In our researches signature was captured by Topaz SigLite T-LBK750-HSB device, where some dynamic features of signature can be also registered. In many transactions, the electronic verification of a person's identity is beneficial, hence it inspires the development of a wide range of automatic identification systems. In this paper the system that automatically authenticates documents based on the owner's handwritten signature is presented.

1. INTRODUCTION

Signature verification is found already as a traditional biometric method, having wide public acceptance, particularly in authentication and authorization within financial and transactions legalisation process. Many works, conducted in past, indicate that signature analysis remains a complex pattern recognition problem. Automatic signature analysis can be categorised into two types: the on-line [11, 12, 13] and the off-line signature verification [8, 9].

In the on-line signature analysis special pens are used, where pressure and dynamic movements can be recorded. For this reason the off-line signature analysis is more complex due to the absence of stable dynamic characteristics.

Signature verification can be treated as a decision-making process, where the original signature is compared to another signature.

Many signature analysis schemes have been investigated in the past years. Among others the Hidden Markov models [5], the Hough transform [3,4] or vector quantisation method [6] were proposed. Fortunately, there are modern devices, which are able to register dynamic features of signature, such as pen pressure, acceleration, velocity or pen location on the surface. For this reason, the off-line methods of signature analysis can be joined with the on-line methods, where unique dynamic signature features can also be analysed. Unfortunately, due to lack of benchmarks with the on-line signatures, in our works our own database has been used, where 30 signatures and its dynamic features are stored. They were captured during four days, creating our database 4×30=120 signatures. Finally, the database includes bitmap image of signature, coordinates of each point of signature and signature

* University of Silesia. Institute of Informatics, 41-200 Sosnowiec, Będzińska 39, Poland
** University of Silesia. Institute of Material Science, 40-007 Katowice, ul. Uniwersytecka 4, Poland

dynamic features for any bitmap. Dynamic features (time and pen pressure and also coordinates of each point) are stored as a textual file, where any feature is expressed by appropriate value.

2. PREVIOUS RESEARCHES

Our previous researches were carried out only for static features. These features were obtained from the scanned images of the signature [3,4]. In the discussed works as the first step special preprocessing was applied, where binarization, cutting edges and thinning procedures were used. In the next stage the Hough algorithm was applied and characteristic signature features were extracted, as: proportion factor, centre of gravity, vertical and horizontal projection. The obtained signature recognition results were very accurate, although the most important dynamic features were not measured. These inaccuracies were overcame.

3. LINEAR REGRESSION. NOTATION AND CONVENTIONS

To find similarity of two signatures the linear regression was applied. The linear regression is a classical statistical problem, where relationship between two random sequences like $X=(x_1,x_2,...,x_n)$ and $Y=(y_1,y_2,...,y_n)$ is searched. The linear regression analyses the distribution between points (x_1,y_1), $(x_2,y_2),\ldots,(x_n,y_n)$ in the Euclidean space. If these both sequences *X* and *Y* have strong linear correlation, then dependences between sequences can be described by means of simple linear equation:

$$y_i = b + ax_i + u_i \tag{1}$$

where:

$$a = \frac{\sum_{i=1}^{n}(x_i - \overline{X})(y_i - \overline{Y})}{\sum_{i=1}^{n}(x_i - \overline{X})^2}$$

$$b = \frac{1}{n}\left[\sum_{i=1}^{n} y_i - a\sum_{i=1}^{n} x_i\right] = \overline{Y} - a\overline{X}$$

and

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i \,, \quad \overline{Y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

It should be noted that in equation (1) the error u_i is defined as $U = (u_1, u_2, ..., u_n)$. Hence, we can estimate the parameters *a* and *b* so that sum of squared-error was minimal. For this reason, sum of $\sum_{i=1}^{n} u_i^2 = \sum_{i=1}^{n} (y_i - b - ax_i)^2$ should be minimized. The quality of fit measure between both sequences *X* and *Y* is well known as *R* factor [6]. The measure *R* indicates the association between the x_i –variable and the y_i –variable. Its absolute value indicates how well the straight line (1) of the best fit approximates the data (Fig. 1).

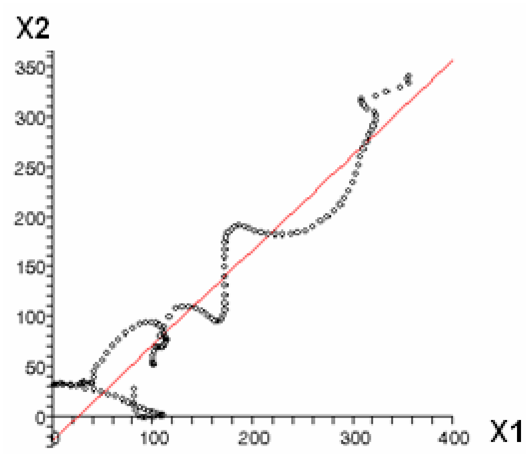

Fig.1. A set of the two signature points which lie along the axes X1 and X2, respectively and linear regression line

In practice, more convenient computations form of the *R* coefficient is [2]:

$$R^2 = \frac{\left[\sum_{i=1}^{n} (x_i - \overline{X})(y_i - \overline{Y})\right]^2}{\sum_{i=1}^{n} (x_i - \overline{X})^2 \sum_{i=1}^{n} (y_i - \overline{Y})^2} \qquad (2)$$

The factor R^2 can be treated as the similarity measure *Sim* between two sequences *X* and *Y*. If we use notation, where $Sim = 1 - R^2$, then values of similarity coefficient will be normalized, hence $Sim \in [0,1]$. If $R^2 = 1$, then both sequences *X*, *Y* have perfect linear correlation. If $R^2 = 0$, then linear relation between sequences do not appear. In other words we obtain similarity with values between 0% - 100%.

In fact, the sequences *X* and *Y* can have different meaning. For example in the signature analysis, the values x_i and y_i create the Euclidean *X–Y* space. Hence, the values x_i and y_i are signature coordinates along the axes *X* and *Y*, respectively.

Usually, signature sequence is neither of the same length nor aligned well even by the same person. For this reason, direct use of the factor R^2 is impossible because number of elements in sequences *X* and *Y* is different. In such case the DTW (Dynamic Time Warping) algorithm can be applied because this technique is able to determine alignment between two sequences with different lengths.

4. THE DTW METHOD. SHORT BACKGROUND

The Dynamic Time Warping (DTW) is a technique that finds the optimal alignment between two sequences if one sequence may be "warped" by stretching or shrinking. This method can be used to find corresponding regions between two sequences or to determine the similarity between two sequences [1,7,10]. The dynamic time warping problem can be described as follows:

The two sequences *X*, and *Y* can be treated as the set of ordered numbers, where cardinality of such datasets can be stated as $n = card(X)$ and $k = card(Y)$. We construct the warp path *W* such, that:

$$W = \{w_1, w_2, ..., w_l\},\ \max(n,k) \le l < n + k \tag{3}$$

where: *l* is the length of the warp path and the m^{th} element of the warp path is defined as $w_m(i, j)$ and *i* and *j* are indexes of elements of the sequences *X*, and *Y*, respectively. To align two sequences using DTW an $n \times k$ costs matrix should be constructed, where elements (*i*,*j*) of the matrix contain the so-called cost values. The cost value is typically computed as the Euclidean distance between the two points x_i and y_j of the sequences *X* and *Y*, respectively:

$$d(x_i, y_j) = \sqrt{(x_i - y_j)^2} \tag{4}$$

Instead of the simple equation (4), the multidimensional Euclidean distance can also be used. For finding the minimum-distance warp path, each cell of the costs matrix must be filled. All cells of the matrix can be filled very efficiently by the recurrence formula:

$$\lambda(i, j) = d(x_i, y_j) + \min\{\lambda(i-1, j-1), \lambda(i-1, j), \lambda(i, j-1)\} \tag{5}$$

where $\lambda(i, j)$ describes the final value of the (*i* ,*j*) cell of the costs matrix.

Hence, cumulative distances in the matrix cell are computed on the basis of the $d(x_i, y_j)$ in the current cell and the minimum of the cumulative distances of the adjacent elements. The cost values of the costs matrix and adjacent elements of the cell with the value 49 are presented in the Fig.2. The same figure presents also final warp path.

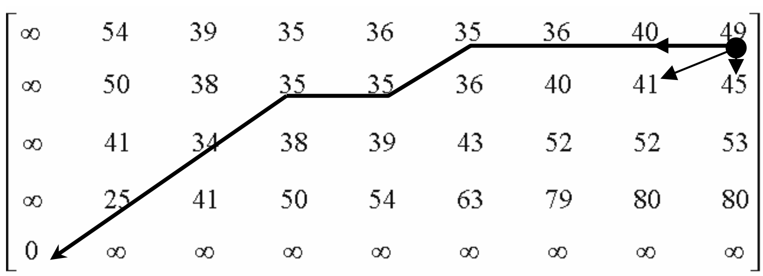

Fig 2. Example of the minimum-distance warp path

The work of the DTW algorithm can be graphically expressed. The Fig. 3a presents the two exemplary signatures. Each of these signatures have different length. The Fig.3b presents the same signatures after the DTW procedure. Now, the two sequences are adjusted and have the same length.

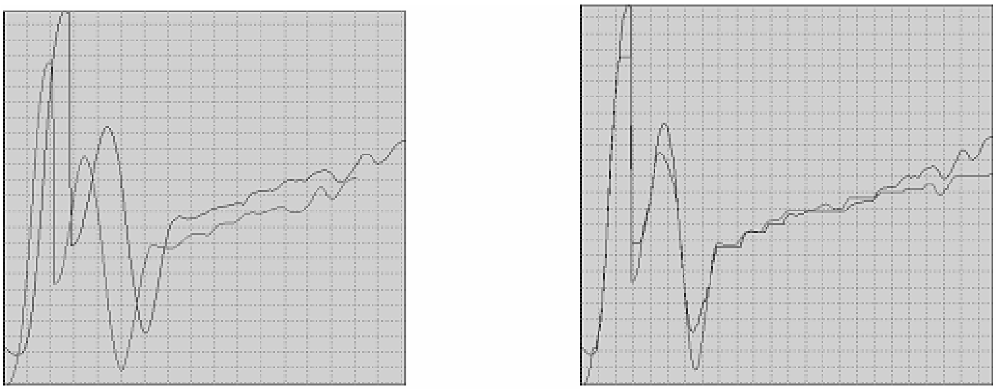

Fig.3. Two signature sequences with different length. Original signatures sequences (left); the same normalized sequences after DTW algorithm (right)

5. RESEARCHES AND OBTAINED RESULTS

Due to the fact that all signatures are stored as textual files (coordinates, values of time and pen-pressure), there is no need to perform any pre-processing (as it is needed for image stored as graphic file). In our database there are four signatures of the same person. In the first step one of these signatures is chosen as a pattern (the most characteristic signature of the person). Each signature (of the same person) is compared to the remaining three signatures (DTW and linear regression are used) and the signature, which is the most similar to the rest signatures is chosen as the pattern.

As it was mentioned above, the static information (coordinates of signature) and dynamic information (time and pen-pressure) are stored in our database. We passed through this database twice and we carried out experiments including pen-pressure factor and excluding that one. We also did some experiments, where time of signature was captured, but the results revealed that time factor was insignificant for the final result of two signatures comparison. We suppose that the factor might be more important during recognizing forged signatures (these investigations will be carried out soon).

The Fig 4 shows results of researches carried out with excluded pen-pressure factor. Results are presented by factors:

False Rejection Rate (FRR) - is stated as the ratio of the number of false rejections (N_{FRR}) divided by the number of total identification attempts T.

False Accept Rate (FAR) - is stated as the ratio of the number of false acceptances (N_{FAR}) divided by the number of total identification attempts T.

Equal Error Rate (EER) – a point where the FAR and FRR intersect (the value of the FAR and the FRR at this point, which is of course the same for both of them).

Compatibility – minimal threshold to consider signature as genuine

Efficiency:

$$Efficiency = \frac{T - (N_{FAR} + N_{FRR})}{T} \times 100\% \qquad (5)$$

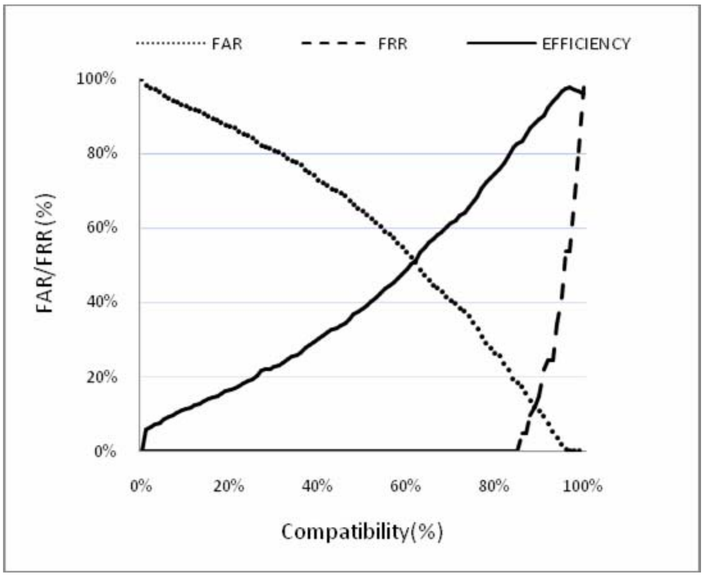

Fig.4. FAR/FRR coefficient distribution for signatures where pen-pressure feature was not measured

The ERR is equal 89% for this investigation (where FRR=FAR=12,2%).

During the next pass of researches the pen-pressure factor was included to DTW algorithm and results are shown on the Fig. 5.

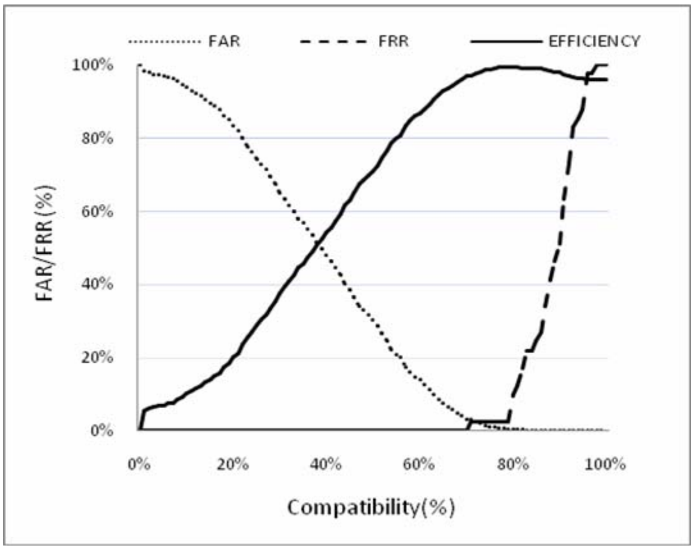

Fig.5. FAR/FRR coefficient distribution for signatures where pen-pressure feature was measured

The ERR is equal 73% for this investigation (FRR=FAR=2,44%).

The pen-pressure factor improved of almost 10% properly recognized signatures. Pen-pressure factor is very sensitive: that is why compatibility threshold might be decreased to 73% for the best results. Comparing to ERR without pen-pressure (89%) the difference is 16%. Decreasing the compatibility to such low level (73%) is very risky. It means that if signatures are similar in 73% to the pattern, then are considered as signatures of the same person. If we try to do this without pen-pressure factor, the results would have a quite high

FAR=37,5%. The investigation with the pen-pressure factor proved that in this case low compatibility threshold is allowed and quite safe.

6. CONCLUSIONS AND FUTURE WORK

In this paper has been presented a handwritten signature recognition system based on both dynamic and static signature features. The system does not need any preprocessing of signature because data of each signature are stored in textual files. It was shown that pen-pressure feature is very meaningful and can improve work of system significantly. On the other hand, pen-pressure is very sensitive and needs quite low compatibility threshold (comparing to static features).

The received results of previous system (which base on the Hough Transform) and current investigations are shown in the Table 1.

Table 1. Comparison of the different methods of the signature recognition

Hough Transform(%)			*R^2 without pen-pressure (%)*			*R^2 with pen-pressure(%)*		
FAR	*FRR*	*Efficiency*	*FAR*	*FRR*	*Efficiency*	*FAR*	*FRR*	*Efficiency*
1,79	*3,57*	*94,60*	*12,1*	*12,3*	*89,00*	*1,2*	*2,4*	*97,48*

As it can be noticed, the results that we have obtained (the Hough Transform) are quite interesting and its effectiveness level is very attractive. Unfortunately, these results are hard to improve without dynamic features. The dynamic features contain many hidden information (e.g. pen- pressure, acceleration, velocity) of signature and these information are much more harder to forge than visual feature (image of signature). It was the reason we included the dynamic features to our investigations and it improved efficiency of our system (Table 1).

During next researches we plan to test other dynamic features (e.g. velocity, acceleration) and to determine set of the most characteristic/important ranges of dynamic features. Moreover, we plan to carry out tests, where forged signatures will also be analysed.

BIBLIOGRAPHY

[1] SALWADOR S., CHAN P., FastDTW: Toward Accurate Dynamic Warping in Linear Time and Space. Proc. of the Int. Conf. on knowledge discovery and data mining – KDD'04, Seattle, USA, pp. 70–80, 2004.

[2] LEI H, et all., ER2: an Intuitive Similarity Measure for On-line Signature Verification. 9th Int. Workshop on Frontiers in Handwriting Recognition – IWFHR'04, Tokyo, Japan, pp. 191–195, 2004.

[3] PORWIK P., PARA T., Some Handwritten Signature Parameters in Biometric Recognition Process. Proc. of the 29th Int. Conf. on Information Technology Interfaces – ITI'07, pp. 185–190. Cavtat, Croatia, 2007.

[4] PORWIK P., The compact three stages method of the signature recognition. Proc of the 6th Int. IEEE Conf. Computer Systems and Industrial Management Applications, CISIM 2007. Ełk, pp. 282–287, 2007.

[5] COETZER J., et all. Offline Signature Verification Using the Discrete Radon Transform and a Hidden Markov Model. EURASIP J. on Applied Signal Processing 2004, pp. 559–57, 2004.

[6] ZHANG B., FU. M, YAN H., Handwritten signature verification based on neural gas based vector quantization. Proc. of 14th Int. Conf. on Pattern Recognition. Vol. 2, pp. 1862 – 1864, 1998.

[7] KORONACKI J., ĆWIK J., Statystyczne systemy uczące się. Wydawnictwa Naukowo-Techniczne, Warszawa, 2005.

[8] SAEED K., ADAMSKI M., Extraction of Global Features for Offline Signature Recognition. Image Analysis, Computer Graphics, Security Systems and Artificial Intelligence Applications, WSFiZ Press, pp. 429–436, 2005.

[9] ADAMSKI M., SAEED K., Signature image recognition by shape context image matching [appear in proceedings of XII Int. Conf. Medical Informatics and Technologies, Osieczany near Cracow, Poland, 2007].

[10] ADAMSKI M., SAEED K., Signature identification by view-based feature extraction and Dynamic Time Warping classifier. Proc. of the 13th Int. MultiConf. on Advanced Computer Systems–ACS–AIBITS/CISIM'06, Miedzyzdroje, Poland, pp. 67–74, 2006.

[11] LEE L., BERGER T., AVICZER E., Reliable On-Line Human Signature Verification Systems. IEEE Trans. on Pattern Analysis and Machine Intelligence , pp. 643–647, 1996.

[12] RHEE T., CHO S., KIM J., On-Line Signature Verification Using Model-Guided Segmentation and Discriminative Feature Selection for Skilled Forgeries. The 6th International Conference on Document Analysis and Recognition (ICDAR), 2001.

[13] TANABE K., YOSHIHARA M., KAMEYA S., MORI S., OMATA S., ITO T., Automatic signature verification based on the dynamic feature of pressure, Document Analysis and Recognition, 2001. Proc. of the sixth Int. Conf. on Volume , Issue , pp. 1045 – 1049, 2001.