A new signature similarity measure

Piotr Porwik, Rafal Doroz and Krzysztof Wrobel
University of Silesia, Institute of Computer Science,
Bedzinska 39, 41-200 Sosnowiec, Poland
{piotr.porwik, rafal.doroz, krzysztof.wrobel}@us.edu.pl

Abstract

The paper presents a new signature similarity measure and new efficient method of recognizing handwritten signatures. Each signature is represented as a set of features such as coordinates of signature points, pen pressure, and speed of writing. Proposed approach consists in dividing signature into windows and calculating similarity values between individual windows. The influence of the size of windows and their location in a signature has been analysed. Additionally, the influence of individual features on the signature similarity value has been examined.

1. Introduction

Biometric techniques are currently among the most dynamically developing areas of science. They prove their usefulness in the era of very high requirements set for security systems. Biometry can be defined as a method of recognition and personal identification based on physical and behavioural features [5,7,8,10]. Physiological biometrics covers data coming directly from a measurement of a part of human body, e.g. a fingerprint, a shape of face, a retina, etc. Behavioural biometrics analyses data obtained on the basis of an activity performed by a given person, e.g. speech, handwritten signature.

Data collection within a signature recognition process can be divided into two categories: static and dynamic. The static system collects data using off-line devices [6,12]. A signature is put on paper, and then is converted into a digital form with the use of a scanner or a digital camera. In this case, the shape of the signature is the only data source, without the possibility of using dynamic data. On the other hand, dynamic systems use on-line devices, which register, apart from the image of the signature, also dynamic data produced during measurement process [10,11].

The most popular on-line devices are graphics tablets. During writing process, tablet can capture many different parameters, such as: pen position and inclination, pen velocity and pressure. In professional devices also time, when pen has contact with surface, can be measured and registered. The position parameter value is given directly by a simple graphics tablet, while speed and acceleration can be obtained from this device or can be calculated on the basis of the position and time parameter. In some cases signatures are digitally captured online and then converted into static images. In our approach only dynamic parameters of a given signature are investigated and signature image is used only as a visual confirmation of recognition.

It will explain more precisely in the next sections of the paper.

The method proposed in this paper consists of a few stages, where rotation of signature, normalization its features, and similarity determination between signatures are carried out. Rotation and standardization of signatures have been described in a lot of works [1,5,6]. Similarity of signatures was determined using many known measures (coefficients) [3,5,13,14].

In the presented study, the modified so-called Czechanowski’s coefficient was applied [4]. The first time this measure was described by Polish scientist in 1909. For the better presentation works in English language are also cited [2,9]. For the first time this measure in anthropological investigations were used and mean differences between objects were measured. Up to now, the usefulness of this coefficient in a signature recognition process has not been studied.

Proposed modification consists of dividing a signature into small windows, where only part of data is gathered. Then, similarity between the windows is determined with the use of mean differences.
Additionally, there has been determined the influence of individual parameters of the method, such as sizes and positions of windows and the weight assigned to individual features, on the signature similarity value. The set of signatures used in the investigations comes from the SVC (Signature Verification Competition). This database is well known in the research community and is available at the website: http://www.cse.ust.hk/svc2004/index.html.

In this article, we focus on online signature verification only. Feature vectors are directly extracted from tablet, but some dynamic parameters can be also computed directly from other features. It is very convenient in a case, when device do not measure appropriate feature – for example velocities or accelerations.

2. Overview of prior work

Numerous methods and approaches have been presented in a lot of survey papers. Earlier work on signature analyze considered only the global features of a signature. Nowadays, signature databases became larger and researchers study more difficult task, where for example, different casual or skilled forgery signatures are analyzed. It can be observed, that not only more elaborate classifiers are applied, but also a new local signature features are measured and the new matching techniques are satisfactorily introduced.

The features that are extracted from static signature images can be classified as the global or local features. Global features describe an entire signature [7,8], and smoothness features [10,11]. Local features are extracted at stroke and include motion and tremor information, pressure or slant features [5,14].

3. Preliminaries

Description of the single signature is represented by a text file containing values of individual features of the signature. The figure 1 presents a sample signature together with values of selected features (x and y coordinates of any signature point, pressure p and measurement time t).

The origin, a continuous image of signature can be described by means of a set of discrete multidimensional vectors $s_i$, $i=1,...,c$ (Fig.1). Any vector $s_i$ has $n$ coordinates, where value $n$ depends on number of parameters which should be analyzed. In example from Fig.1 $n=6$. Hence, signature description consists of $c \times n$ real numbers, stored in a text file.

![Fig.1. Sample of signature and its selected features.](image)

Let $S$ be a given signature, consisting of $s_i$, $i=1,...,c$ points. In proposed approach, it was assumed that any signature point was describe by the next values: point coordinates $x$, $y$, pen pressure $p$ at the point $(x,y)$ and velocities $v_x$, $v_y$ and $v_p$ in its consecutive points. The velocities in selected directions were computed from well known, simple formulas:

$$v_\text{xy} = \frac{x_{i+1} - x_i}{t_{i+1} - t_i}, \quad v_\text{yx} = \frac{y_{i+1} - y_i}{t_{i+1} - t_i}, \quad v_\text{yp} = \frac{p_{i+1} - p_i}{t_{i+1} - t_i}$$

for $i=1,...,c-1$

$$v_\text{xx} = v_\text{xy}_{i-1}, \quad v_\text{yy} = v_\text{yx}_{i-1}, \quad v_\text{vp} = v_\text{yp}_{i-1}, \quad \text{for } i=c$$

Taking into account above considerations signature $S$ from Fig.1 can be described by the set of the multidimensional vectors:

$$S = \{s_1, s_2, ..., s_c\}$$

where:

$$s_i = [x_i^S, y_i^S, p_i^S, v_{x_i}^S, v_{y_i}^S, v_{p_i}^S], \quad \text{for } i=1,...,c$$

is a six-dimensional vector.

For simplification, elements of vector will be called features.

In the next parts of the paper usefulness of the features of the vector $s_i$, in the signature recognition process, will be also stated. Features of the vector $s_i$ are normalised to the range $[0,1]$ according to the formula:

$$\lambda_{\text{norm}} = \frac{\lambda - \lambda_{\text{min}}}{\lambda_{\text{max}} - \lambda_{\text{min}}}$$

(1)

where:

$\lambda$ – value before normalization,

$\lambda_{\text{min}}$ – the minimal value of a given feature in the signature,

$\lambda_{\text{max}}$ – the maximal value of a given feature in the signature.
4. Similarity measure which uses mean differences

In order to determine the similarity between signatures, the new measure based on mean differences has been introduced. Let \( s_i \in S \) be a multidimensional vector of a signature \( S \) features and \( q_j \in Q \) be the vector of the signature \( Q \) features. The signatures \( S \) or \( Q \) can have different dimension. Hence, these signatures can be represented by means of the set of the vectors:

\[
S = \{ s_1, s_2, \ldots, s_n \} \quad \text{and} \quad Q = \{ q_1, q_2, \ldots, q_m \}, \quad \text{respectively.}
\]

Let mean difference between the given vectors \( s_i \) and \( q_j \) be determined as follows:

\[
DD(s_i, q_j) = \frac{1}{n} \left( \sum_{i=1}^{n} \left| x_i^s - x_j^q \right| + \ldots + \left| y_p^s - y_p^q \right| \right) \quad (2)
\]

where:

\( n \) – number of features being compared.

In this case \( DD(s_i, q_j) = DD(q_j, s_i) \).

The formula (2) can be simple re-written in another form. Such a modification allows comparing similarity measure with other popular measures, where value 1 means that objects are identical, and value 0 means that objects are completely dissimilar:

\[
DO(s_i, q_j) = 1 - DD(s_i, q_j) \quad (3)
\]

In order to specify the influence of a given feature on the result of the comparison, the weight of the \( i \)’th feature \( w_i \) was introduced. In the next part of the paper the values of the individual features will be called:

\( w_x \) – weight of \( x \) feature, \( w_y \) – weight of \( y \) feature, \( w_p \) – weight of \( p \) feature, \( w_v \) – weight of \( v_x \) feature, \( w_y \) – weight of \( v_y \) feature, \( w_{vp} \) – weight of \( vp \) feature.

These weights should fulfill the following conditions:

\[
w_x, w_y, w_p, w_v, w_{vp} \in [0, 1] \quad (4)
\]

\[
w_x + w_y + w_p + w_v + w_{vp} + w_{vp} = 1
\]

In proposed approach it was assumed, that weight is the same, for the same feature in signatures \( S \) and \( Q \). Hence, after introduction of the weights, the formula (2) has a new form:

\[
DD(s_i, q_j) = \frac{1}{n} \left( \sum_{i=1}^{n} \left| x_i^s - x_j^q \right| w_x + \ldots + \left| y_p^s - y_p^q \right| w_{vp} \right) \quad (5)
\]

To compute the similarity between two signatures \( S \) and \( Q \), the following equation is introduced:

\[
sim_{DO}(S, Q) = \frac{1}{c} \sum_{i=1}^{c} \max \{DO(s_i, q_{ij}), DO(s_i, q_{ij}), \ldots, DO(s_i, q_{ij})\} \quad (5)
\]

where:

\( S = \{ s_1, s_2, \ldots, s_n \} \) – the set of vectors describing the first signature.

\( Q = \{ q_1, q_2, \ldots, q_m \} \) – the set of vectors describing the second signature.

A single vector (data collection), which describe a one signature point consist of the six features, e.g.

\[
s_i = \{ x_i^s, y_i^s, p_i^s, v_x^s, v_y^s, v_p^s \}, \quad i = 1, \ldots, c, \quad \text{and} \quad q_j = \{ x_j^q, y_j^q, p_j^q, v_x^q, v_y^q, v_p^q \}, \quad j = 1, \ldots, v.
\]

The formula (5) is known as Czekanowski’s coefficient [4].

The similarity measure presented by equation (5) has the following characteristics:

1. \( sim_{DO}(S, Q) \in [0, 1] \). Value 1 means that the signatures are identical.

2. The similarity measure is not symmetrical, which means that \( sim_{DO}(S, Q) \neq sim_{DO}(Q, S). \)

In practice, two different signatures may have points of the similar coordinates, so these signatures can be insufficient different each other. For this reason, similarity measure can be wrongly determined. These troubles can be overcome by proposed in this paper the windowing technique.

5. Split signature into windows

The number of selected signature points will be marked as \( h \). These points are located into a window \( win \).

Any window is identified by two indexes which identified number of window and signature. For example \( win_k^s \) identifies the \( k \)-th window of the signature \( S \).

For any signature, number of windows can be clearly determined. If signature consists of \( c \) points, then \( h = c + h + 1 \). The first window includes the first point of the signature and the \( h \) next points. The second window is shifted right, and starts at the second point of the same signature, and so on for the next signature points. It can be observed that all windows overlapped.

The principle of the windows shifting against the signature background is depicted in Fig. 2.
Fig. 2. The main principles of the windows shifting for the signature $S$. In this example every window includes $h=5$ signature points.

In the next step, the mean difference is calculated for the window of the signatures being compared. In order to calculate the mean difference for the data of the two windows $win^S_i$ and $win^Q_j$, it is necessary to calculate the mean differences between the corresponding data included in the windows (Fig. 3).

For every pair of windows being compared, the mean difference has been calculated from the formula:

$$DD_{ij} = \frac{1}{h} \sum_{k=0}^{h-1} (DD(s_{ik}, q_{jk}) - DD(s_{ik}, q_{jk}))$$ (6)

where:
- $i$ – number of the window in the signature $S$ and $1 \leq i \leq c - h + 1$,
- $j$ – number of the window in the signature $Q$, and $1 \leq j \leq v - h + 1$,
- $h$ – number of signature points in the window.

In addition, the equation (6) was modified by means of two factors which represented:
- standard deviation $\sigma$ between the values of mean differences in the compared windows,
- distance $d$ between the windows being compared.

The mean value does not allow us to explain the discrepancy between individual values $DD$ in the window. Therefore, the result $DW$ of the comparison is additionally conditioned on the value of the standard deviation $\sigma$. It was assumed that if the higher value of standard deviation occurs, then the smaller similarity between the windows can be observed. Moreover, the distance $d$ between windows of signatures being compared was taken into consideration. Our verifier is constructed as follows.

6. A new signature similarity measure

The standard deviation $\sigma$ between values of mean differences in the windows can be calculated from the formula:

$$\sigma_{i,j}^k = \sqrt{\frac{1}{h} \sum_{l=0}^{h-1} (DD(s_{ik}, q_{lk}) - DW_{i,j}^k)^2}$$ (7)

The distance $d(win^S_i, win^Q_j)$ between windows is defined as follows:

$$d(win^S_i, win^Q_j) = 1 - \frac{i}{c-h+1} - \frac{j}{v-h+1}$$ (8)

and

$$d(win^S_i, win^Q_j) = \begin{cases} 0, & d(win^S_i, win^Q_j) < dist \\ d(win^S_i, win^Q_j), & d(win^S_i, win^Q_j) \geq dist \\ \end{cases}$$ (9)

In performed investigations various percentage values of the constant $dist$ were tested. After considering the above modifications, equation (6) has a new form:

$$DW_{i,j}^k = DW_{i,j}^k \cdot (1 - \sigma_{i,j}^k) \cdot d(win^S_i, win^Q_j)$$ (10)

In the next step, a single window in the first signature is compared with every window in the second signature (Fig. 4). From among all comparisons the most similar window is selected. This procedure is repeated for every single window of the first signature. Finally, the similarity measure which considers all the above modifications is calculated by means of the following equation:

$$sim_{S,Q} = \frac{\sum_{i=1}^{c-h+1} \max(DW_{i,1}^1, DW_{i,2}^2, ..., DW_{i,v-h+1}^v)}{c-h+1}$$ (11)

for assumption, that $h < c$ and $h < v$.

For the parameter value $h = 1$, the equation (11) is equal to the equation (5).
Taking into account also the established parameters combination of weights was considered. Times determined. In each measurement, different measure was (4). In our case there are 6 different weights. For this from 0 to 1, with step 0.2, according to the condition into account. Each weight was changed in the range: values of weights were also taken into account. During preformed investigations, parameters were selected from the range: \( h \in \{1,10,20,40\} \), \( \text{dist}(\%) \in \{10,85,90,95,100\} \). In addition, various values of weights were also taken into account. Each weight was changed in the range from 0 to 1, with step 0.2, according to the condition (4). In our case there are 6 different weights. For this assumption, in conducted examinations similarity measure was \( \delta \sum (w_x+w_y+w_y+w_y+w_y+w_y)\geq 1=252 \) times determined. In each measurement, different combination of weights was considered. Taking into account also the established parameters \( h \) and \( \text{dist} \), \( (252\times 4)\times 5=5040 \) combination of parameters was tested. It means that for two signatures, their similarity measure was \( 5040\times 2=10080 \) times calculated. It follows from the fact, that similarity coefficient used in the studies was not symmetrical. The comparison of signatures was performed as the each-with-each. It gave in total \( 50\times 50=2500 \) signatures comparison. In all investigations \( 5040\times 2500=1260000 \) comparisons have been carried out. From analyze of the gathered results follows, that the best signatures recognition level was obtained for the next weights combination: 
\[ w_x=0.2, w_y=0.4, w_p=0.2, w_y=0, w_y=0, w_y=0, w_y=0, w_y=0, w_y=0, w_y=0 \]  and for parameters \( h=20 \) and \( \text{dist}=10 \). Values of the weights parameters result in observation that instead of 6 signature features only 4 features can be used. The weights \( w_y=w_y=0 \). This means that pen velocity in the direction \( y \), and velocity of pen pressure changes can be ignored. Hence, the same result of signature recognition can be obtained by means of reduced number of features. It is very convenient because algorithm time complexity can be decreased. In the next stage of investigations the well known the false rejection rate (FRR), the false acceptance rate (FAR), and the equal error rate (EER) have been used as quality performance measures. The FRR is used for genuine signatures and the FAR for forgery signatures. Because these two factors are inversely related, the EER factor is often reported. Realistic measurement of FRR and FAR is not straightforward because it is hard to obtain unquestioned signature database that would contain comprehensive signature samples. For instance, genuine signatures are generally collected in single session. Then, one part of them is used as training set, and the rest to measure of the FRR factor. In our method, signatures used to train the system were collected in different, long-time sessions. Obtaining forgeries is more difficult, because professional forgeries would be really motivated to break the system. From this reason, two forgery types have been defined: 
- A casual forgery (a simple forgery) is produced when the forger is familiar with the writer’s name, but does not have access to a sample of the actual signature stylistic.
- A skilled forgery is signed by a person who has had access to a genuine signature for practice. In this case the professional forgeries are also possible.

In this paper the random forgeries were also considered, although this type of fraud is very simple to detect. It follows from, that random forgery is signed without having any information about the signature of the person whose signature is forged. Other inconveniences, announced in many works, were different types methodologies of investigations are discussed. For example, in many cases only genuine signatures are analyzed and frauds are not respected in tests [8]. In this paper we analyze this problem more realistic, because genuine as well as forgery signatures have been considered. In order to determine the quality of the method, in the first step, the EER (Equal Error Rate) value was calculated. If the EER value is small, then signatures recognition error is also small. For above mentioned the best parameters the EER=1.29%.

The obtained results show also that making the modification consisting in the use of the windowing

![Fig.4. Example of comparing windows of the signature S and Q, respectively.](image-url)
method has decreased the EER value. The smallest error was obtained for the standard Czekanowski’s method: 6.52%. In this method, the dist parameter is not taken into account, and the window size parameter $k=1$.

When analysing the influence of window sizes on method error, it can be noticed that the smallest EER error was obtained for windows with the size of 10 and 20. For a window with the size of 40, the value of error increased to 2.23%. At the same time, significance of the dist parameter increases along with an increase in the size of a window. In case of windows with the size of 1, 10, and 20, EER is practically independent of the dist value.

8. Final remarks

Comparison of the obtained results with other methods described in the literatures shows a high usefulness of the Czekanowski’s measure in a signature recognition process. The obtained error is relatively small, which is shown in the Table 1. Comparison of the Czekanowski’s coefficient with other similarity measures has been planned for the next stage of the studies.

<table>
<thead>
<tr>
<th>Methods (for genuine signatures)</th>
<th>ERR [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed technique (the best result)</td>
<td>1.29</td>
</tr>
<tr>
<td>Data glove, Kamel et al. (2008)</td>
<td>2.37</td>
</tr>
<tr>
<td>Maramatsu et al. (2003)</td>
<td>2.60</td>
</tr>
<tr>
<td>Kholmatov et al. (2005)</td>
<td>2.80</td>
</tr>
<tr>
<td>Nakashita et al. (2005)</td>
<td>3.30</td>
</tr>
<tr>
<td>Shintaro et al. (2006)</td>
<td>4.10</td>
</tr>
<tr>
<td>Fierez-Aguilar et al. (2005)</td>
<td>7.2</td>
</tr>
</tbody>
</table>

9. References


