

Przemysław KUDŁACIK<sup>1</sup>

## IMPROVING A SIGNATURE RECOGNITION METHOD USING THE FUZZY APPROACH

The paper introduces a significant improvement of the signature recognition method based on characteristic preprocessing of an input data. The original approach transforms an input data into a sorted set of points obtained from intersections of a signature with generated lines going through it's center point. For further analysis the discrete Walsh transform was used. The solution presented in this paper divides points obtained in the preprocessing phase into groups. This step allows the method to preserve more unique features, which positively reflects on the results. Preprocessed data is used to build a fuzzy structure called the fuzzy signature. The method considering a natural imprecision makes the verification system flexible.

### 1. INTRODUCTION

The problem of automatic signature verification was always very important for many researchers [2, 3, 4]. The task is complex, because each signature is unique and even in case of one individual a significant differences can be found. Dissimilarity can be caused by natural processes like aging or disease and other like different positions during writing or even emotions. The fact reflects in many solutions developed through all the years of biometric systems applications. Each solution has it's advantages and disadvantages, which always depends on the environment of application and destined purpose in general.

The approach presented in this paper extends the solution introduced in 2008 by Porwik and Wróbel [10]. The characteristic preprocessing phase of the method was analyzed, improved and used to create a fuzzy verification system [1, 6, 8, 9] (the authors of [10] used the discrete Walsh transform for further processing).

The obtained results after some preliminary test are promising, but the method still needs to be precisely examined in order to find the optimal values of parameters, the structure of the fuzzy signature and the method of signature assessment. Computations were performed using SVC database [11], which is available online.

### 2. THE PREPROCESSING PHASE

A signature can be treated as a set of discrete points  $(x_j, y_j)$  laying on the Cartesian  $X$ - $Y$  plane, where  $j=1, 2, \dots, N$ , which describes piecewise-linear graphical form. The number  $N$  can vary for different signatures.

The solution's first step is calculating a signature's center point  $(\bar{x}, \bar{y})$  as so-called center of gravity, by the following equations [10]:

$$\bar{x} = \frac{1}{N} \sum_{j=1}^N x_j; \quad \bar{y} = \frac{1}{N} \sum_{j=1}^N y_j; \quad (1)$$

---

<sup>1</sup> University of Silesia, Institute of Computer Science, ul. Będzińska 39, 41-200 Sosnowiec, Poland,  
email: przemyslaw.kudlacik@us.edu.pl

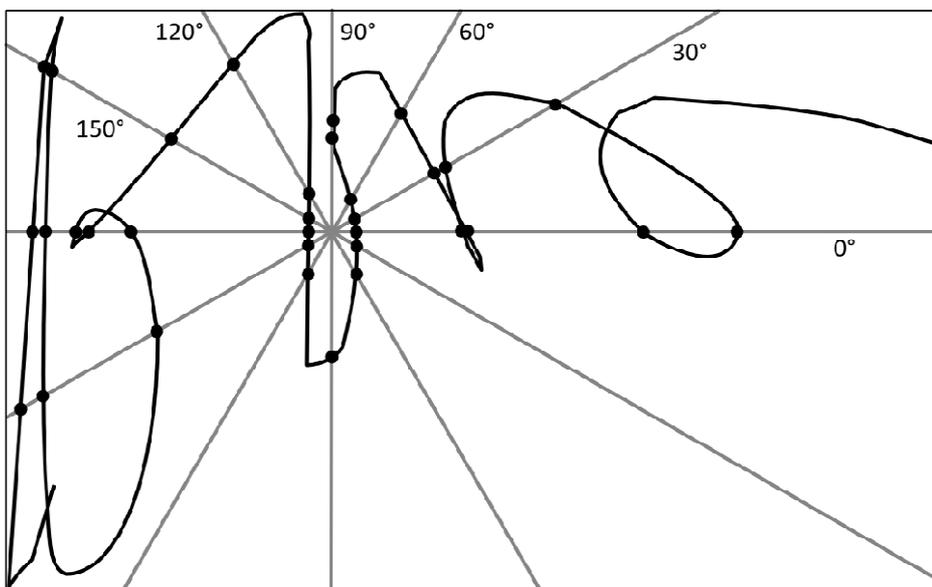


Fig. 1. Intersections of a sample signature with generated lines.

In the next step a new set of points  $(x_k, y_k)$  is obtained, where  $k=1,2,\dots,M$  and  $M \ll N$ . The new points are calculated from an intersection of a signature and lines generated at different angles and passing through the center point. The phase is shown in Fig. 1. The number of generated lines depends on an angle step  $\Delta\alpha$ , which is a parameter of the method. Fig. 1 contains visualization for  $\Delta\alpha = 30^\circ$  and for that reason 6 lines are drawn at  $0^\circ, 30^\circ, 60^\circ, 90^\circ, 120^\circ$  and  $150^\circ$ .

For each point of intersection  $(x_k, y_k)$  the distance from the center point  $d_k$  is calculated [10]:

$$d_k = \sqrt{(x_k - \bar{x})^2 + (y_k - \bar{y})^2} \quad (2)$$

and normalized [10] :

$$l_k = \frac{d_k}{d_{\max}}, \quad d_{\max} = \max\{d_1, d_2, \dots, d_M\} \quad (3)$$

The normalized  $l_k$  values create the  $\Omega_{S_i}$  set, obtained from  $S_i$  signature. The set is an output of the preprocessing phase and a basis for further computation.

### 3. ANALYSIS AND IMPROVEMENTS OF THE PREPROCESSING PHASE

Considering the introduced method of preprocessing it is obvious that the phase produces  $\Omega_{S_i}$  sets differing in size. The fact can be observed in large scale even for signatures of the same person. It can be caused by even small differences of shape and position between the input data. The Fig. 2 depicts such situation, where  $\Omega_{S_i}$  sets ( $i=1,\dots,5$ ) obtained for the first five signatures of SVC database [11] are presented. The number of elements in this case vary from 28 to 34. However, many local similarities can be found. Particularly in the b) part of Fig. 2, where all presented charts are normalized to the same width.

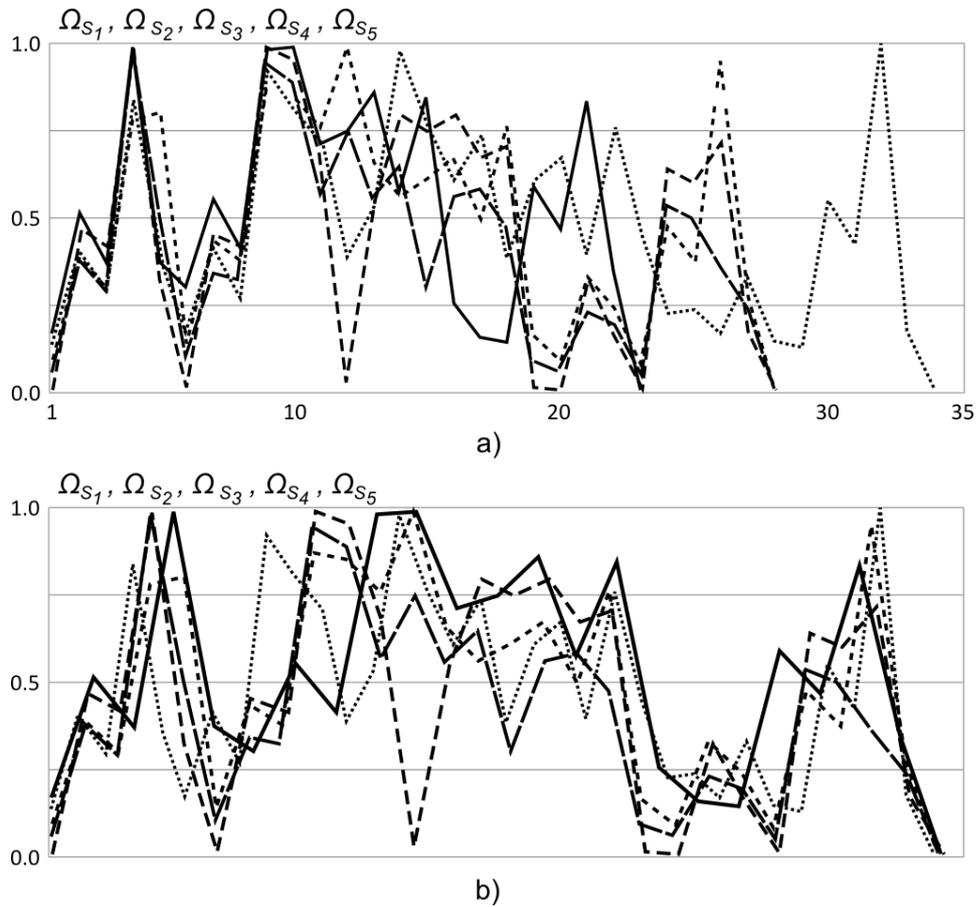


Fig. 2. Sample  $\Omega_{S_i}$  sets: a) charts without modifications, b) charts normalized to the same width.

Another approach analysed in [10] assumes that values within  $\Omega_{S_i}$  sets are sorted in decreasing order. The situation is depicted in Fig. 3. for 5 sample signatures.

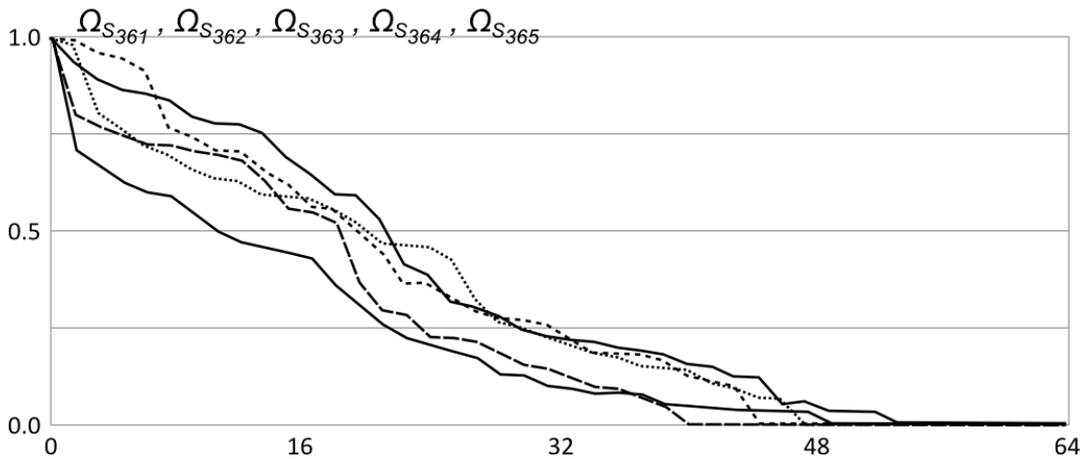


Fig. 3. Charts of sorted  $\Omega_{S_i}$  sets obtained from five first signatures of the 10<sup>th</sup> person in SVC database.

It can be noticed that shapes of functions are similar, which makes the sets easier to compare. If the charts were stretched, like in the example in Fig. 2b, the difference would be even smaller. However, in this case all local similarities existing in non-sorted version of  $\Omega_{S_i}$  sets are lost. Assuming soft methods of samples' comparison for many cases the sets could be hard to distinguish. The problem can be noticed in Fig. 4, where average sets  $\Omega'_{S_i}$  are presented, representing ten different people. Each of the  $\Omega'_{S_i}$  sets,  $i=1, \dots, 10$ , consists of average values calculated from five first signatures of  $i$ -th person.

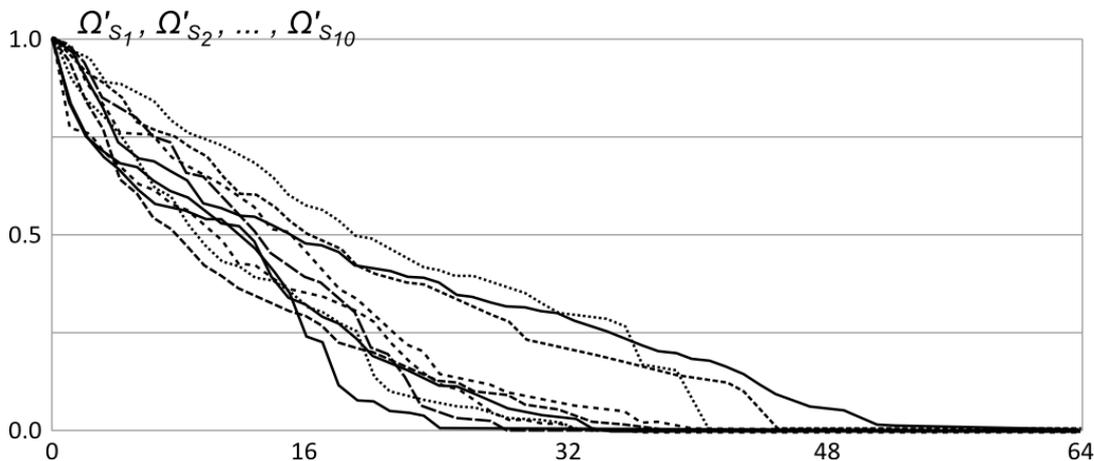


Fig. 4. Charts of  $\Omega'_{S_i}$  sets, obtained from 5 signatures of 10 different people.

After analysis of the problem it was assumed that the designed solution could use the effect of local similarities observed very well for normalized version of not sorted  $\Omega_{S_i}$  sets. On the other hand, sorted  $\Omega_{S_i}$  sets are easier to compare even in cases where they are not normalized. Therefore, it would be convenient if the approach profited from both observations.

The main modification assumes dividing  $\Omega_{S_i}$  sets into smaller  $\Omega_{S_i}^\alpha$  subsets, where  $\alpha = \alpha_1, \alpha_2, \dots, \alpha_G$  represents a particular angle for which the elements of a set were collected. Therefore, for the case with  $\Delta\alpha = 30^\circ$  depicted in Fig. 1,  $\Omega_{S_i}$  will consist of six following subsets:  $\Omega_{S_i}^0, \Omega_{S_i}^{30}, \Omega_{S_i}^{60}, \Omega_{S_i}^{90}, \Omega_{S_i}^{120}, \Omega_{S_i}^{150}$  obtained for  $\alpha = 0, 30, 60, 90, 120$  and  $150$  respectively.

Such modification gives an additional parameter: the size of  $\Omega_{S_i}^\alpha$ . The parameter is crucial to the method, because it provides a simple tool very helpful in finding more and less important areas (angles) of  $\Omega_{S_i}$  for signatures taken from one individual. The possibility of adjustment makes the method flexible and is a very good starting point for developing a solution based on soft computing.

The approach also assumes sorting elements within each  $\Omega_{S_i}^\alpha$  subset in decreasing order. This process is a form of generalization but only within one small subset. Considering described partitioning of  $\Omega_{S_i}$ , this solution does not lose precious information coming from obtaining subsequent cross-points but only prepares data for faster further analysis.

#### 4. THE PROCESS OF BUILDING THE FUZZY SYSTEM

As it was mentioned in previous section, the partitioning of  $\Omega_{S_i}$  produces an additional parameter: a size of each generated  $\Omega_{S_i}^\alpha$ . Let it be described as  $L_{S_i}^\alpha$ , which is obviously a natural number:

$$L_{S_i}^\alpha = \#\Omega_{S_i}^\alpha \in \mathbb{N}. \quad (4)$$

This information can be used by the recognition system in preliminary phase of comparison. The idea uses the fact of similar values of relevant  $L_{S_i}^\alpha$  in different samples obtained from one individual. The values are similar but obviously will differ in some range, which should be considered as the unique property of each person.

Let there be three sets  $\Omega_{S_1}, \Omega_{S_2}, \Omega_{S_3}$  calculated for the same  $\Delta\alpha = 30^\circ$  from  $S_1, S_2, S_3$  samples of signatures obtained from one person. Values of the sets are properly clustered into  $\Omega_{S_i}^\alpha$  subsets and sorted as it was described earlier.

For each  $\Omega_{S_i}^\alpha$  set the  $L_{S_i}^\alpha$  numbers are calculated and grouped into  $M^\alpha$  matrices for the same  $\alpha$  in all samples as shown in the following example:

$$M^0 = \begin{bmatrix} L_{S1}^0 \\ L_{S2}^0 \\ L_{S3}^0 \end{bmatrix}, M^{30} = \begin{bmatrix} L_{S1}^{30} \\ L_{S2}^{30} \\ L_{S3}^{30} \end{bmatrix}, M^{60} = \begin{bmatrix} L_{S1}^{60} \\ L_{S2}^{60} \\ L_{S3}^{60} \end{bmatrix}, M^{90} = \begin{bmatrix} L_{S1}^{90} \\ L_{S2}^{90} \\ L_{S3}^{90} \end{bmatrix}, M^{120} = \begin{bmatrix} L_{S1}^{120} \\ L_{S2}^{120} \\ L_{S3}^{120} \end{bmatrix}, M^{150} = \begin{bmatrix} L_{S1}^{150} \\ L_{S2}^{150} \\ L_{S3}^{150} \end{bmatrix}. \quad (5)$$

Therefore, each  $M^\alpha$  contains an information about changes of  $\Omega_{S_i}^\alpha$  size for different samples of the same individual. It is assumed that the smaller variance of values within  $M^\alpha$ , the more valuable it is and can be considered as an important, unique property of person's signature.

Values within  $M^\alpha$  can be used to create a fuzzy set  $A_{S_\alpha}$  describing a soft constraint for  $L_{S_i}^\alpha$  numbers differing for particular  $\alpha$

$$A_{S_\alpha} = \{(x, \mu_{S_\alpha}(x)) : x \in X_{S_\alpha}\}. \quad (6)$$

Let the membership function  $\mu_{S_\alpha}$  is defined as follows

$$\mu_{S_\alpha}(x_i) = \max\left(\frac{O_{x_i}}{N_L}, \frac{1}{2}\right), \quad x_i \in M^\alpha, \quad (7)$$

where  $O_{x_i}$  represents the number of occurrence of  $x_i$  values within  $M^\alpha$  set and  $N_L = \#M^\alpha$ . The equation (7) promotes multiple occurrence of the same values assuming 0.5 as the possible minimum. Unfortunately, it assigns membership levels only for sizes encountered within  $M^\alpha$ . However, an important advantage of any fuzzy approach is considering values that are near the analyzed area, making the solution more flexible;  $L_{S_i}^\alpha$  values in this case. Described solution takes into account the mentioned problem by accepting additional values considered within the defined  $\beta$  range from maximum and minimum  $L_{S_i}^\alpha$  values, as it is shown in Fig. 5.

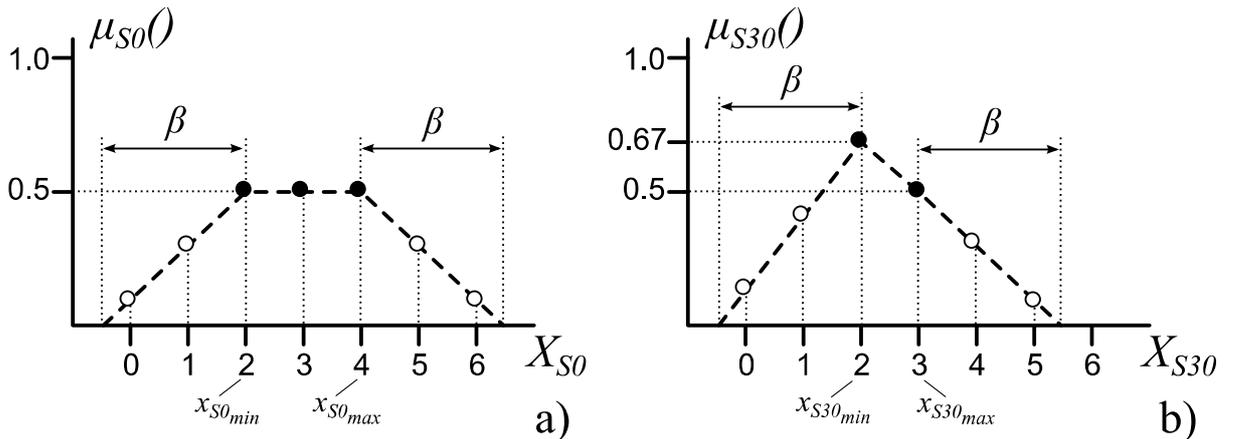


Fig. 5. Sample membership functions of  $A_{S_\alpha}$  fuzzy sets assuming parameter  $\beta = 2.5$  and :

$$\text{a) } M^0 = [2, 4, 3]; \quad \text{b) } M^{30} = [2, 3, 2].$$

The first phase of signature recognition for presented method is based on the structure of  $A_{S_\alpha}$  fuzzy sets. It is the most important element of the system, because the sets carry unique information about signatures, preserving local similarities. Therefore, the first phase allows the recognition system to decrease false acceptances significantly. However, using other information stored within  $\Omega_{S_i}$  can further

improve system parameters. That is why another phase is needed, where  $\Omega_{S_i}^\alpha$  values are used to obtain a second group of fuzzy sets.

Let the three sets  $\Omega_{S_1}, \Omega_{S_2}, \Omega_{S_3}$  again represent the output of the preprocessing phase obtained for the same  $\Delta\alpha = 30^\circ$  from  $S_1, S_2, S_3$  samples of the same individual. Let the  $K^\alpha$  represent a matrix storing values of  $\Omega_{S_i}^\alpha$  sets for all samples, like presented in (8):

$$K^0 = \begin{bmatrix} k_{S_1}^{0:1} & k_{S_1}^{0:2} & \dots & k_{S_1}^{0:L0} \\ k_{S_2}^{0:1} & k_{S_2}^{0:2} & \dots & k_{S_2}^{0:L0} \\ k_{S_3}^{0:1} & k_{S_3}^{0:2} & \dots & k_{S_3}^{0:L0} \end{bmatrix}, K^{30} = \begin{bmatrix} k_{S_1}^{30:1} & k_{S_1}^{30:2} & \dots & k_{S_1}^{30:L30} \\ k_{S_2}^{30:1} & k_{S_2}^{30:2} & \dots & k_{S_2}^{30:L30} \\ k_{S_3}^{30:1} & k_{S_3}^{30:2} & \dots & k_{S_3}^{30:L30} \end{bmatrix}, K^{60} = \begin{bmatrix} k_{S_1}^{60:1} & k_{S_1}^{60:2} & \dots & k_{S_1}^{60:L60} \\ k_{S_2}^{60:1} & k_{S_2}^{60:2} & \dots & k_{S_2}^{60:L60} \\ k_{S_3}^{60:1} & k_{S_3}^{60:2} & \dots & k_{S_3}^{60:L60} \end{bmatrix}, \dots, \quad (8)$$

where  $k_{S_i}^{\alpha:j}$  represent  $j$ -th value of  $\Omega_{S_i}^\alpha$  set (symbol  $\alpha:j$  uniquely describes each column in all  $K^\alpha$  matrices).

It is important to emphasize, that all  $\Omega_{S_i}^\alpha$  sets are normalized to the same size within each  $K^\alpha$  matrix, indicated in (7) by  $L\alpha$ . The parameter is defined by the largest size of  $\Omega_{S_i}^\alpha$  for particular  $\alpha$ . All remaining  $\Omega_{S_i}^\alpha$  sets in  $K^\alpha$  matrix are extended, where 0 values are inserted at missing positions. All sets are in decreasing order and 0 is the smallest value possible.

Analogously to the process of obtaining  $A_{S_\alpha}$ , relevant values within  $K^\alpha$  matrix can be used to calculate another group of  $A_{\alpha:j}$  fuzzy sets:

$$A_{\alpha:j} = \{(x, \mu_{\alpha:j}(x)) : x \in X_{\alpha:j}\}. \quad (9)$$

This time their membership functions are Gaussian and are defined by the following equation

$$\mu_{\alpha:j}(x; m_{\alpha:j}, \sigma_{\alpha:j}) = e^{-\frac{(x-m_{\alpha:j})^2}{2\sigma_{\alpha:j}^2}}, x \in X_{\alpha:j}. \quad (10)$$

Indexes suggest a direct connection between  $A_{\alpha:j}$  sets and  $k_{S_i}^{\alpha:j}$  values, because each  $A_{\alpha:j}$  set is based on one column of  $K^\alpha$  matrix (relevant values in all learning samples). Parameters used in (10) are defined as follows:

$m_{\alpha:j}$  – arithmetic mean of  $k_{S_i}^{\alpha:j}$  values in  $\alpha:j$  column of  $K^\alpha$  matrix

$\sigma_{\alpha:j}$  – range between  $k_{S_i}^{\alpha:j}$  values in  $\alpha:j$  column of  $K^\alpha$  matrix obtained by (11)

$$\sigma_{\alpha:j} = \begin{cases} \gamma \frac{k_{\max}^{\alpha:j} - k_{\min}^{\alpha:j}}{2} & ; \sigma_{\alpha:j} \geq \sigma_{\min} \\ \sigma_{\min} & ; \sigma_{\alpha:j} < \sigma_{\min} \end{cases}, \quad (11)$$

where  $k_{\max}^{\alpha:j}$  and  $k_{\min}^{\alpha:j}$  are respectively maximum and minimum values of all  $k_{S_i}^{\alpha:j}$  for given  $\alpha:j$  (one column of  $K^\alpha$ ). The fuzzyfication ratio  $\gamma$  is a global parameter of the system and allows the user to influence all fuzzy sets by increasing or decreasing their width (for  $0 < \gamma < 1$  or  $\gamma > 1$  respectively). The minimum width  $\sigma_{\min}$  is another global parameter and limits calculated ranges from below in case they are too small.

Two groups of fuzzy sets, defined by (6) and (9), represent the fuzzy model of individual's signature. The model is used in recognition phase described in the following section.

## 5. RECOGNITION PHASE

The database of the fuzzy system stores the fuzzy models based on signatures collected during the enrolment phase. Obviously, each fuzzy model, called a fuzzy signature, corresponds with only one individual. Number of samples collected during the enrolment phase is not constant. It is assumed, that number of samples should equal at least 3 and shouldn't be greater than 10. Small number of samples usually makes fuzzy models too precise. On the other hand, large size of a learning set can have a negative impact on generalization of the model. Tests for described approach considered 5 samples in each learning set.

The main idea of implemented recognition is based on matching input signatures with fuzzy models stored within the database. During the process a level of conformity is calculated for each fuzzy model and tested signature, which is denoted by  $S_{test}$ . If the obtained level exceeds configured trigger value, the signature is considered as matched with this particular fuzzy signature and therefore, it is matched with particular person enrolled in the database.

The recognition process itself also consists of several phases. The first one confronts sizes of  $\Omega_{S_{test}}^\alpha$  subsets of tested sample with relevant  $A_{S_\alpha}$  fuzzy sets within the model. At this phase partial results  $R_{S_\alpha}$  are obtained for each fuzzy set

$$R_{S_\alpha} = \mu_{S_\alpha}(\#\Omega_{S_{test}}^\alpha). \quad (12)$$

Next, values of each  $\Omega_{S_{test}}^\alpha$  subset are matched with relevant  $A_{\alpha:j}$  fuzzy set, which produces groups of  $R_{\alpha:i}$  results

$$R_{\alpha:i} = \mu_{\alpha:i}(k_{S_{test}}^{\alpha:i}). \quad (13)$$

The last phase aggregates partial results of comparison. Obtained value represents the output level of conformity  $R$ . Assuming any triangular norm  $*_T$  as an aggregation operator, the final result  $R$  can be obtained by the following equation

$$R = \underset{\alpha=0}{\overset{\alpha_{\max}}{*}_T} \left\{ R_{S_\alpha} *_T \left\{ \underset{i=0}{\overset{n_\alpha}{*_T}} R_{\alpha:i} \right\} \right\}, \quad (14)$$

where  $\alpha_{\max}$  represents maximum degree existing within the model, while  $n_\alpha$  represents smaller number of either  $k_{S_{test}}^{\alpha:i}$  values or  $A_{\alpha:j}$  fuzzy sets for particular  $\alpha$ , because the numbers obviously don't have to be equal.

Assuming an operation of multiplication applied as T-norm in (14), the equation will take the following simple form

$$R = \underset{\alpha=0}{\overset{\alpha_{\max}}{\prod}} \left\{ R_{S_\alpha} \cdot \left\{ \underset{i=0}{\overset{n_\alpha}{\prod}} R_{\alpha:i} \right\} \right\}. \quad (15)$$

Accepting or rejecting the tested sample as matching the analyzed fuzzy signature is based on calculated level of conformity  $R$ , which can be described by simple equation

$$R \geq Tr. \quad (16)$$

The trigger level  $Tr$  is fixed in presented approach during analysis of all models stored within the database. However, future research will consider a solution with dynamic trigger, which is adjusted separately for each fuzzy signature. The problems of examinations and obtained results are described in detail in subsequent section.

## 6. RESULTS OBTAINED

Presented solution was implemented in Java programming language using the FUZZLIB library [5,6]. Like for the referenced work [10], tests were performed on SVC [11] database. A great advantage of this project is open online access. Researchers can use the same source of data and, what is very important, compare their results. The database contains 40 signatures of 40 different individuals. First 20 signatures of each person are original and remaining 20 are professionally forged.

The main goal of this work was to improve results obtained by [10] using the same interesting preprocessing phase. That is why the tests mainly focused on verification if the system based on presented method can be practically applied. However, more extensive examinations are planned for future research, which consider influence of all presented parameters of the system:  $\Delta\alpha$ , type of T-norms,  $\beta$ ,  $\gamma$ , dynamic trigger level  $Tr$  and number of learning samples.

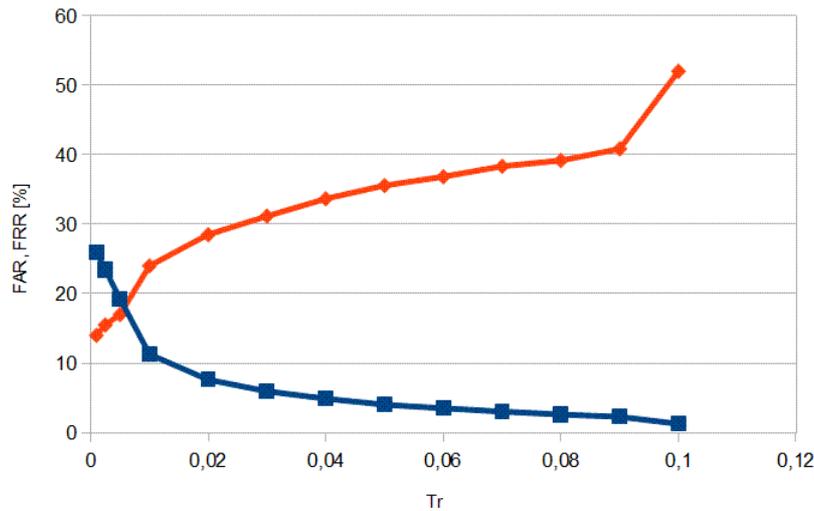


Fig. 6. FAR and FRR coefficients depending on the trigger level  $Tr$ .

Fig. 6 depicts dependence of FAR (False Acceptance Ratio) and FRR (False Rejection Ratio) ratios on the trigger level. Other parameters were arbitrarily set to the following values:  $\Delta\alpha = 30^\circ$ ,  $\beta = 2,5$ ,  $\gamma = 1,5$ . Number of learning samples was set to 5 and function minimum was chosen as T-norm operation.

## 7. CONCLUSION

Results of performed examinations show that the presented fuzzy approach is characterized by a relatively high level of FRR ratio. This reveals that the fuzzy model and proposed method of verification makes the solution too precise. Nevertheless, it shows that the original preprocessing phase makes the samples distinguishable enough.

Fortunately, the approach is very flexible. Several parameters can be adjusted to influence fuzziness of the model. Moreover, other methods can be used in verification phase. This methods can be based on operators like S-norms and means.

The objectives established for the project were achieved. The method certainly shows the potential of the approach with described preprocessing phase. Such solution has several important advantages like lower computational complexity and safety of signature data stored within the database. However, further research are needed to make the designed approach competitive with the best methods available.

**BIBLIOGRAPHY**

- [1] CZOGAŁA E., ŁĘSKI J., Fuzzy and Neuro-Fuzzy Intelligent Systems, Physica-Verlag, Springer-Verlag Comp, 2000.
- [2] DOROZ R., PORWIK P., PARA T., WRÓBEL K., Dynamic signature recognition based on velocity changes of some features, *Int. J. Biometrics*, 2008, Vol. 1, No. 1, pp. 47-62.
- [3] HUANG K., YAN H., On-line signature verification based on geometric feature extraction and neural network classification, *Pattern Recognition*, 1997, Vol. 30, No. 1, pp. 9-17.
- [4] JAIN A.K., FLYNN P., ROSS A., *Handbook of Biometrics*, Springer Verlag New York inc., Secaucus, New York, USA, 2007.
- [5] KUDŁACIK P., Operations on fuzzy sets with piecewise-linear membership function, *Studia Informatica*, 2008, Vol. 29 No. 3A(78), pp. 91-111.
- [6] KUDŁACIK P., Structure of a knowledge base in the FUZZLIB library, *Studia Informatica*, 2010, Vol. 31, No. 2A(89), pp. 469-478.
- [7] KUDŁACIK P., Advantages of an Approximate Reasoning Based on a Fuzzy Truth Value, *Journal of Medical Informatics and Technologies*, 2010, Vol. 15, pp. 57-61.
- [8] ŁĘSKI J., *Neuro-Fuzzy Systems*, Wydawnictwa Naukowo-Techniczne (WNT), Warsaw, 2008.
- [9] MAMDANI E.H., ASSILAN S., An Experiment in Linguistic Synthesis with a Fuzzy Logic Controller, *International Journal of Man-Machine Studies*, 1975, Vol. 20, pp. 1-13.
- [10] PORWIK P., WRÓBEL K., Signature preprocessing based on Walsh coefficients, *Journal of Medical Informatics and Technologies*, 2008, Vol.12, pp. 57-61.
- [11] <http://www.cse.ust.hk/svc2004/>

