signature recognition,
window method,
fuzzy sets

Przemysław KUDŁACIK[1], Rafał DOROZ[1]

DETERMINING VALUABLE RANGES OF HANDWRITTEN SIGNATURE USING FUZZY APPROACH AND WINDOW METHOD

The paper proposes possible improvements in signature recognition approach based on window method. The analysis focuses on a stage of window preprocessing using fuzzy sets in order to choose significant ranges of each signature. Proposed extension allows the solution to improve in two areas. First of all minimizing a number of processed windows significantly reduces computation time. Secondly, filtered signatures with valuable information about significant ranges allow the system to recognize signatures of a poor or good quality. Developed method of signature quality assessment can be used in any signature recognition system, regardless of used method of analysis. Merging the information about signature quality and choosing only important signature ranges should also improve the overall detection results, however, more examinations are needed to confirm this statement.

1. SIGNATURE RECOGNITION BASED ON THE WINDOW METHOD

The window method of signature recognition was firstly introduced in [7] and since then developed [2], [6], [8]. The general idea is based on treating each signature as a set of points in two-dimensional plane. The best way to obtain signatures in such form is using a specialized tablet, however, it is also possible to convert a signature stored in a graphical form into a set of points.

The main concept of the window method is analyzing signature samples in chunks to find similarities between defined subsets of given samples, which are called windows. The name windows directly correspond with a graphical representation of analyzed subsets, which can be presented in a form of a frame, as it is shown in Fig. 1.

Formally the signature representation can be defined by the following equation

$$S = \{S_1, S_2, \cdots, S_n\}, \tag{1}$$

where

$$S_i = \{x_i, y_i : x_i, y_i \in \mathbb{R}\},\ i = [1, \cdots, n], \tag{2}$$

while i-th window of S signature containing $k+1$ elements can be defined as follows

$$Win_i = \{S_i, S_{i+1}, \cdots, S_{i+k}\}. \tag{3}$$

[1]University of Silesia, Institute of Computer Science, Będzińska 39, 41-200 Sosnowiec, Poland

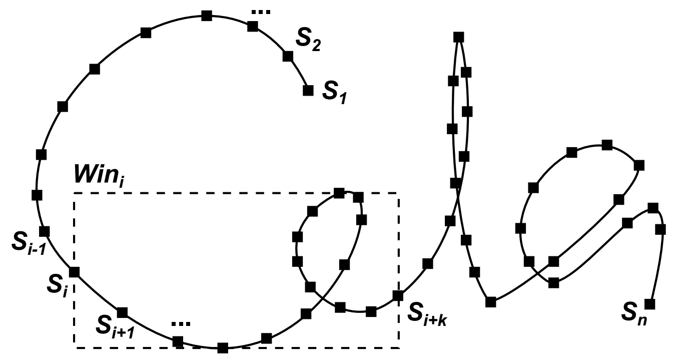

Fig. 1. Sample signature S with marked window Win_i consisting of $k+1$ points.

Assuming that $n \geq (k+1)$, the number of Win_i windows that can be defined within the S signature equals $n-k$.

Details of the window method considering all necessary stages of signature recognition are extensively described in [2], [6], [8] and will not be addressed in this paper. The aim of this work is to focus on the preprocessing phase providing more sophisticated signature analysis, where samples are pruned off the invaluable ranges, which will improve computation time of further analysis and may positively reflect on efficiency of the window method itself. The second advantage of the phase is obtaining the information describing an overall signature quality in terms of recognition and can be used to asses samples given by a particular individual.

Subsequent sections focus on problem analysis, where drawbacks of the original window method are described and a solution is proposed. The main concept of modification is taking advantage of uncertainty provided by fuzzy sets [9], [1]. The extension allows to analyze signatures finding the most valuable ranges in terms of recognition. In the aspect of uncertain environment of handwritten signature the solutions employing fuzzy sets proved to be useful [4], [3].

1.1. DATABASE

The database used in examinations, as the original research [2], consists of signatures collected for 100 people [5]. Each person provided 10 samples, therefore, the total number of samples equals 1000. Signatures were collected using a specialized tablet storing a subsequent x,y coordinates of signature points. This method is flexible enough to allow researchers examine solutions treating a signature as a set of points or as a linear version after converting to the piece-wise linear form.

As it was described in the previous section, the window method used as the base approach in this paper, analyses each signature as a set of subsequent points.

2. PROBLEM OF WINDOW SIMILARITY

Within first stages of signatures' analysis by the window method the approach calculates window similarity of two given samples S^A and S^B, where the first one, marked as S^A, is being compared with the other, marked as S^B. In this way each analyzed window of signature S^A is paired with the most similar in signature S^B. The procedure is repeated for all windows of S^A. This data is further used in the recognition process, which is not important for the analysis in this paper.

Considering that two signatures, S^A and S^B, are taken from the same individual, it seems natural that a given window $Win_i^A = \{S_i^A, S_{i+1}^A, \ldots, S_{i+k}^A, \}$, representing some range of signature S^A, should be characterized by the highest level of similarity with windows of the S^B signature only within a similar range. Therefore, the range of S^B can be described as

$$Win^B_{i\pm u} = \{S^B_{i\pm u}, S^B_{i+1\pm u}, \ldots, S^B_{i+k\pm u},\}, \tag{4}$$

where u represents an integer representing a relatively small shift - depending on the total number of points within a signature.

However, in many cases a window representing a given range of the S^A signature gains the highest levels of similarity for not relevant ranges in S^B. The first reason of such behavior is involved with to many differences between samples - where overall similarity is poor. The second reason is involved with multiple occurrence of similar periods within the same signature.

Therefore, comparing signatures taken from the same individual using the first phase of the window method would reveal problematic signature ranges. The author believes, that the overall result of the window method could be improved when only the most valuable ranges of signatures would be taken into account. On the other hand, in case of too small number of valuable ranges, signatures of an individual could be treated as not sufficient or reliable method of verification.

3. FUZZY SIGNATURE FILTERING

As it was mentioned earlier, the original window method analyses only the most similar windows in compared signatures. To obtain more objective view on the similarity the developed method analyses not one but 10 most similar windows. Therefore, considering two samples obtained from the same individual, S^A and S^B, each window of S^A is paired with a group of 10 most similar windows in S^B. The similarity level denoted by sim is normalized and defined in the $[0, 1]$ range ($sim \in [0, 1]$). Because all signatures can differ in number of description points (1), instead of window number of S^B the original method returns the similarity level for normalized window position denoted by pos, also in the $[0, 1]$ range [7] (0 denotes the beginning of a signature and 1 means the end). Therefore, let the similarity level be defined by the following expression

$$sim^{A_i}_{B_j} = Sim^A_i(pos^B_j), \tag{5}$$

where Sim^A_i represents a function returning the similarity with i-th window in A signature, in this case the similarity for j-th window of B signature.

The database used in examinations [5] contains 10 signatures stored for each individual. Let these 10 signatures of sample person be denoted as $S^A, S^B, \cdots, S^J$. Therefore, each window of S^A can be assigned with 10 most similar windows in each of 9 remaining signatures from S^B to S^J. This results in 90 pairs describing position and similarity $(pos^X_j, sim^{A_i}_{X_j})$, where $X \in (B, C, \cdots, J)$. To simplify further description let this set of 90 pairs obtained for A_i window be denoted as

$$SD_{A_i} = (pos_j, sim_j),\ j \in [1, 2, \ldots, 90]\,. \tag{6}$$

Described set of pairs can be presented as similarity distribution for A_i window of S^A signature. An example of such distribution for sample window is presented in Fig. 2. The original values are shown in form of black squares connected with black thin lines (SD_{A_i}). The thick lightgray line show softened result (SD'_{A_i}) and thick darkgray line presents normalized chart - with minimum value equal 0 and maximum value equal 1 (SD''_{A_i}).

The softened version denoted by SD'_{A_i} is obtained using arithmetic average. Each value of the original node, which is sim_j, is converted into an arithmetic average denoted by sim'_j and

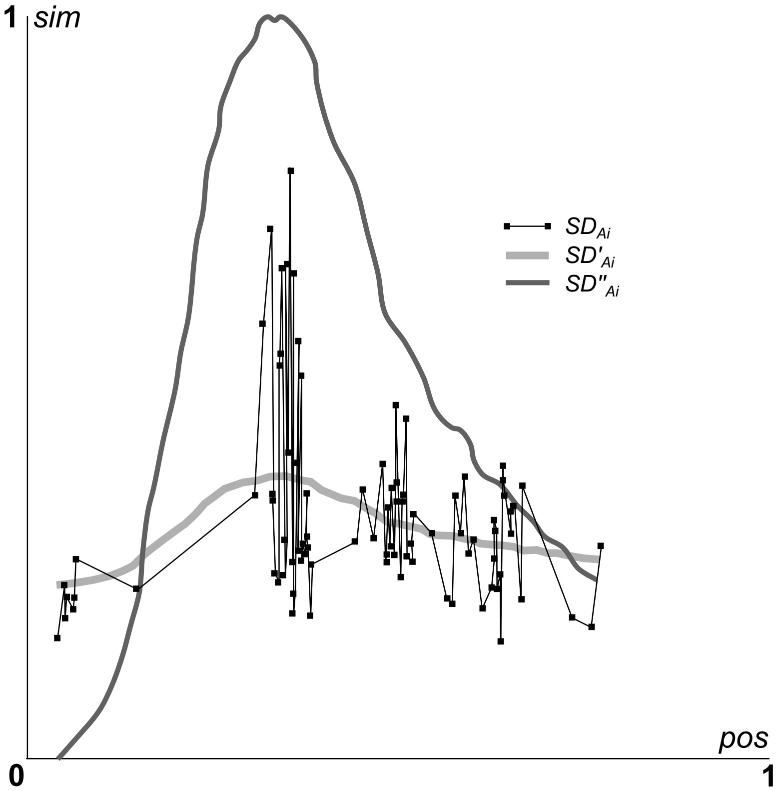

Fig. 2. Fuzzy set obtained from similarity distribution of one sample window compared with 9 other signatures of the same person. Black squares connected with thin black lines represent original values (SD_{A_i}). The thick lightgray line represents softened version of the original chart using arithmetic average (SD'_{A_i}). The thick darkgray line is the softened version after normalization (SD''_{A_i}).

calculated within -0.1 and $+0.1$ range from the node position as follows

$$sim'_j = \frac{\sum_{k=m}^{n} sim_k}{n-m+1}, \tag{7}$$

where $k \in [m, m+1, \cdots, n]$ represent positions for which

$$\forall_{k\in[m,m+1,\cdots,n]} \, |pos_j - pos_k| \leq 0.1\,. \tag{8}$$

Therefore, the SD'_{A_i} can be defined as follows

$$SD'_{A_i} = (pos_j, sim'_j), \; j \in [1, 2, \ldots, 90]\,. \tag{9}$$

To obtain softened chart shown in Fig. 2 the procedure was performed twice.

The most important chart shown in Fig. 2 is presented with thick darkgray line (SD''_{A_i}). It shows normalized values, denoted by sim''_j, directly obtained from the softened version of the original (sim'_j), which can be described by the following equation

$$sim''_j = \frac{\left(sim'_j - min\right)}{max - min} \tag{10}$$

where min and max represent the minimum and maximum values of sim'_j, $j \in [1, 2, \ldots, 90]$ respectively.

This output chart is the most valuable because it can be interpreted as a fuzzy set representing the possibility of A_i position in the context of the best similarity (computed on the basis of ten signatures from the same individual). The main concept of the approach is based on this fuzzy set, which for the purposes of description is denoted as F_{A_i}.

The membership function of F_{A_i} set, denoted as μ_{A_i} can be equal to SD''_{A_i}, which is a discrete function. For the purposes of further description let the μ_{A_i} be defined by the (pos_j, sim''_j) pairs of SD''_{A_i} set, therefore,

$$\mu_{A_i}(pos_j) = sim''_j,\ j \in [1, 2, \ldots, 90]\,. \tag{11}$$

Continuous version of μ_{a_i} can be obtained by using piece-wise linear interpolation of SD''_{A_i} nodes.

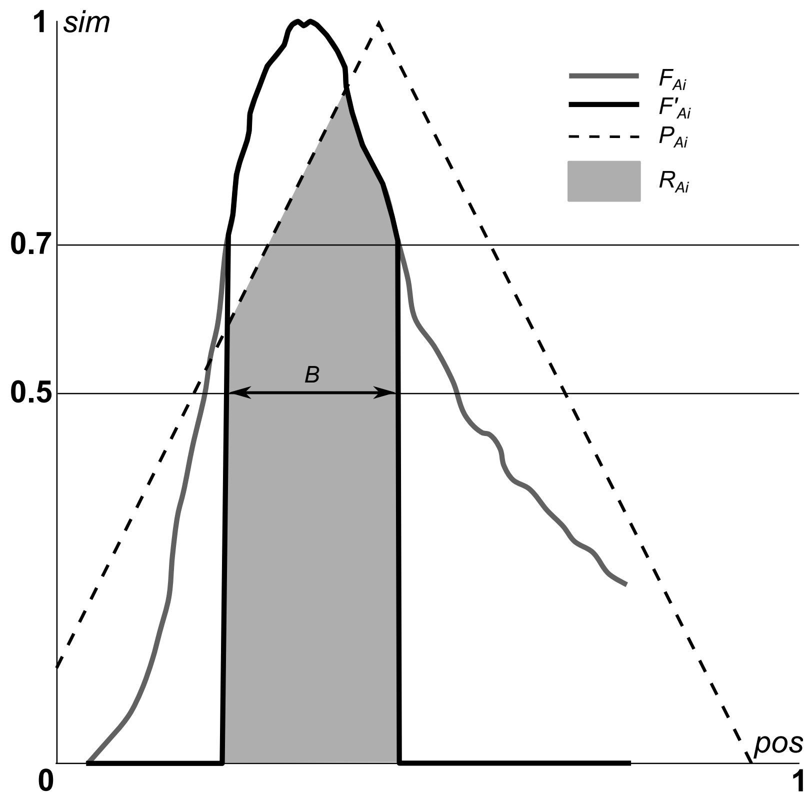

Fig. 3. Fuzzy set F_{A_i} and its modifications.

Fig. 3 depicts further steps of obtaining the final result in fuzzy analysis of the problem. First of all the original fuzzy set F_{A_i}, obtained from SD''_{A_i} presented earlier in Fig. 2, is modified by zeroing all values lower than specified level l. The process generates a new set, denoted by F'_{A_i} with μ'_{A_i} membership function, which is defined as follows

$$\forall_{j\in[1,2,\ldots,90]}\ \mu'_{A_i}(pos_j) = \begin{cases} \mu_{A_i}(pos_j) & , \mu_{A_i}(pos_j) \geq l \\ 0 & , \mu_{A_i}(pos_j) < l \end{cases}. \tag{12}$$

In the case presented in Fig. 3 the parameter was arbitrarily chosen as $l = 0.7$. The concept of this operation is to leave only the highest values representing the most relevant ranges. Therefore, in the opinion of the author, the parameter l should not be set below 0.5. On the other hand, the l value can not be set too high as well in order to preserve information about the character of the F_{A_i} set. In author's opinion the level $l = 0.8$ should not be exceeded. The result of described transformation is depicted in Fig. 3 as F'_{A_i}.

The output F'_{A_i} set can be used to assess the A_i window in context of potential recognition. There are two important aspects in this area. Firstly, the position (pos) of highest values of μ'_{A_i} indicates range or ranges in which the A_i window is matched best. Secondly, the large core of the F'_{A_i} set (large size of ranges) indicates that too many positions were selected. This information allows to compare the result with normalized position of the A_i window and check

whether they correspond to each other.
In case of a very good match the area where $\mu'_{A_i} > 0$ should contain the position of A_i, denoted by pos_{A_i}. On the other hand, pos_{A_i} inside the area where $\mu'_{A_i} = 0$ should indicate worse compatibility.
In order to eliminate ranges found far from pos_{A_i}, the output F'_{A_i} set is compared with the additional P_{A_i} set, whose triangular membership function μ_{PA_i} is directly generated from pos_{A_i} as follows

$$\mu_{PA_i}(x) = \begin{cases} 0 & , x < pos_{A_i} - \frac{1}{\lambda} \\ \lambda\,(x - pos_{A_i}) + 1 & , pos_{A_i} - \frac{1}{\lambda} \leq x \leq pos_{A_i} \\ \lambda\,(pos_{A_i} - x) + 1 & , pos_{A_i} \leq x \leq pos_{A_i} - \frac{1}{\lambda} \\ 0 & , x > \frac{1}{\lambda} - pos_{A_i} \end{cases} \quad , x \in [0,1]. \tag{13}$$

For the case presented in Fig. 3 the $\lambda = 2$ and $pos_{A_i} = 0.43$.

Described μ_{PA_i} is depicted in Fig. 3 with a dashed line. Common area underneath μ_{PA_i} and μ'_{A_i} is marked as gray. It is obtained as an intersection of F'_{A_i} and P_{A_i} fuzzy sets [9], [1]. In this example intersection was based on the minimum function. Let the resulting fuzzy set after intersection be denoted as R_{A_i}.

To simplify final analysis the method compares two values: bandwidth of R_{A_i} and F'_{A_i} sets [1]. Bandwidth was depicted in Fig. 3 as B - in this case it is the same value for R_{A_i} and F'_{A_i} sets.

If bandwidth of R_{A_i}, denoted by B_R, is small it could mean two things: either the range of the most similar windows is small, or it could be the result of reduced area after intersection with the triangular P_{A_i} set. Therefore, the bandwidth of F'_{A_i}, denoted by $B_{F'}$ helps to recognize the situation when significant area of the set is reduced by the intersection.
The best results would be then characterized by relatively small B_R (around 0.3 and less) and small difference $B_{F'} - B_R$. However, B_R must exist ($B_R > 0$). Therefore, the final assessment of A_i window quality can be based on the following parameter α_{A_i}

$$\alpha_{A_i} = \begin{cases} 1 & , B_R = 0 \\ max(B_R, \frac{B_{F'} - B_R}{B_{F'}}) & , B_R > 0 \end{cases} \quad , x \in [0,1]. \tag{14}$$

Therefore, $alpha_{A_i} \in (0, 1]$ and the lower value the better quality of A_i window. It is important to notice that the parameter α_{A_i} will never be equal 0. Higher levels indicate either the larger difference between bandwidths or larger values of B_R itself.

This final result allows to filter windows of given signature focusing only on those ranges of windows for which the parameter α is not higher then specified value.

4. RESULTS

The method of window assessment described in previous section allows the user to tune the mechanism by adjusting two parameters: γ, defining triangular fuzzy set P and l, representing the level of noise reduction. For the purposes of examinations these parameters were arbitrarily set to previously described levels: 2 and 0.7 respectively.

To verify presented analysis the tests were performed on the subset of signatures in the database [5] obtained from 5 individuals. The aim of examinations was to notice the decreasing number of filtered windows according to decreasing level of accepted α. The obtained results are presented in chart in Fig. 4.

The results show that the method correctly chooses less windows with decreasing the level of maximum α. Additionally, it can be observed that the number of filtered windows raises

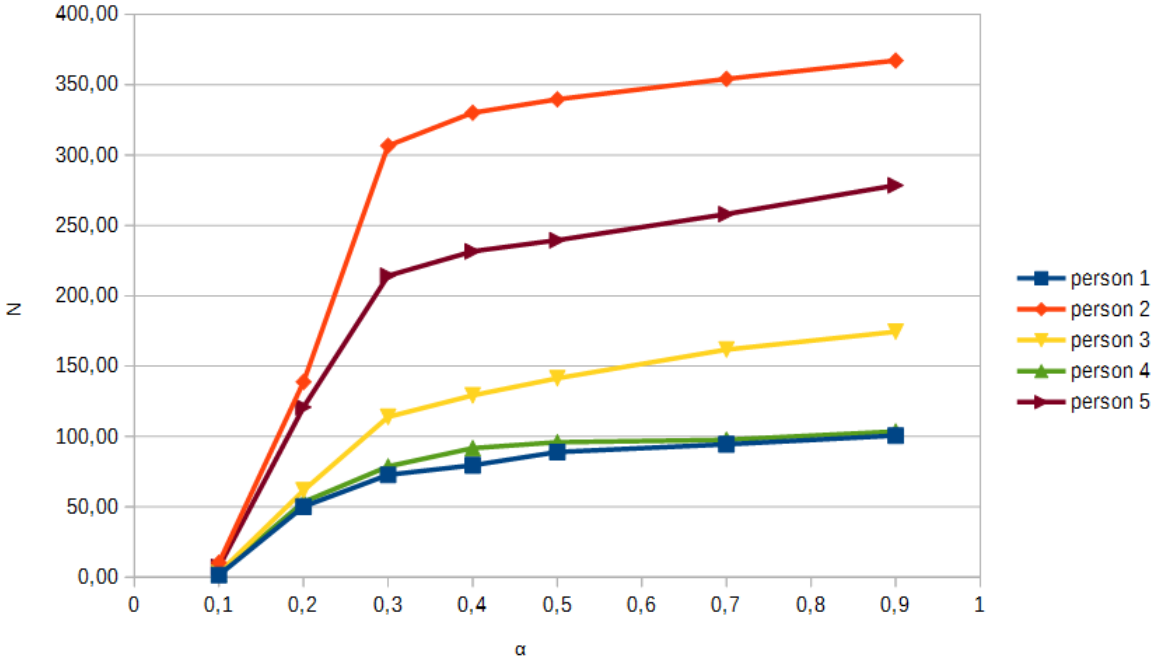

Fig. 4. Average number of filtered windows according to maximum level of α (trigger). Results obtained for signatures of 5 different individuals.

significantly until the trigger value of α gains the level 0.3. For higher values of α the number of filtered windows raises much slower.

5. CONCLUSION

The article presented the method of signature analysis in order to find valuable ranges for further processing by recognition systems. Focusing only on chosen ranges will allow recognition algorithms to reduce computation time. However, the influence of filtering at different levels on recognition efficiency, or other aspects, have to be further examined. The natural first step is to verify results of original window method employing the filtering procedure using different ranges of parameters defined in previous sections.

Results of examinations performed in order to verify theoretical analysis confirm the validity of deliberations. Future research will focus on using the method as an assessment tool to classify quality of signatures in terms of recognition process regardless to used approach.

BIBLIOGRAPHY

[1] Czogała E., Łęski J. Fuzzy and neuro-fuzzy intelligent systems. 2000. Physica-Verlag, Springer-Verlag Comp.

[2] Doroz R., Wróbel K. Dynamic signature recognition based on modified windows technique. Lecture Notes in Computer Science (LNCS), Springer, 2012, Vol. 7564. pp. 158–167.

[3] KUDŁACIK P. Improving a signature recognition method using the fuzzy approach. Journal of MIT, 2012, Vol. 21. pp. 85–93.

[4] Kudłacik P., Porwik P. A new approach to signature recognition using the fuzzy method. Pattern Analysis and Applications, 2014, Vol. 17, Issue 3. pp. 451–463.

[5] 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. Mcyt baseline corpus: a bimodal biometric database. Vision, Image and Signal Processing, IEE Proceedings -, Dec 2003, Vol. 150. pp. 395–401.

[6] Porwik P., Doroz R., Wróbel K. A new signature similarity measure based on windows allocation technique. Int. Journal of Computer Information Systems and Industrial Management Applications (IJCISIM), 2010, Vol. 2. pp. 297–305.

[7] PORWIK P., WRÓBEL K., DOROZ R. Signature recognition method by means of the windows technique. Image Processing and Communications, 2009, Vol. 14. pp. 43–50.

[8] WRÓBEL K., DOROZ R. The method of signature recognition based on least squares contour alignment and windows technique. Journal of Medical Informatics and Technologies, 2010, Vol. 15. pp. 35–39.

[9] ZADEH L. A. Fuzzy sets. Information and Control, 1965, Vol. 8. pp. 338–353.