A COMPREHENSIVE EXPERIMENTAL COMPARISON OF THE AGGREGATION TECHNIQUES FOR FACE RECOGNITION

P. KARCZMAREK, W. PEDRYCZ, A. KIERSZTYN, M. DOLECKI

Abstract. In face recognition, one of the most important problems to tackle is a large amount of data and the redundancy of information contained in facial images. There are numerous approaches attempting to reduce this redundancy. One of them is information aggregation based on the results of classifiers built on selected facial areas being the most salient regions from the point of view of classification both by humans and computers. In this study, we report on a series of experiments and offer a comprehensive comparison between various methods of aggregation of outputs of these classifiers based on essential facial features such as eyebrows, eyes, nose, and mouth areas. For each of them, we carry the recognition process utilizing the well-known Fisherfaces transformation. During the comparisons of the vectors representing the features of images (faces) after the transformations, we consider 16 similarity/dissimilarity measures for which we select the best aggregation operator. The set of operators to compare was selected on a basis of the comprehensive literature review regarding aggregation functions.

1. Introduction

Face classification has been one of the most intensively, but still not completely successful, investigated disciplines of a modern biometrics in the recent years by the researchers from the whole world. This is implied by numerous factors such as its wide usage in security systems, criminal investigations, and others. The accuracies reported in the modern face recognition literature are very high and the methods seem to be efficient. However, there are still a few problems to be fully solved, e.g., coping with aging process, pose, illumination and partial occlusion of the face, and finally, large datasets to retrieve in a real-time applications. In a more general perspective, the important problems of modern biometrics are listed in [6].

Information fusion techniques, in general, and the aggregation operators (functions) in particular, realizing these techniques, are used in many fields of science such as economics, biology, education, computer science, and artificial intelligence (mainly in multi-criteria decision making theory), etc., see [66]. In the field of face recognition many works have been devoted to the problem of classification involving more than a single classifier. The classifiers can utilize facial regions images, whole face 2D or 3D images, or an infrared image (see the section devoted to related work). The need for partitioning face into meaningful parts comes from the

Received: June 2005; Revised: July 2006; Accepted: November 2006

Key words and phrases: Face recognition, Aggregation functions, Facial features, Similarity/dissimilarity measures, Information aggregation.

Dataset	Function	Acc.
AT&T	$f(x,y) = \min(x,y)\min(1, x^p + y^q)$ with $p = 1$, $q = 2$	91.52
	Extreme value copula (EV-copula) $f(x,y) = \min\left(x\sqrt{y}, y\right)$	91.5
	Farlie-Gumbel-Morgenstern copula $C_\lambda^{FGM}(x,y) = xy + \lambda xy(1-x)(1-y)$ with $\lambda = \frac{3}{4}$	91.46
	General nonassociative symmetric copula $C(x,y) = pxy + (1-p)\min(x,y)$ with $p = 0.75$	91.46
	Weighted radical mean $RM_{\mathbf{w},\gamma}(x_1,\ldots,x_n) = \frac{1}{\left(\log_\gamma \sum\limits_{i=1}^{n} w_i \gamma^{\frac{1}{x_i}}\right)}$ with $\gamma = 0.75$	91.45
	$f(x,y) = \min(x,y)\min(1, x^p + y^q)$ with $p = 2$, $q = 1$	91.45
FERET	Choquet integral with membership grades $\mu_{ij} = \frac{1}{1+\frac{d_{ij}}{\overline{d_i}}}$ where i denotes the classifier number, j is the index of the training image, $\overline{d_i}$ means the average distance within ith classifier, and d_{ij} is a distance between a given image and jth training image within ith classifier	84.59
	Median $\text{med}(x_1,\ldots,x_n)$	84.54
	Choquet integral with membership grades $\mu_{ij} = \frac{1}{2}\left(1 + \frac{\frac{d_{ij}}{\overline{d_i}}}{1\frac{d_{ij}}{\overline{d_i}}}\right)$	84.52
	Choquet integral with membership grades $\mu_{ij} = \frac{1}{2}\left(1 + \frac{d_{ij}}{1+\frac{1}{d_{ij}}}\right)$	84.5
	Voting: One vote for each minimum in one of 4 classifiers	83.98
	Weighted average $m_w(x_1,\ldots,x_n) = \frac{1}{n}\frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i}$	82.61

Table 1. Accuracy values obtained for Euclidean distance

following two facts. First, some small parts of the face can be visible, for instance a person can wear sunglasses, a balaclava, a helmet, a veil, a mask, or the face might not accurately aligned. Second, choosing a few parts of the face can significantly reduce the size of data and in consequence the computation time. Bearing these facts in mind, it can be seen that the need of an appropriate choice of aggregation function may be fundamental for the construction of face recognition system. Such function or operator should efficiently minimize the possibility of incorrect classification and maximize the likelihood of the correct finding the proper class of the unknown person, namely his/her identification. Therefore, at the stage of aggregation of the results of particular classifiers such as eyes, nose, or mouth regions, the aggregation function should ignore the single cases of misclassification being the

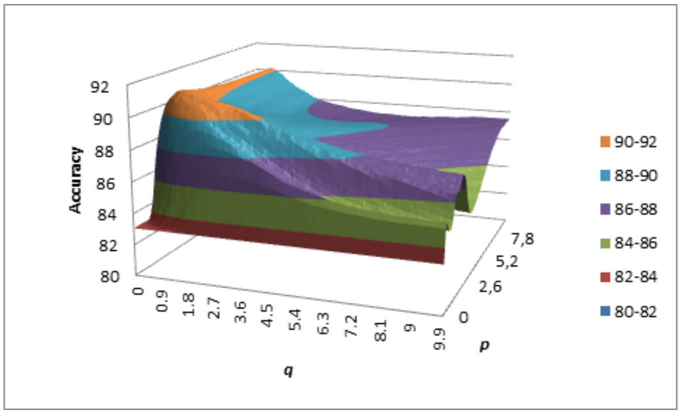

FIGURE 1. Accuracies of repeated series of experiments for the winning function $f(x, y)$ (see Table 1)

results of individual classifiers and, at the same time, it should emphasize the summary results for the cases of individual's classifications being positioned instantly at the top of rankings of most of the classifiers.

The main objective of this study is to determine the best aggregation functions in the process of integration of the information coming from various classifiers depending on the distance measure used to compare the features representing the testing and training images. We carefully examine the accuracies obtained in the processes of recognition based on four facial regions, i.e. eyes, eyebrows, nose, and mouth, using the well-known face recognition approach which is Fisherfaces. In a series of experiments with the AT&T (formerly ORL) and FERET datasets we have found the best aggregation techniques and corresponding distance measures giving the highest results of recognition. It should be stressed that although the main concepts identified above are well-known in the existing literature, a comprehensive and quantitatively sound analysis leading to meaningful design guidelines has not been provided so far. The originality of this study lies in a comprehensive and very thorough, in-depth comparative study. We believe that this work may help in making an appropriate choice as to the method of aggregation in any biometric system involving a family of classifiers, e.g. in multimodal face recognition, etc. Our motivation is not only to obtain the best possible biometric classification result but to shed the light on the choice of aggregation operator in general, i.e., in any system where the results of classification need to be aggregated. The study is highly experimental and its novelty is implied by the scale of the conducted experiments. Over $1,000$ various aggregation operators were tested. The work demonstrates that the choice may vary depending upon the dataset or the form of classifier. Therefore, the results may be of interest to researchers working on aggregation problems in a broader context where the multi-criteria decision making theory has to address the needs stemming from practical applications. The datasets of faces consist of

Dataset	Function	Acc.
AT&T	$f(x,y)=\min(x,y)\min(1,x^p+y^q)$ with $p=\frac{9}{10}$, $q=\frac{9}{10}$	91.54
	Weighted Yager t-norm $T^Y_{\lambda,w}=\max\left(0,\left(1-\sum_{i=1}^{n}w_i(1-x_i)^{\lambda}\right)^{1/\lambda}\right)$ with $\lambda=10$	91.54
	EV-copula $f(x,y)=\min\left(x\sqrt{y},y\right)$	91.5
	$f(x,y)=\min(x,y)\min(1,x^p+y^q)$ with $p=\frac{9}{10}$, $q=\frac{9}{10}$	91.5
	Commutative, non-associative copula	91.41
FERET	The same order of functions as in Table 1 with the average accuracies $84.59, 84.29, 84.2, 84.03, 82.49$, respectively	

TABLE 2. Accuracy obtained for squared Euclidean distance

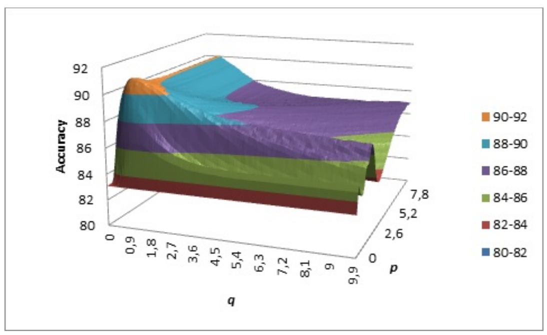

FIGURE 2. Winning function and the values of aggregation results with respect to parameters p and q (squared Euclidean distance)

relatively small set of images. However, these data are sufficient to arrive at some regularities of the behavior of the selected aggregation operators.

The bibliography of the paper is extended. Nevertheless, we are also interested in presentation in a form of a shortened survey this wide spectrum of aggregation techniques applied to the computational face recognition problem and specification of the most important, in our opinion, directions of development.

The paper is structured as follows. The related studies are described in Section 2. A discussion on the main properties of aggregation operators and a general processing scheme is presented in Section 3. Various distance measures used in the experiments are discussed in Section 4. The experimental results are presented in detail in Section 5. Finally, in Section 6 we offer some conclusions.

2. **Related studies**

Since face recognition is one of the most challenging problems in computer vision for more than last three decades, there are many approaches to this field (see, e.g. a wide survey in [73]). The most known examples are eigenfaces [67], Fisherfaces [7], local descriptors and their modifications [1, 37] or recently intensively developed

Dataset	Function	Acc.
AT&T	Power-based OWA $OWA_p(x_1,\ldots,x_n) = \left(\sum_{i=1}^{n}\left(w_i x_{(i)}^r\right)\right)^{1/r}$ with $r = 0.1$, where $w_i = \frac{2i}{n(n+1)}$ and $\left(x_{(1)},\ldots,x_{(n)}\right)$ is decreasing	90.7
	Ordered weighted geometric mean $OWGM(x_1,\ldots,x_n) = \prod_{i=1}^{n} x_{(i)}^{w_i}$ with $w_i = 1 - \frac{i-1}{n}$ and $\left(x_{(1)},\ldots,x_{(n)}\right)$ is increasing	90.66
	Family $T_\alpha(x,y) = \left(1 + \frac{\left[(1+x)^{-\alpha}-1\right]\left[(1+y)^{-\alpha}-1\right]}{2^{-\alpha}-1}\right)^{-1/\alpha} - 1$ with $\alpha = 10$	90.6
	$OWGM(x_1,\ldots,x_n)$ with $w_1 = \frac{1}{2}, w_2 = \frac{1}{2}mw_3 = \frac{1}{4}, w_4 = \frac{1}{8},\ldots$ (the rest parameters are as above)	90.58
	Frank t-norm $T_\lambda^F(x,y) = \log_\lambda\left(1 + \frac{(\lambda^x-1)(\lambda^y-1)}{\lambda-1}\right)$ with $\lambda = 0.0001$	90.58
FERET	Function $G_\alpha(x,y) = \log\left(e^{\alpha x} + e^{\alpha y}\right)$ with $\alpha = 0.99$	86.48
	Exponential mean $EM(x_1,\ldots,x_n) = \frac{1}{\alpha}\log\left(\frac{1}{n}\sum_{i=1}^{n} e^{\alpha x_i}\right)$ with $\alpha = 10$	86.36
	Weighted Frank t-norm $T_{\lambda,w}^F(x_1,\ldots,x_n) = \log_\lambda\left(1 + \frac{\prod_{i=1}^{n}(\lambda^{x_i}-1)^{w_i}}{(1-\lambda)^{\sum_{i=1}^{n} w_i - 1}}\right)$ with $\lambda = 0.25$	86.27
	Kolesárová function $M(x_1,\ldots,x_n) = \frac{GM(x_1,\ldots,x_n)}{(x_1,\ldots,x_n)-(1-x_1,\ldots,1-x_n)}$ where $GM(x_1,\ldots,x_n)$ is the geometric mean	86.26
	Weighted radical mean $RM_{w,\gamma}(x_1,\ldots,x_n) = \left(\log_\gamma\left(\sum_{i=1}^{n} w_i \gamma^{1/x_i}\right)\right)^{-1}$ with $\gamma = 5$	86.26

TABLE 3. Accuracy for modified Euclidean distance

sparse representation [68] and deep learning [31, 64] studied on the LFW [30] or CelebFaces [63] datasets. Finally, there are many algorithms, which try to combine the results of classification based on more than one source of information on the subject, in particular in the case when the parts of the face are the inputs to the classification. One of them is an information fusion approach [45], where the final result is obtained by combining the results of classification based on the particular facial features, i.e. eyes, nose, mouth, and the whole face using the Choquet fuzzy integral and fuzzy measure (see [17, 38, 54]) as a vehicle for aggregation operation for the results of the specific classifiers.

The general theory of the aggregation operators can be found in several monographs [8, 12, 13, 25, 66] and in the papers and books devoted to the particular topics related to aggregation functions such as triangular norms and copulas [4, 41, 42], fuzzy sets, fuzzy measure, and fuzzy integral [24, 47, 57], OWA (ordered weighted averaging operators [70], and other techniques known from the multi-criteria decision making theory, e.g. [19].

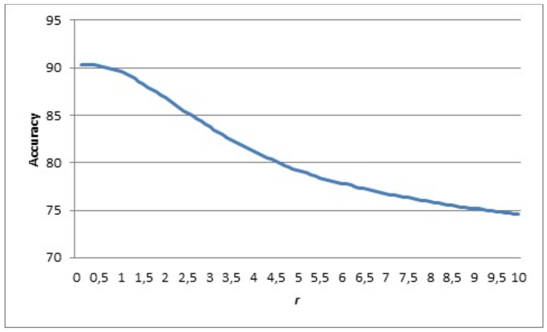

FIGURE 3. Power-based OWA and the dependence of its results on the parameter r (modified Euclidean distance)

Dataset	Function	Acc.
AT&T	Weighted Yager t-norm $T^Y_{\lambda,w}$ with $\lambda = 10$ (see Table 2)	85.32
	EV-copula $f(x, y) = \min\left(x\sqrt{y}, y\right)$	85.29
	Ordered weighted geometric mean $OWGM(x_1, \dots, x_n)$ with $w_i = 1 - \frac{i-1}{n}$ (see Table 3)	85.25
	Power-based OWA (see Table 3)	85.24
	Frank t-norm $T^F_\lambda(x, y)$ with $\lambda = 0.0001$	85.21
FERET	Choquet integral with membership grades $\mu = \frac{1}{1+\frac{d_{ij}}{d_i}}$ (see Table 1)	80.31
	Median $med(x_1, \dots, x_n)$	79.94
	Choquet integral with membership grades $\mu = \frac{1}{2}\left(1 + \frac{\frac{d_{ij}}{d_i}}{1+\frac{d_{ij}}{d_i}}\right)$ (see Table 1)	79.75
	Choquet integral with membership grades $\mu = \frac{1}{2}\left(1 + \frac{d_{ij}}{1+d_{ij}}\right)$ (see Table 1)	79.42
	Weighted Yager t-norm $T^Y_{\lambda,w}(x, y)$ with $\lambda = \frac{1}{4}$	78.08

TABLE 4. Accuracy for weighted squared Euclidean distance

In the pattern recognition literature, there are two approaches to the information fusion. First of them is a data-level approach which is based on the combination of the information contained in the considered features (images, facial parts, etc.). The second one is a score-level approach. Here the results of classification form the inputs to the final classification result. The last one is obtained using aggregation operations only.

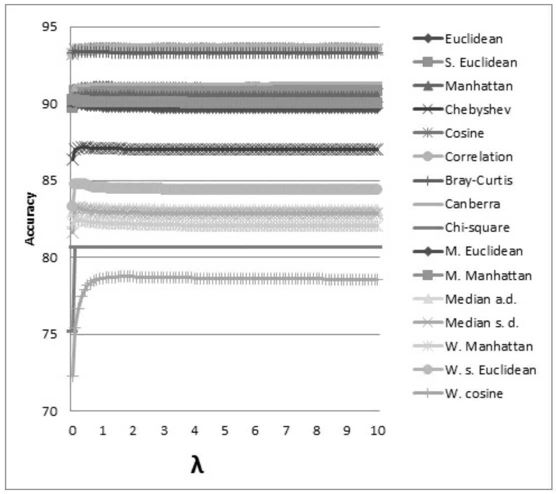

FIGURE 4. Results for modified Hamacher function as aggregator of classifiers in various similarity measures

In the face recognition literature many applications of the aggregation functions or, generally speaking, information aggregation, are described. The first type of applications is to combine the outcomes of various classifiers based on the same whole face image but using various techniques of description or comparison of one image with another images. The other approach is to combine the results of the classifiers based on the different images of the same person. Here the important problem is to aggregate the information obtained from the standard 2D photo of the person and his/her infrared image or from 2D and 3D images. Finally, the last approach is to combine the results of classification obtained utilizing only the information contained in two or more facial features or, optionally, contained in the whole facial image. Let us recall the main results reported in the literature. In [11] template matching strategy was applied to produce significant improvement of recognition result when the particular feature (eyes, nose, and mouth) scores and the whole face score are added. In [58] the eigenfeatures approach, i.e., the eigenfaces method for the most important facial features, was used. The authors proposed a few strategies to aggregate classification results. Similar methods were presented in [23] and [39]. In [26] radial basis function (RBF) neural network-based classifiers for various facial domains were applied and then fused together through the majority rule. The system based on the aggregation of classifiers such as eyes and forehead, nose, and mouth images, using fuzzy measure and Choquet integral for fusion of classifiers was proposed in [53]. The authors of [45] integrated the results of classification based on the Fisherfaces method applied to the images of particular facial features, i.e. eyes, nose, mouth, and the whole face, using Choquet fuzzy integral and fuzzy measure. Similarly, they proposed the aggregation for the Fisherfaces results of four classifications for the respective 3 kinds of wavelet decompositions of the images and

Dataset	Function	Acc.
AT&T	Modification of Hamacher t-norm $T_\lambda(x,y)=\frac{xy}{\lambda+(1+\lambda)(x+y-xy)}$ with $\lambda=\frac{3}{5}$	90.91
	Modification of Hamacher t-norm with $\lambda=\frac{1}{2}$	90.91
	Weighted radical mean $RM_{w,\gamma}(x_1,\ldots,x_n)$ with $\gamma=1.01$ (see Table 1)	90.9
	Weighted harmonic mean $\sum_{i=1}^{n} w_i/x_i$	90.9
	EV-copula $f(x,y)=\min\left(x\sqrt{y},y\right)$	90.89
FERET	Choquet integral (see Table 1)	82.96
	Median $med(x_1,\ldots,x_n)$	82.9
	Choquet integral with membership grades $\mu=\frac{1}{2}\left(1+\frac{\frac{d_{ij}}{d_i}}{1+\frac{d_{ij}}{d_i}}\right)$	82.88
	Choquet integral with membership grades $\mu=\frac{1}{2}\left(1+\frac{d_{ij}}{1+d_{ij}}\right)$	82.79
	Voting: One vote for each minimum in one of 4 classifiers	82.18

TABLE 5. Accuracy produced for Manhattan distance

the original images from the database [44, 45]. In [52] neural networks were applied to the facial modules such as eyes, nose, and mouth regions, then the results were integrated using the fuzzy integral in a similar way. In [21] the 8×8 blocks of pixels are described using the discrete cosine transform and the final combined result is obtained using the sum rule [40]. Similar results for 14 facial parts are obtained in [22]. In [72] it was presented combination by majority voting of the eigenfaces method with eyes, mouth, nose, and forehead eigenfeature. The LDA method for the whole face and its main components was discussed in [18]. The authors produced the final results using an OR gate. In [33] an aggregation based on majority voting and Bayesian product for the PCA, LDA, kernel-PCA, and Isomap methods for dimensionality reduction and histogram equalization and Sobel edge detector at the stage of image description was proposed. Color, local spatial and global frequency information fusion was comprehensively discussed in [48]. At the decision level the authors proposed the weighted sum rule for fusion of similarity matrices. In the work [38] an application of the Sugeno fuzzy measure [62] to the process of aggregation of the classifiers based on eyebrows, eyes, nose, mouth, right and left cheeks regions was examined. The aggregation was prepared both at the level of data and at the level of classification. An automatic manner of determination of fuzzy measure parameters were proposed in [34, 35]. A similar processes with an application of a so-called generalized Choquet integral were examined in [36]. In [32] it was proposed to join a pixel-level and a feature-level fusion at the top-level's wavelet sub-bands. The authors reduced the problem of finding the best pixel-level fusion coefficients to the PCA- and LDA-based optimization problems. Furthermore, there are many results involving SVMs. For instance, in [28] proposed was a system where at the phase of detection, the facial features are inputs to the SVM classifiers, and next, at the phase of recognition, the normalized results are again the inputs for the SVM classifier. Similarly, in [27] 9 facial features and the global feature were taken into account. The gray pixel values of the components were combined into a single vector. A set of respective 10 second-degree polynomial SVM classifiers was trained using these vectors. Finally, the outputs of the SVM classifiers were compared. A similar method was presented in [29]. Newer methods of face recognition based on data-level fusion or score-level fusion were discussed in [3] (spatially optimized fusion of texture and shape), [15] (kernel fusion of multiple descriptors), [46] (application of SVM and its modifications for 3D face recognition), [56] (aggregation of polynomial-based RBF neural networks), [71] (feature fusion based on polynomial correlation filter), [74] (fusion of the results of proposed by the authors 2D-PCA modifications), [61] (a selection and fusion of ensembles produced through Boolean classifiers combination), or [16] (fusion of local descriptions of before-makeup and after makeup images). An interesting example of study of descriptive capabilities of 15 facial features from a forensic point of view can be found in [65]. Again, Choquet integral has been recently efficiently utilized as an integration method in modular neural networks in [50, 51]. [43] proposed Granular Computing techniques to work with various features of a face. Various aggregation functions were interestingly used to model craniofacial correspondence in craniofacial superimposition in [14]. In [20] proposed was a set of utility functions, which can be effectively used as aggregation operators. A three-valued logic and fuzzy set-based decision mechanisms were proposed in [2]. Finally, there are some studies discussing the impact of various distance measures on the quality of biometric systems, see, e.g. [1, 9, 10, 55, 59, 69]. Their general conclusion is that the appropriate choice of the distance measure may have a significant influence on the efficiency of the method.

3. Aggregation of classifiers

Let us consider the following situation. Given is a face recognition system comparing the facial regions images. Assume that someone has been classified as person A when taking into account his/her eyes region. But he/she has been also classified as a person B when taking into account his/her nose region. Of course, we can consider many other examples of misclassification. But the main question remains the same. How to correctly classify this person? One of the natural ways of solving this problem may be a choice of an aggregation function, i.e. the function returning as a result the proper class taking as the input the results of the particular classifiers, e.g. the distances between the features vectors or the rankings. However, this is only our wish, since even if we have the best aggregation function, the results of the particular classifiers could be incorrect, and then the final classification result will probably be incorrect as well. On the other hand, it is intuitively obvious, that the general accuracy of classification should be improved if the number of classifiers increases (at least to some level).

Dataset	Function	Acc.
AT&T	Weighted radical mean $RM_{w,\gamma}(x_1,\ldots,x_n)$ with $\gamma=1.01$	90.61
	Weighted harmonic mean	90.59
	T-norm $T_{\alpha,\beta}(x,y)=\frac{1}{\beta}\left(1+\frac{\left((1+\beta x)^{-\alpha}-1\right)\left((1+\beta y)^{-\alpha}-1\right)}{(1+\beta)^{-\alpha}-1}\right)^{\frac{-1}{\alpha}}-\frac{1}{\beta}$ with $\alpha=-10,\beta=-\frac{1}{2}$	90.5
	Ordered weighted harmonic mean $\sum_{i=1}^{n}\left(\frac{w_i}{x_{(i)}}\right)$ with $w_1=1,w_2=\frac{9}{10},w_3=\frac{4}{5},w_4=\frac{7}{10}$	90.48
	Family $F_\alpha(x,y)=\frac{\alpha}{\log\left(e^{\alpha/x}+e^{\alpha/y}-e^{\alpha}\right)}$ with $\alpha=\frac{1}{10}$	90.48
FERET	Extension of the Wiener-Shannon law $F_\alpha(x,y)=-\frac{1}{\alpha}\log\left(e^{-\alpha x}+e^{-\alpha y}\right)$ with $\alpha=0.99$	84.99
	Quasi-arithmetic mean $M_g(x_1,\ldots,x_n)=\begin{cases}\frac{\sqrt[n]{\prod_{i=1}^{n}x_i}}{\sqrt[n]{\prod_{i=1}^{n}x_i}+\sqrt[n]{\prod_{i=1}^{n}(1-x_i)}}, & if\{0,1\}\not\subseteq\{x_1,\ldots,x_n\}\\ 0, & otherwise\end{cases}$	84.7
	Weighted radical mean $RM_{w,\gamma}(x_1,\ldots,x_n)$ with $\gamma=7$	84.7
	Kolesárová function $M(x_1,\ldots,x_n)$ (see Table 3)	84.7
	Weighted Frank t-norm with $\lambda=\frac{1}{4}$ (see Table 3)	84.68

Table 6. Accuracies for modified Manhattan distance

In the simplest form, an aggregation function is defined as the function $f: [0,1]^n \to [0,1]$ exhibiting the following properties

$$f(0,0,\ldots,0)=0, \quad f(1,1,\ldots,1)=1,$$

and

$$\forall \mathbf{x},\mathbf{y}\in[0,1]^n \ \mathbf{x}\le\mathbf{y} \implies \mathbf{f}(\mathbf{x})\le\mathbf{f}(\mathbf{y}),$$

i.e. it has the properties of bounds preservation and monotonicity [8]. Typical representations are the basic mathematical means. There are many classes of aggregation operators such as averaging, conjunctive, disjunctive, or mixed. As a consequence, they can have various properties such as idempotency, symmetry, neutral element, etc. From the point of view of potential applications, the properties of continuity and stability may be very important. These two conditions assure that the small change in the values of the input variables implies a small change in the result produced by the aggregation function. Of course, these conditions can be defined in a more specific manner (say, a function can be continuous, Lipschitz continuous, etc.).

The situation considered above is a typical problem of the multicriteria decision making theory and information fusion. In such kinds of applications the authors of [25] advise the usage of aggregation functions like weighted arithmetic mean or

Dataset	Function	Acc.
AT&T	Weighted harmonic mean	82.94
	Weighted radical mean $RM_{w,\gamma}(x_1,\ldots,x_n)$ with $\gamma = 1.01$	82.93
	OWGM with $w_1 = 1, w_2 = \frac{3}{4}, w_3 = \frac{1}{2}, w_4 = \frac{1}{4}$	82.6
	Power-based OWA (see Table 3 with $w_1 = \frac{1}{2}, w_2 = 1, w_3 = \frac{3}{2}, w_4 = 2$	82.6
	Mod. Hamacher t-norm $T_\lambda(x,y) = \frac{xy}{\lambda+(1+\lambda)(x+y-xy)}$ with $\lambda = \frac{1}{2}$	82.56

Table 7. Accuracies for weighted Manhattan distance

Dataset	Function	Acc.
AT&T	Mod. Hamacher t-norm with $\lambda = \frac{1}{2}$ or $\lambda = \frac{3}{5}$	87.67
	Mod. Hamacher t-norm with $\lambda = 1$	87.6
	Fuction $f(x_1,\ldots,x_n) = \prod_{i=1}^{n} x_i^{w_i}$, weighted geometric mean, weighted product	87.58
	Weighted Gini mean $G(x_1,\ldots,x_n) = \left(\prod_{i=1}^{n} x_i^{w_i x_i^p}\right)$ with $p = \frac{1}{10}$	87.57
	Mod. Hamacher t-norm with $\lambda = 2$	87.56

Table 8. Accuracies for Chebyshev distance

other means or averages, integral-type aggregation functions (e.g. Choquet integral), etc. On the other hand, at the stage of data fusion in this situation the authors advise to use, for instance, Choquet integral.

In this paper, we comprehensively revisit the aggregation problem presented above. The main processing phases are as follows. First, an interior face without the ears, hair, and neck is cropped, scaled, and an image histogram is equalized. Next, positions of the most important face areas, namely eyes, eyebrows, nose, and mouth are manually determined. The specific parameters of these areas were chosen using the results of some introductory experiments. Then, for the four kinds of images containing selected face parts the well-known Fisherfaces dimensionality reduction method is executed. The feature vectors being the result of the method and corresponding to the images contained in the training and testing sets are then compared using 16 distance measures (which are listed in the next section). The distances are normalized to the interval [0,1]. Finally, these results (i.e. the distances between the training and testing images) are aggregated using one of over 1000 aggregation operators chosen from the database of aggregation functions. As the database of these functions we used the well-known monographs devoted to the aggregation operators and their applications, namely [4, 8, 25, 57]. Obviously, we do not present all the formulas and confine our presentation to demonstrate the best results only.

One of the most discussed aggregation operators in this study is Choquet integral. Let us recall this concept. Consider that $X = \{x_1,\ldots,x_n\}$ denotes the whole face, where $x_1,\ldots,x_n$ mean the particular facial segments such as eyes, nose, etc. Then the fuzzy measure can be defined as a set function $g : P(X) \to [0,1]$ fulfilling the relations:

(1) $1g(\emptyset) = 0, \quad g(X) = 1$;
(2) $2g(A) \le g(B)$ for $A \subset B$, where $A, B \in P(X)$.

λ-fuzzy measure introduced by Sugeno (1974) satisfies the following rule

$$g(A \cup B) = g(A) + g(B) + \lambda g(A)\, g(B), \quad \lambda > -1,$$

where A and B are disjoint sets. The value of λ can be easily obtained as a unique solution for $\lambda > -1$, $\lambda \neq 0$, of the equation [62]

$$1 + \lambda = \prod_{i=1}^{n} (1 + \lambda g_i), \quad g_i = g(\{x_i\}).$$

The values g_i have to be set. If we denote $A_i = \{x_1, \ldots, x_n\}$, $A_{i+1} = \{x_1, \ldots, x_n, x_{n+1}\}$, one can use the formula for overlapping regions

$$g(A_{i+1}) = g(A_i) + g_{i+1} + \lambda g(A_i)\, g_{i+1},$$

assuming $g(A_1) = g_1$. Finally, the Choquet integral is defined as

$$\mathrm{Ch} \int h \circ g = \sum_{i=1}^{n} (h(x_i) - h(x_{i+1}))\, g(A_i), \; h(x_{n+1}) = 0,$$

where $h : X \to [0,1]$ is a measurable function and the values of it are rearranged to satisfy the inequality $h(x_i) \ge h(x_{i+1})$, $i = 1, \ldots, n$. In our experiments, we use the nomenclature from [45], namely

$$h(y_{ik}) = \frac{1}{N_k} \sum_{\mu_{ij} \in C_k} \mu_{ij},$$

where N_k is a number of images in kth class C_k, μ_{ij} is calculated based on the $\overline{d_i}$, $\overline{d_i}$ stands for the average distance within ith classifier, and d_{ij} is a distance between a given image and jth training image within ith classifier (see Table 1).

4. Distance measures

There are many distance measures, or, more generally, similarity/dissimilarity measures which may be applied to quantify the closeness of two vectors $\mathbf{x} = (x_1, \ldots, x_n)$, $\mathbf{y} = (y_1, \ldots, y_n) \in \mathbb{R}^n$. In the series of experimental tests we use the following functions

(1) Euclidean $d(\mathbf{x}, \mathbf{y}) = \left(\sum_{i=1}^{n} (x_i - y_1)^2 \right)^{\frac{1}{2}}$,
(2) Squared Euclidean $d(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} (x_i - y_1)^2$,
(3) Manhattan $d(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} |x_i - y_1|^2$,
(4) Chebyshev $d(\mathbf{x}, \mathbf{y}) = \max_i |x_i - y_i|$,
(5) Cosine $d(\mathbf{x}, \mathbf{y}) = 1 - \frac{\sum_{i=1}^{n} (x_i y_i)}{\left(\sum_{i=1}^{n} x_i^2\right)^{\frac{1}{2}} \left(\sum_{i=1}^{n} y_i^2\right)^{\frac{1}{2}}}$,
(6) Weighted cosine $d(\mathbf{x}, \mathbf{y}) = 1 - \frac{\sum_{i=1}^{n} \frac{x_i y_i}{\sqrt{w_i}}}{\left(\sum_{i=1}^{n} x_i^2\right)^{\frac{1}{2}} \left(\sum_{i=1}^{n} y_i^2\right)^{\frac{1}{2}}}$,

Dataset	Function	Acc.
AT&T	Mod. Hamacher t-norm with $\lambda = \frac{1}{2}$	93.81
	Mod. Hamacher t-norm with $\lambda = \frac{3}{5}$	93.81
	Mod. Hamacher t-norm with $\lambda = 1$	93.8
	$A(x,y) = \begin{cases} x^{\alpha}y^{1-\alpha} & for\ x \leq y \\ x^{1-\beta}y^{\beta} & otherwise \end{cases}$ with $\alpha = \frac{3}{4}, \beta = \frac{1}{4}$	93.8
	EV-copula $f(x,y) = \min\left(x\sqrt{y}, y\right)$	93.79
FERET	Voting	90.96
	Choquet integral with membership grades $\mu_{ij} = \frac{1}{2}\left(1 + \frac{d_{ij}}{1+d_{ij}}\right)$	90.79
	Choquet integral with membership grades $\mu_{ij} = \frac{1}{2}\left(1 + \frac{\frac{d_{ij}}{d_i}}{1+\frac{d_{ij}}{d_i}}\right)$	90.72
	Median	90.66
	Choquet integral (see the first one in Table 1)	90.35

TABLE 9. Accuracies for cosine distance

(7) Correlation $d(\mathbf{x},\mathbf{y}) = 1 - \frac{\sum_{i=1}^{n}\left(\left(x_i - \frac{1}{n}\sum_{j=1}^{n}x_j\right)\left(y_i - \frac{1}{n}\sum_{j=1}^{n}y_j\right)\right)}{\left(\sum_{i=1}^{n}\left(x_i - \frac{1}{n}\sum_{j=1}^{n}x_j\right)^2\right)^{\frac{1}{2}}\left(\sum_{i=1}^{n}\left(y_i - \frac{1}{n}\sum_{j=1}^{n}y_j\right)^2\right)^{\frac{1}{2}}}$,

(8) Bray-Curtis $d(\mathbf{x},\mathbf{y}) = \frac{\sum_{i=1}^{n}|x_i - y_i|}{\sum_{i=1}^{n}|x_i + y_i|}$,

(9) Canberra $d(\mathbf{x},\mathbf{y}) = \sum_{i=1}^{n}\frac{|x_i - y_i|}{|x_i| + |y_i|}$,

(10) χ^2-statistics $d(\mathbf{x},\mathbf{y}) = \sum_{i=1}^{n}\frac{(x_i - y_i)^2}{x_i + y_i}$,

(11) Modified Euclidean $d(\mathbf{x},\mathbf{y}) = \frac{\sum_{i=1}^{n}(x_i - y_i)^2}{\sum_{i=1}^{n}x_i^2\sum_{i=1}^{n}y_i^2}$,

(12) Modified Manhattan $d(\mathbf{x},\mathbf{y}) = \frac{\sum_{i=1}^{n}|x_i - y_1|^2}{\sum_{i=1}^{n}|x_i|^2\sum_{i=1}^{n}|y_1|^2}$,

(13) Median of absolute differences $d(\mathbf{x},\mathbf{y}) = \text{med}_i\,|x_i - y_i|$,

(14) Median of square differences $d(\mathbf{x},\mathbf{y}) = \text{med}_i\,(x_i - y_i)^2$,

(15) Weighted Manhattan $d(\mathbf{x},\mathbf{y}) = \sum_{i=1}^{n} = \frac{|x_i - y_i|}{\sqrt{w_i}}$,

(16) Weighted squared Euclidean $d(\mathbf{x},\mathbf{y}) = \sum_{i=1}^{n} = \frac{(x_i - y_i)^2}{\sqrt{w_i}}$,

where w_i, $i = 1, \ldots, n$ are the corresponding eigenvalues calculated when LDA method is processed.

The choice of these functions is motivated by their common usage in the literature and their presence in survey papers (see the related work section). We are interested in studying the dependence between the choice of distance measure at the stage of classification based on facial features and the aggregation function, which integrates these results giving the final score.

5. Experiments

5.1. **Experimental setup.** The main aim of the series of experiments reported in this work is to determine the accuracies of recognition after an aggregation of the

Dataset	Function	Acc.
AT&T	OWA function with $w_1=\frac{1}{4}, w_2=\frac{1}{8}, w_3=\frac{1}{12}, w_4=\frac{1}{16}$	84.79
	Special OWGM $OWGM(x_1,\ldots,x_n) = (\min(x_1,\ldots,x_n))^{1-\alpha}((x_1,\ldots,x_n))^{\alpha}$ with $\alpha=\frac{1}{10}$	84.65
	OWA function with $w_1=1, w_2=\frac{3}{4}, w_3=\frac{1}{2}, w_4=\frac{1}{4}$	84.61
	OWA function with $w_1=1, w_2=\frac{1}{42}, w_3=\frac{1}{4}, w_4=\frac{1}{8}$	84.56
	Weighted exponential mean $E(x_1,\ldots,x_n) = \log\sum_{i=1}^{n}\frac{w_i\log\alpha^{x_i}}{\log\alpha}$ with $\alpha=10$	84.31
FERET	First four functions as in Table 9 and weighted average with results $87.77, 87.71, 87.7, 87.67, 87.67$, respectively.	

Table 10. Accuracies for weighted cosine distance

Dataset	Function	Acc.
AT&T	EC-copula $f(x,y)=\min\left(x\sqrt{y},y\right)$	93.8
	$A(x,y)=\begin{cases} x^{\alpha}y^{1-\alpha}, & for\ x\le y \\ x^{1-\beta}y^{\beta}, & otherwise \end{cases}$ with $\alpha=0.75, \beta=0.25$	93.79
	Mod. Hamacher t-norm $T_\lambda(x,y)=\frac{xy}{\lambda+(1+\lambda)(x+y-xy)}$ with $\lambda=1$	93.76
	Weighted geometric mean	93.76
	Mod. Hamacher t-norm $T_\lambda(x,y)$ with $\lambda=\frac{1}{2}$	93.75
	Weighted Gini mean $G(x_1,\ldots,x_n)=\left(\prod_{i=1}^{n}x_i^{w_i x_i^p}\right)^{\frac{1}{\sum_{i=1}^{n}w_i x_i^p}}$ with $p=0.01$	93.75
FERET	The order of functions is the same as in Table 9 with the results as follows $90.73, 90.59, 90.51, 90.49, 90.14$	

Table 11. Accuracies produced for correlation distance

results of classification for 4 facial segments such as eyebrows, eyes, nose, and mouth areas. The properties of aggregation functions applied to integration of the partial results obtained by the classifiers corresponding to the most salient facial parts are comprehensively examined and compared. The performance of the aggregation functions is evaluated with the use of the well-known AT&T [5] and FERET [60] databases. The AT&T database contains 400 images of 40 subjects taken with different pose, expression, or illumination. When using the FERET database, we limited the set of the images to the so-called *ba*, *bk*, and *bj* subsets of the dataset. It contains 600 images of 200 people taken under different illumination and expression conditions.

In our previous work [38] we found the recognition rates for PCA followed by LDA method based on these facial regions images. The accuracies are the input weights for all the weighted aggregation functions, e.g. weighted average, Choquet integral, etc. The weights for, respectively, eyebrows, eyes, nose, and mouth are: 0.26, 0.28, 0.25, 0.21 (eigenfaces for AT&T), 0.28, 0.28, 0.23, 0.21, (Fisherfaces for AT&T), 0.49, 0.26, 0.18, 0.07 (eigenfaces for FERET), and 0.42, 0.3, 0.18, 0.1 (Fisherfaces for FERET).

Dataset	Function	Acc.
AT&T	Function $T(x,y) = \frac{xy}{x+y-xy}$	93.83
	EV-copula $f(x,y) = \min\left(x\sqrt{y}, y\right)$	93.69
	Weighted harmonic mean	93.63
	Weighted radical mean $RM_{w,\gamma} = (x_1, \ldots, x_n)$ with $\gamma = 1.01$	93.63
	$A(x,y)$ (see Table 9) with $\alpha = \frac{3}{4}, \beta = \frac{1}{4}$	93.61
FERET	Choquet integral with membership grades $\mu_{ij} = \frac{1}{2}\left(1 + \frac{d_{ij}}{1+d_{ij}}\right)$	90.2
	Choquet integral with membership grades $\mu_{ij} = \frac{1}{2}\left(1 + \frac{\frac{d_{ij}}{d_i}}{1+\frac{d_{ij}}{d_i}}\right)$	90.15
	Median	90.13
	Voting	90.13
	Choquet integral (see the first one in Table 1)	90.05

TABLE 12. Accuracies reported for Bray-Curtis distance

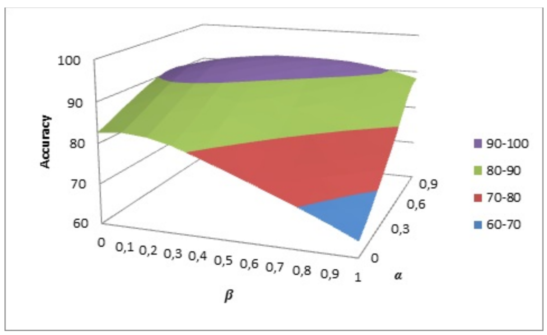

FIGURE 5. $A(x,y)$ results for PCA+LDA in Bray-Curtis similarity measure

As the results of the main experimental series, we present the average accuracies obtained for the combinations of the results of classifiers based on the four facial regions using the PCA followed by LDA method. Each of the testing cases was executed separately for the above mentioned 16 distance measures. Moreover, we make the partitioning of the collection AT&T images choosing randomly five photographs of each person to the training set and five photos remaining to the testing image set, respectively. The experiments with the FERET database we do similarly taking two images of the same person to the training set and the last one to the testing set. Such iteration was repeated 200 times and the final rates of recognition are computed as the averages of all the results. In the sequel, we repeat the tests with the new random choice of the images for the cases of functions having additional parameters to find the proper and optimal values for them. The computations were repeated for the selected functions and their modifications and only for the selected databases and distance measures. Recall that the weights w_i, $i = 1, \ldots, n$,

Dataset	Function	Acc.
AT&T	Mod. Hamacher t-norm $T_\lambda(x,y) = \frac{xy}{\lambda+(1+\lambda)(x+y-xy)}$ with $\lambda = 20$	91.43
	Mod. Hamacher t-norm $T_\lambda(x,y)$ with $\lambda = e$ or $\lambda = 4$	91.41
	Weighted Yager t-norm $T^Y_{\lambda,w}$ (see Table 2) with $\lambda = 2$	91.4
	Mod. Hamacher t-norm $T_\lambda(x,y)$ with $\lambda = 2$	91.4
	Mod. Hamacher t-norm $T_\lambda(x,y)$ with $\lambda = 1$	91.37
FERET	Voting	83.92
	Weighted average	83.76
	Choquet integral with membership grades $\mu_{ij} = \frac{1}{2}\left(1+\frac{d_{ij}}{1+d_{ij}}\right)$	83.44
	Choquet integral with membership grades $\mu_{ij} = \frac{1}{2}\left(1+\frac{\frac{d_{ij}}{d_i}}{1+\frac{d_{ij}}{d_i}}\right)$	83.36
	Median	83.32

Table 13. Accuracies for Canberra distance

are the corresponding eigenvalues calculated when the LDA method is processed unless their values are explicitly given.

5.2. **Experimental results.** The results obtained for the first distance measure, namely the Euclidean distance, are presented in Table 1. Similarly, the results obtained using squared Euclidean distance are gathered in Table 2. Top lists of results for modified Euclidean, weighted squared Euclidean, Manhattan, modification of the Manhattan, weighted version of Manhattan, Chebyshev, cosine, weighted cosine, correlation, Bray-Curtis, Canberra, χ^2-statistic, medians of absolute differences, and square differences similarity measures are presented in Table 3, Table 4, Table 5, Table 6, Table 7, Table 8, Table 9, Table 10, Table 11, Table 12, Table 13, Table 14, Table 15, and Table 16, respectively. The average standard deviations with respect to all discussed similarity/dissimilarity measures obtained during the tests are gathered in Table 17. The last statistics (Table 18) presented here are the all results obtained with, in our opinion, 3 very efficient aggregation operators such as median, voting, and Choquet integral (the latter calculated as in Table 1 best result for FERET dataset).

5.3. **Analysis of the results.** In the case of AT&T dataset and Euclidean distance (Table 1) a dominating function is $f(x,y) = \min(x,y)\min(1, x^p + y^p)$ with $p = 1$, $q = 2$, or inversely, $p = 2$, $q = 1$. Here, we prepared additional series of tests to find the best choice of the parameters p and q. One can see that the best results are obtained when p and q are relatively close to each other (lay near the main diagonal) in between of values 1 and 3 (see Figure 1). On the other hand, in the case of FERET dataset Choquet integral gives very good results (two of its three best versions are built using the mechanism of compensation). Moreover, one can see that the classic functions like median or weighted average may be helpful here. It is worth to recall the definition of fuzzy measure, particularly λ-fuzzy measure. Relatively similar results can be obtained using squared Euclidean

Dataset	Function	Acc.
AT&T	Mod. Hamacher t-norm $T_\lambda(x,y)$ with $\lambda = 0.0001$	82.06
	Power-based OWA (see Table 3)	81.7
	$f(x,y) = \min(x,y)\min(1, x^p + y^q)$ with $p = \frac{1}{2},\ q = \frac{1}{2}$	81.63
	$f(x,y) = \min(x,y)(x^p + y^q - x^p y^q)$ with $p = \frac{1}{2},\ q = \frac{1}{2}$	81.63
	Clayton copula $C_\alpha(x,y) = \max\left(\left(x^{-\alpha} + y^{-\alpha} - 1\right)^{-\frac{1}{\alpha}}, 0\right)$ with $\alpha = \frac{1}{4}$	81.59
	Family $F_\alpha(x,y) = \left(x^{-\alpha} + y^{-\alpha} - 1\right)^{-\frac{1}{\alpha}}$ with $\alpha = \frac{1}{4}$	81.59
	Family $F_\alpha(x,y) = 1 - e^{-\left[(-\log(1-x))^{-\alpha} + (-\log(1-y))^{-\alpha}\right]^{-\frac{1}{\alpha}}}$ with $\alpha = \frac{1}{4}$	81.59
	Dombi t-norm with generator $g_\lambda^T(t) = \left(\frac{1-t}{t}\right)^\lambda$ with $\lambda = \frac{1}{4}$	81.59
	Family $F_\alpha(x,y) = \frac{1}{\left(1+\left(1+\left(\frac{1}{x}-1\right)^\alpha + \left(\frac{1}{y}-1\right)^\alpha\right)^{\frac{1}{\alpha}}\right)}$ with $\alpha = \frac{1}{4}$	81.59

Table 14. Accuracies for χ^2-statistics distance

Dataset	Function	Acc.
AT&T	Mod. Hamacher t-norm $T_\lambda(x,y)$ with $\lambda = \frac{1}{2}$	83.12
	Weighted harmonic mean	83.09
	Weighted radical mean $RM_{w,\gamma}(x_1, \ldots, x_n)$ with $\gamma = 1.01$	83.08
	Mod. Hamacher t-norm $T_\lambda(x,y)$ with $\lambda = 1$	83.07
	Aczél-Alsina t-norm (Gumbel-Hougaard copula) with the generator $g_\lambda^{AA}(t) = \frac{1}{(-\log t)^\lambda}$ with $\lambda = 2$	83.03

Table 15. Accuracies for median of absolute differences

distance (see Table 2). As it can be seen in the table the results for FERET dataset are similar to the case of classic Euclidean distance. However, in the case of AT&T database we have the situation that the multiplication of minima of the form $f(x,y) = \min(x,y)\min(1, x^p + y^p)$ produces the best results. For this function we have prepared the additional set of experiments to find the proper parameters for computations with aggregation of classifiers. Its structure of best results is similar to the plot obtained in case of Euclidean distance (see Figure 2). In the case of modified Euclidean distance (Table 3), we have a domination of various ordered weighted means and families of Frank functions. Repeated results for Power-based OWA related to the parameter r are covered in Figure 3. The most appropriate choices are the values of r less than 0.4. For the FERET dataset the

Dataset	Function	Acc.
AT&T	Aczél-Alsina t-norm (Gumbel-Hougaard copula) with the generator $g_\lambda^{AA}(t) = (-\log t)^{-\lambda}$ with $\lambda = 2$	83.03
	Hamacher t-norm $T_\lambda(x, y)$ with a generator $\log\left(\frac{\lambda+(1-\lambda))t}{t}\right)$ with $\lambda = 0.1$	83.00
	Dombi t-norm $T(x, y) = \frac{1}{1+\left(\left(\frac{1-x}{x}\right)^\alpha+\left(\frac{1-y}{y}\right)^\alpha\right)^{\frac{1}{p}}}$ with $p = 0.5$	82.98
	Family $F_\alpha(x, y) = \left(1+\left(1+\left(\frac{1}{x}-1\right)^\alpha+\left(\frac{1}{y}-1\right)^\alpha\right)^{\frac{1}{\alpha}}\right)^{-1}$ with $\alpha = 2$	82.98
	Family $F_\alpha(x, y) = \exp\left(1-((1-\log x)^\alpha+(1-\log y)^\alpha-1)^{1/\alpha}\right)$ with $\alpha = 2$	82.96

TABLE 16. Accuracies for median of square differences

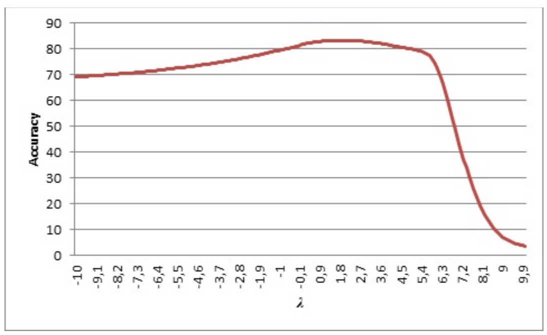

FIGURE 6. Aczél-Alsina t-norm results of application for PCA+LDA-based face recognition (median of square differences)

best results are obtained by using the function $G_\alpha(x, y)$ with $\alpha = 0.99$. However, such a good result can be produced only for α near the value of 1. Table 4 (weighted squared Euclidean distance) shows the importance of many t-norms, combinations of multiplication and power functions, and finally, OWA and OWGM functions. In case of FERET one can see, again, that an application of the Choquet integral to the process of aggregation may be a very good choice. In case of the Manhattan distance (Table 5) we can observe that the modified version of Hamacher t-norm and weighted means are again very good aggregators. As in some previous cases, we have repeated the tests for detailed values of parameters for winning function. The modified Hamacher results comparison is depicted in Figure 4. The plots show that for vast majority of parameters this function is relatively stable aggregator for almost all similarity/dissimilarity measures. Modified Manhattan distance (Table 6) gives other results, namely we see that many various families of means can be chosen. LDA is well supported by weighted radical mean, weighted harmonic mean, and an extension of Wiener – Shannon low. When comparing results for weighted Manhattan distance function (Table 7) we see that they are similar to

Dataset	AT&T	FERET
Distance function	PCA+LDA	PCA +LDA
Euclidean	2.67	2.58
Squared Euclidean	2.81	2.62
Modified Euclidean	2.55	2.43
Weighted squared Euclidean	3.85	2.78
Manhattan	2.42	2.59
Modified Manhattan	2.32	2.51
Weighted Manhattan	3.19	2.75
Chebyshev	3.88	3.05
Cosine	2.03	2.37
Weighted cosine	5.51	2.85
Correlation	2.02	2.38
Bray-Curtis	2.08	2.6
Canberra	2.08	2.6
χ^2-statistics	4.51	-
Median of absolute differences	2.69	-
Median of square differences	2.72	-

TABLE 17. Average standard deviations obtained in tests

Name	Median	Voting	Choquet	Median	Voting	Choquet
Eucl.	88.9	87	87.4	84.5	84	84.6
S. Eucl.	88.4	87	86.5	84.3	82.5	84.6
Manh.	87.8	86.1	86.5	82.9	82.2	83
Cheb.	83.9	81.6	83.3	79	77.8	79.1
Cos.	92.3	88.6	92.7	90.7	91	90.3
Corr.	92.2	88.4	92.6	90.5	90.7	90.1
Bray-C.	91.9	88.4	92.6	90.1	90.1	90
Canb.	89	83.8	91.3	83.3	83.9	83.3
Chi-s.	73.9	77.5	67	51.3	40.9	63.2
M. Eucl.	86.2	86	81.6	83.1	79.3	84.3
M. Manh.	86.7	85.5	84.8	82.9	81.3	83.1
Med. a.	78	72.2	80.2	68.4	67.5	68.3
Med. s.	77.3	72.2	78.9	68.1	64.8	68.4
W. Manh.	77.1	75.7	75.3	77.5	76.3	77.6
W. Eucl.	79.7	79.8	76.7	79.9	77	80.3
W. cos.	78.1	75.1	81.7	87.7	87.8	87.6

TABLE 18. Recognition rates for commonly encountered aggregation operators, i.e. median, voting, and Choquet integral. The first three columns correspond to PCA+LDA on the AT&T dataset, while last three columns refer to PCA+LDA for FERET

the version of the function with no weights. Modified Hamacher t-norm, Gini mean, and various weighted averages give the best accuracies in case of Chebyshev distance (Table 8). Considering cosine measure (Table 9) one can note that voting, median, and various Choquet integrals are the best alternatives in case of the cosine distance for FERET dataset. Weighted geometric mean and modified Hamacher function are the best in case of AT&T database. In case of weighted cosine measure (Table 10) one can observe that the results are very similar to the non-weighted case. Again, good choices are the modified Hamacher t-norm, voting, median, and Choquet integral. Very similar results are obtained for correlation distance (Table 11). Weighted geometric mean, ev-copula are the best for AT&T. The Fisherfaces method for many parts of face can be supported by aggregation functions such as median, voting, and Choquet integral. Choquet integral is dominating aggregation operator in the case of FERET and, well-known from previous examples, $A(x,y)$ and $T(x,y)$ are dominating in the case of AT&T dataset for Bray-Curtis distance (Table 12). A plot showing detailed efficiency of $A(x,y)$ is Figure 5. Next, voting, Choquet integral, and many forms of modified Hamacher functions are good choices in case of Canberra distance (Table 13). In the case of χ^2-statistics distance again, the modification of a Hamacher function and $f(x,y) = \min\left(x\sqrt{y}, y\right)$ are good choices. However, the results at the level of $80-81\%$ accuracy are, in general, not satisfying. Table 14 shows that the median of absolute differences is a helpful aggregating function only in the case of Fisherfaces for AT&T dataset. Here, a modified form of Hamacher t-norm and, among others, weighted harmonic and radical mean are winning in our experiments. Aczél-Alsina t-norm appears in the case of median of square differences (Table 15) and median of absolute differences in top-5 results. However, it does not give the results at a satisfactory level (at least 90% recognition rate). The plot showing the dependence of the recognition rate on the parameter λ is shown in Figure 6. The best choice is the value $\lambda = 1.6$. Let us now consider the average standard deviations with respect to all discussed similarity or dissimilarity measures obtained during the tests (see Table 17). When calculating the averages we have taken into account only these measures for which the overall accuracy was greater than 70%. The standard deviations show that the order of winning functions might be slightly different if the tests would be repeated again. On the other hand, the presence of many functions such as Choquet integral in top-5 results often than one is, in our opinion, not accidental. Finally, the last statistics (Table 18) presented here are the all results obtained with, in our opinion, 3 very efficient aggregation operators such as median, voting, and Choquet integral (the latter calculated as in Table 1 best result for FERET dataset). PCA+LDA for this dataset can be realized by both median and Choquet integral. However, in the most difficult case of PCA+LDA for FERET the Choquet integral wins in 10 similarity measures.

As one can observe, the results are, in general, different for the AT&T and FERET datasets. The reason may be as follows. First, in the case of AT&T there are 10 images per person while in the case of FERET there are only 3 images (one testing and two training). It implies that the general results of classification are relatively worse. Moreover, the variations in expression are greater in the case of FERET. Finally, the FERET images are preprocessed using scaling, cropping and histogram equalization while the AT&T images are only scaled and cropped.

6. Conclusions

In this study, we have selected the best aggregating functions applied to the fundamental face recognition method such as Principal Component Analysis followed by Linear Discriminant Analysis (Fisherfaces) when the area of discussion is constrained to four domains, namely eyes, eyebrows, nose, and mouths. The method is well known and widely used because of the reduction of dimensionality it produces. The representations of the original images come as one-dimensional vectors, which can be compared using the nearest neighbor classifiers. The comparisons are realized using various norms or similarity/dissimilarity measures. Here, we have completed comprehensive comparative studies for 16 measures with intent to deliver experimentally supported conclusions. In many cases, the Choquet integral, voting, median, modifications of Hamacher norms and a few other functions have given very accurate results when aggregating the outputs of four classifiers based on the key above-mentioned facial features.

The future work can be focused on the selection of parameters of the most appropriate aggregation functions using the well-known optimization tools such as genetic algorithms, particle swarm optimization, or differential evolution. Another approach worth considering could be to exploit data fitting to determine the parameters of the aggregation operators (see, e.g., [8]) or to thoroughly examine a relatively new concept of so-called pre-aggregation functions [49]. Finally, it is worth to consider the role of the preprocessing influence on the aggregation processes.

Acknowledgements. The authors are supported by National Science Centre, Poland (grant no. 2014/13/D/ST6/03244). Support from the Canada Research Chair (CRC) program and Natural Sciences and Engineering Research Council is gratefully acknowledged (W. Pedrycz).

References

[1] T. Ahonen, A. Hadid, and M. Pietikäinen, *Face recognition with local binary patterns.* In: Proceedings of the 8th European Conference on Computer Vision, Lecture Notes in Computer Science, **3021** (2004), 469–481.

[2] R. Al-Hmouz, W. Pedrycz, K. Daqrouq, and A. Morfeq, *Development of multimodal biometric systems with three-way and fuzzy set-based decision mechanisms*, International Journal of Fuzzy Systems **20** (1) (2017), 128–140.

[3] F. R. Al-Osaimi, M. Bennamoun, and A. Mian, *Spatially optimized data-level fusion of texture and shape for face recognition*, IEEE Transactions on Image Processing, **21** (2012), 859–872.

[4] C. Alsina, M. J. Frank, and B. Schweizer, *Associative functions. Triangular norms and copulas.* World Scientific, New Jersey, 2006.

[5] AT&T Laboratories Cambridge. The Database of Faces. http://www.cl.cam.ac.uk/ research/dtg/attarchive/facedatabase.html. Accessed 10 May 2016.

[6] A. I. Awad and A. E. Hassanien, *Impact of some biometric modalities on forensic science.* In: A. K. Muda, Y.-H. Choo, A. Abraham, and S. N. Srihari (Eds.), Computational intelligence in digital forensics: Forensic investigation and applications. Studies in Computational Intelligence, **555** (2014), 47–-62.

[7] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, *Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection*, IEEE Transactions on Pattern Analysis and Machine Intelligence, **19** (1997), 711-–720.

[8] G. Beliakov, A. Pradera, and T. Calvo, *Aggregation functions: A guide for practitioners.* Springer-Verlag, Berlin Heidelberg, 2007.

[9] M. Bereta, P. Karczmarek, W. Pedrycz, and M. Reformat, *Local descriptors in application to the aging problem in face recognition*, Pattern Recognition, **46** (2013), 2634–2646.

[10] S. D. Bharkad and M. Kokare, *Performance evaluation of distance metrics: Application to fingerprint recognition*, International Journal of Pattern Recognition and Artificial Intelligence, **25** (2011), 777-–806.

[11] R. Brunelli and T. Poggio, *Face recognition: Features versus templates*, IEEE Transactions on Pattern Analysis and Machine Intelligence, **15** (1993), 1042—1052.

[12] H. Bustince, J. Fernandez, R. Mesiar, and T. Calvo, *Aggregation functions in theory and in practice.* Springer, Berlin, 2013.

[13] T. Calvo, G. Mayor, and R. Mesiar, *Aggregation operators. New trends and applications.* Physica-Verlag, Heidelberg, 2014.

[14] C. Campomanes-Alvarez, O. Ibáñez, and O. Cordón, *Experimental study of different aggregation functions for modeling craniofacial correspondence in craniofacial superimposition.* In: 2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), 437–444, 2016.

[15] C. H. Chan, M. A. Tahir, J. Kittler, and M. Pietikäinen, *Multiscale local phase quantization for robust component-based face recognition using kernel fusion of multiple descriptors*, IEEE Transactions on Pattern Analysis and Machine Intelligence, **35** (2013), 1164–1177.

[16] C. Chen, A. Dantcheva, and A. Ross, *An ensemble of patch-based subspaces for makeup-robust face recognition*, Information Fusion, **32** (2016), 80-–92.

[17] J. H. Chiang, *Aggregating membership values by a Choquet-fuzzy-integral based operator*, Fuzzy Sets and Systems, **114** (2000), 367—375.

[18] J. A. Dargham, A. Chekima, and M. Hamdan, *Hybrid component-based face recognition system.* In: S. Omatu et al. (Eds.), Distributed computing and artificial intelligence, Advances in Intelligent and Soft Computing, **151**, 573-–580, 2012.

[19] S. Das and D. Guha, *Power harmonic aggregation operator with trapezoidal intuitionistic fuzzy numbers for solving MAGDM problems*, Iranian Journal of Fuzzy Systems, **12 (6)** (2015), 41-74.

[20] M. Dolecki, P. Karczmarek, A. Kiersztyn, and W. Pedrycz, *Utility functions as aggregation functions in face recognition.* In: Proc. 2016 IEEE Symposium Series on Computational Intelligence (SSCI 2016), 2016.

[21] H. K. Ekenel and R. Stiefelhagen, *Local appearance based face recognition using discrete cosine transform.* In: Proceedings of the 13th European Signal Processing Conference (EUSIPCO), Antalya, 2484-–2488, 2005.

[22] H. K. Ekenel and R. Stiefelhagen, *Generic versus salient region-based partitioning for local appearance face recognition.* In: M. Tistarelli and M. S. Nixon (Eds.), Advances in Biometrics, Lecture Notes in Computer Science, **5558**, 367-–375, 2009.

[23] R. Gottumukkal and V. K. Asari, *An improved face recognition technique based on modular PCA approach*, Pattern Recognition Letters, **25** (2004), 429-–436.

[24] M. Grabisch, *Fuzzy integral in multicriteria decision making*, Fuzzy Sets and Systems, **69** (1995), 279-–298.

[25] M. Grabisch, J.-L. Marichal, R. Mesiar, and E. Pap, *Aggregation functions.* Cambridge University Press, Cambridge, 2009.

[26] J. Haddadnia and M. Ahmadi, *N-feature neural network human face recognition*, Image and Vision Computing, **22** (2004), 1071-–1082.

[27] B. Heisele and V. Blanz, *Morphable models for training a component-based face recognition system.* In: W. Zhao and R. Chellappa (Eds.), Face processing: Advanced modeling and methods, **1055** (2006) 439–462.

[28] B. Heisele, P. Ho, J. Wu, and T. Poggio, *Face recognition: Component-based versus global approaches*, Computer Vision and Image Understanding, **91** (2003), 6-–21.

[29] B. Heisele, T. Serre, and T. Poggio, *A component-based framework for face detection and identification*, International Journal of Computer Vision, **74** (2007), 167–181.

[30] G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller, *Labeled faces in the wild: A database for studying face recognition in unconstrained environments*. In: Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition, ID: inria-**00321923** (2007), 1–14.

[31] G. B. Huang, H. Lee, E. Learned-Miller, *Learning hierarchical representations for face verification with convolutional deep belief networks*. In: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, 2518–2525, 2012.

[32] Z.-H. Huang, W.-J. Li, J. Wang, and T. Zhang, *Face recognition based on pixel-level and feature-level fusion of the top-level's wavelet sub-bands*, Information Fusion, **22** (2015), 95-–104.

[33] G. Jarillo, W. Pedrycz, and M. Reformat, *Aggregation of classifiers based on image transformations in biometric face recognition*, Machine Vision and Applications, **19** (2008), 125-–140.

[34] P. Karczmarek, A. Kiersztyn, and W. Pedrycz, *An evaluation of fuzzy measure for face recognition*. In: L. Rutkowski, M. Korytkowski, R. Scherer, R. Tadeusiewicz, L. Zadeh, and J. Zurada J. (Eds.), ICAISC 2017, LNCS, **10245** (2017), 668–676.

[35] P. Karczmarek, A. Kiersztyn, and W. Pedrycz, *On developing Sugeno fuzzy measure densities in problems of face recognition*, International Journal of Machine Intelligence and Sensory Signal Processing, **2** (2017), 80—96.

[36] P. Karczmarek, A. Kiersztyn, and W. Pedrycz, *Generalized Choquet integral for face recognition*, International Journal of Fuzzy Systems, **20** (3) (2017), 1047–1055.

[37] P. Karczmarek, A. Kiersztyn, W. Pedrycz, and M. Dolecki, *An application of chain code-based local descriptor and its extension to face recognition*, Pattern Recognition, **65** (2017), 26-–34.

[38] P. Karczmarek, W. Pedrycz, M. Reformat, and E. Akhoundi, *A study in facial regions saliency: A fuzzy measure approach*, Soft Computing, **18** (2014), 379-–391.

[39] T.-K. Kim, H. Kim, H., W. Hwang, and J. Kittler, *Component-based LDA face description for image retrieval and MPEG-7 standardisation*, Image and Vision Computing, **23** (2005), 631—642.

[40] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas, *On combining classifiers*, IEEE Transactions on Pattern Analysis and Machine Intelligence, **20** (1998), 226–239.

[41] E. P. Klement and R. Mesiar, *Logical, algebraic, analytic, and probabilistic aspects of triangular norms*. Elsevier, Amsterdam, 2005.

[42] E. P. Klement, R. Mesiar, and E. Pap, *Triangular norms*. Kluwer Academic Publishers, Dordrecht, 2000.

[43] D. Kurach, D. Rutkowska, E. Rakus-Andersson, *Face classification based on linguistic description of facial features*. In: L. Rutkowski, M. Korytkowski, R. Scherer, R. Tadeusiewicz, L. A. Zadeh, and J. M. Zurada (Eds.), ICAISC 2014, Part II, LNAI, **8468** (2014), 155-–166.

[44] K.-C. Kwak and W. Pedrycz, *Face recognition using fuzzy integral and wavelet decomposition method*, IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, **34** (2004), 1666—1675.

[45] K.-C. Kwak and W. Pedrycz, *Face recognition: A study in information fusion using fuzzy integral*, Pattern Recognition Letters, **26** (2005), 719-–733.

[46] Y. Lei, M. Bennamoun, and A. A. El-Sallam, *An efficient 3D face recognition approach based on the fusion of novel local low-level features*, Pattern Recognition, **46** (2013), 24-–37.

[47] W. L. Liu, X. Q. Song, Q. Z. Zhang, S. B. Zhang, *(T) fuzzy integral of multi-dimensional function with respect to multi-valued measure*, Iranian Journal of Fuzzy Systems, **9 (3)** (2012), 111–126.

[48] Z. Liu and C. Liu, *Fusion of color, local spatial and global frequency information for face recognition*, Pattern Recognition, **43** (2010), 2882-–2890.

[49] G. Lucca, J. A., Sanz, G. P. Dimuro, B. Bedregal, R. Mesiar, A. Kolesárová, and H. Bustince, *Pre-aggregation functions: Construction and an application*, IEEE Transactions on Fuzzy Systems, **24** (2016), 260-–272.

[50] G. E. Martínez, P. Melin, O. D. Mendoza, and O. Castillo, *Face recognition with Choquet integral in modular neural networks*. In: O. Castillo, P. Melin, W. Pedrycz, and J. Kacprzyk (Eds.), Recent advances on hybrid approaches for designing intelligent systems. Studies in Computational Intelligence, **547** (2014), 437–449.

[51] G. E. Martínez, P. Melin, O. D. Mendoza, and O. Castillo, *Face recognition with a Sobel edge detector and the Choquet integral as integration method in a modular neural networks*. In: P. Melin, O. Castillo, and J. Kacprzyk (Eds.), Design of intelligent systems based on fuzzy logic, neural networks and nature-inspired optimization. Studies in Computational Intelligence **601** (2015), 59–70.

[52] P. Melin, C. Felix, O. Castillo, *Face recognition using modular neural networks and the fuzzy Sugeno integral for response integration*, International Journal of Intelligent Systems, **20** (2005), 275-–291.

[53] A. R. Mirhosseini, H. Yan, K.-M. Lam, and T. Pham, *Human face image recognition: An evidence aggregation approach*, Computer Vision and Image Understanding, **71** (1998), 213-–230.

[54] T. Murofushi and M. Sugeno, *An interpretation of fuzzy measures and the Choquet integral as an integral with respect to a fuzzy measure*, Fuzzy Sets and Systems, **29** (1989), 201-–227.

[55] C. Naveena, V. N. Manjunath Aradhya, and S. K. Niranjan, *The study of different similarity measure techniques in recognition of handwritten characters*. In: Proceedings of the International Conference on Advances in Computing, Communications and Informatics (ICACCI-2012), ACM, 781-–787, 2012.

[56] S.-K. Oh, S.-H. Yoo, and W. Pedrycz, *Design of face recognition algorithm using PCA -LDA combined for hybrid data pre-processing and polynomial-based RBF neural networks : Design and its application*, Expert Systems with Applications, **40** (2013), 1451—1466.

[57] W. Pedrycz and F. Gomide, *An introduction to fuzzy sets: Analysis and design*. The MIT Press, Cambridge, 1998.

[58] A. Pentland, B. Moghaddam, and T. Starner, *View-based and modular eigenspaces for face recognition*. In: Computer Vision and Pattern Recognition. Proceedings CVPR '94., 1994 IEEE Computer Society Conference on, 84–91, 1994.

[59] V. Perlibakas, *Distance measures for PCA-based face recognition*, Pattern Recognition Letters, **25** (1994), 711—724.

[60] P. J. Phillips, J. Wechsler, J. Huang, and P. Rauss, *The FERET database and evaluation procedure for face recognition algorithms*, Image and Vision Computing, **16** (1998), 295—306.

[61] P. V. W. Radtke, E. Granger, R. Sabourin, and D. O. Gorodnichy, *Skew-sensitive boolean combination for adaptive ensembles – An application to face recognition in video surveillance*, Information Fusion, **20** (2014), 31-–48.

[62] M. Sugeno, *Theory of fuzzy integral and its applications*. Dissertation. Tokyo Institute of Technology, Tokyo, 1974.

[63] Y. Sun, X. Wang, and X. Tang, *Hybrid deep learning for face verification*, IEEE Transactions on Pattern Analysis and Machine Intelligence, **38** (2016), 1997–2009.

[64] Y. Sun, X. Wang, and X. Tang, *Deep learning face representation from predicting 10,000 classes*. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1891–1898, 2014.

[65] P. Tome, R. Vera-Rodriguez, and J. Fierrez, *Analysing facial regions for face recognition using forensic protocols*. In: J. M. Corchado et al. (Eds.), Highlights on practical applications of agents and multi-agent systems, **365** (2013), 223–230.

[66] V. Torra and Y. Narukawa, *Modeling decisions. Information fusion and aggregation operators*. Springer-Verlag, Berlin Heidelberg, 2007.

[67] M. Turk and A. Pentland, *Eigenfaces for recognition*, Journal of Cognitive Neuroscience, **3** (1991), 71-–86.

[68] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, *Robust face recognition via sparse representation*, IEEE Transactions on Pattern Analysis and Machine Intelligence, **31** (2009), 210–227.

[69] Y. Xue, C. S. Tong, and W. Zhang, *Survey of distance measures for NMF-based face recognition.* In: Y. Wang, Y. Cheung, and H. Liu (Eds.), Computational Intelligence and Security, CIS 2006, Lecture Notes in Computer Science, **4456** (2007), 1039–1049.

[70] R. R. Yager and J. Kacprzyk, *The ordered weighted averaging operators: Theory and Applications.* Springer Science+Business Media, New York, 2012.

[71] D. Yan, Y. Shen, Y. Yan, and H. Wang, *Non-linear feature fusion based on polynomial correlation filter for face recognition.* In: Lecture Notes in Computer Science, **8261**, 312–319, 2013.

[72] Y. Yan and L. A. Osadciw, *Intradifference-based segmentation and face identification.* In: A. K. Jain and N. K. Ratha (Eds.), Biometric technology for human identification, Proceedings of SPIE, **5404** (2004), 502-–510.

[73] W. Zhao, R. Chellappa, P. J. Phillips, A. Rosenfeld, *Face recognition: A literature survey*, ACM Computing Surveys, **35** (2003), 399–458.

[74] Q. Zhu and Y. Xu, *Multi-directional two-dimensional PCA with matching score level fusion for face recognition*, Neural Computing and Applications, **23** (2013), 169-–174.

Paweł Karczmarek *, Institute of Institute of Computer Science, Lublin University of Technology, ul. Nadbystrzycka 36B, 20-618 Lublin, Poland, Institute of Mathematics and Computer Science, The John Paul II Catholic University of Lublin, ul. Konstantynów 1H, 20-708, Lublin, Poland

Email address: pawel.karczmarek@gmail.com

Witold Pedrycz, Department of Electrical & Computer Engineering, University of Alberta, Edmonton T6R 2V4 AB, Canada, Department of Electrical and Computer Engineering, Faculty of Engineering, King Abdulaziz University, Jeddah, 21589, Saudi Arabia, Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland

Email address: wpedrycz@ualberta.ca

Adam Kiersztyn, Institute of Institute of Computer Science,, Lublin University of Technology, ul. Nadbystrzycka 36B, 20-618 Lublin, Poland, Institute of Mathematics and Computer Science, The John Paul II Catholic University of Lublin, ul. Konstantynów 1H, 20-708, Lublin, Poland

Email address: adam.kiersztyn.pl@gmail.com

Michał Dolecki, Institute of Mathematics and Computer Science, The John Paul II Catholic University of Lublin, ul. Konstantynów 1H, 20-708, Lublin, Poland

Email address: michal.dolecki@kul.pl

*Corresponding Author