

Generalized Choquet Integral for Face Recognition

Paweł Karczmarek¹ · Adam Kiersztyn¹ · Witold Pedrycz^{2,3,4}

Received: 28 April 2017/Revised: 9 June 2017/Accepted: 7 July 2017/Published online: 10 August 2017
© The Author(s) 2017. This article is an open access publication

Abstract In this study, we introduce a recent multicriteria decision theory concept of a new, generalized form of Choquet integral function and its application, in particular to the problem of face classification based on the aggregation of classifiers. Such function may be constructed by a simple replacement of the product used under the Choquet integral sign by any t-norm. This idea brings forward a broad class of aggregation operators, which can be incorporated into the decision-making theory. In this context, in a series of experiments we compare the most known t-norms and thoroughly examine their performance in the process of combining individual classifiers based either on facial regions or classic face recognition methods. Such

kind of generalization can successfully improve the classification process provided that the parameters of the t-norms are carefully adjusted.

Keywords Face recognition · Choquet integral · Generalized Choquet integral · Aggregation · Pre-aggregation functions

1 Introduction

Biometrics and face recognition, in particular, have been one of the most intensively explored computer vision areas in the computer science community. It is motivated by a vast area of applications in border checking, surveillance systems, mobile devices, passport and driver's license verification, etc. A choice of the biometrics method can depend upon an application, hardware, memory resources, time constraints and other factors. Modern approaches based on deep learning [17], sparse representation [46] or multimodal biometrics [20] led to good results. Classic methods such as principal component analysis (PCA [43]), linear discriminant analysis (LDA [6]), local descriptors [1] and others are still being enhanced. Moreover, there exist hybrid constructs such as the proposed in [48] or the proposals incorporating experts [23].

The strategies which save memory resources and potentially improve the accuracy of recognition processes are aggregation techniques. They can be divided into a few relatively similar groups, namely aggregations of classifiers based on particular non-overlapping facial regions where the final classification is a result of combining the specific results. In case of overlapping areas, we may talk information fusion. The next group is constituted by the multimodal biometrics. Slightly different class of methods are

✉ Paweł Karczmarek
pawelk@kul.pl
Adam Kiersztyn
adam.kiersztyn@kul.pl
Witold Pedrycz
wpedrycz@ualberta.ca

¹ Institute of Mathematics and Computer Science, The John Paul II Catholic University of Lublin, ul. Konstantynów 1H, 20-708 Lublin, Poland
² Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6R 2V4, Canada
³ Department of Electrical and Computer Engineering, Faculty of Engineering, King Abdulaziz University, Jeddah 21589, Saudi Arabia
⁴ Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland

the approaches based on aggregation of information at the data level. A vast plethora of applications of fuzzy logic, e.g., [5, 30, 41] offers an attractive possibility of using fuzzy measure-based constructs to aggregate the classification results.

Let us briefly recollect some essential results reported in the literature. In [8], eyes, nose, mouth and the whole face scores are added when template matching strategy is applied producing significant improvement in recognition result. Again, a few aggregation proposals for the eigenfaces method for the selected facial areas are presented in [40]. Classifiers based on the RBF neural networks are fused through the majority rule in [15]. In [19], majority voting and Bayesian product are used to aggregate chosen methods of dimensionality reduction. The work [28] at the decision level suggests the weighted sum rule for fusion of similarity matrices. Utility functions serve as aggregation operators [12]. Three-valued and fuzzy set-based decision mechanisms are proposed as aggregation techniques in multimodal biometrics [2]. Finally, fuzzy measure and fuzzy integral find their applications in many facial and pattern recognition studies as an efficient aggregation techniques, namely [21, 24–26, 31–38]. Additionally, an interesting comparison of triangular norms in terms of their statistical non-distinguishability was present in [16].

Aggregation functions are widely studied because of their applications, mainly in decision-making theory, see [14]. They, in general, should satisfy the condition of monotonicity and special boundary conditions (see next section). We are interested in a generalization of Choquet integral in the context of introduction of some kind of structural flexibility by applying t-norms instead of product under the integral sign. In [9, 29], this weaker class of functions (when compared to the aggregation conditions) was introduced and called pre-aggregation functions. They are monotonic along a fixed direction which is called directional monotonicity.

The main goal of this study is to apply this broad class of functions to the problem of face recognition and, to compare their efficiency as the aggregation vehicle, to the initial Choquet integral operator which is one of the commonly encountered aggregation functions, which was proved in above-mentioned works. In particular, we are interested in building twenty-five classes of this kind of integral which are constructed on the basis of an application of t-norms instead of product under the integral sign. Using four well-known face recognition datasets: FERET [44], AT&T [4], Yale Face Database [47] and cropped version [45] of the Labeled Faces in the Wild [18], we compare the performance of the generalized Choquet integrals in an application to the methods such as PCA, LDA, LBP [1], multiscale block LBP [10, 27], chain code-

based local descriptor (CCBLD, [22]) and full ranking [11] when applied to various facial part images and/or the images of the whole face.

The paper is organized as follows. The general properties of aggregation functions and the role of the fuzzy measure and Choquet integral in the process of face recognition are covered in Sect. 2. Section 3 discusses the experimental results while Sect. 4 presents conclusions and directions for the future studies.

2 Aggregation Functions and a General Processing Scheme

Let us recall the basic definitions and properties of pre-aggregation functions.

Definition 1 ([7]) We say $f : [0, 1]^n \rightarrow [0, 1]$ is an aggregation function if it satisfies the following conditions:

- (i) If $x_i \leq x$, then $f(x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_n) \leq f(x_1, \dots, x_{i-1}, x, x_{i+1}, \dots, x_n)$ for each $x_i, i = 1, \dots, n$;
- (ii) $f(0, \dots, 0) = 0$; (iii) $f(1, \dots, 1) = 1$.

Definition 2 ([29]) A function $f : [0, 1]^n \rightarrow [0, 1]$ is said to be \mathbf{r} -increasing if for all points (x_1, \dots, x_n) and for all $c > 0$

$$f(x_1, \dots, x_n) \leq f(x_1 + cr_1, \dots, x_n + cr_n),$$

where $\mathbf{r} = (r_1, \dots, r_n)$.

Now, we recall the concepts of λ -fuzzy measure and Choquet integral. Since in this study we explore them in the context of face recognition, we use the terminology coming from this application domain.

Definition 3 ([39]) Let $X = \{x_1, \dots, x_n\}$ denote the whole face with x_1, \dots, x_n being the particular facial areas such as eyes, eyebrows, mouth. A set function $g : P(X) \rightarrow [0, 1]$ is called a fuzzy measure if it satisfies the following properties: (i) $g(\emptyset) = 0$; (ii) $g(X) = 1$; (iii) $g(A) \leq g(B)$ for $A \subset B$, where $A, B \in P(X)$.

If A and B are disjoint sets, λ -fuzzy measure [42] is a fuzzy measure that fulfills the following rule: $g(A \cup B) = g(A) + g(B) + \lambda g(A)g(B)$, $\lambda > -1$. The parameter $\lambda \neq 0$ can be uniquely determined based on the following equality [42] $1 + \lambda = \prod_{i=1}^n (1 + \lambda g_i)$, where $g_i = g(\{x_i\})$ are fixed. Using the notation $A_i = \{x_1, \dots, x_i\}$, $A_{i+1} = \{x_1, \dots, x_i, x_{i+1}\}$ for overlapping facial areas, one can write the recursive relation $g(A_{i+1}) = g(A_i) + g_{i+1} + \lambda g(A_i)g_{i+1}$, where $g(A_1) = g_1$.

Definition 4 ([39]) Using the above notation, Choquet integral is defined as follows

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

where the function $h : X \rightarrow [0, 1]$ and $h(x_{n+1}) = 0$. Its values are reordered in a non-decreasing order so that $h(x_i) \geq h(x_{i+1}), i = 1, \dots, n$.

Remark 1 In the experimental part, the following notation [26] is applied: $h(y_{ik}) = \frac{1}{N_k} \sum_{\mu_{ij} \in C_k} \mu_{ij}$ where N_k stands for a number of images in k -th class C_k , and the membership grades μ_{ij} are given by $\mu_{ij} = \frac{1}{1 + \frac{d_{ij}}{\bar{d}_i}}$. Here i is a classifier number, j is the training image index, \bar{d}_i is the average distance within i th classifier, and finally, d_{ij} means a distance between a given testing image and j th facial image within i th classifier.

Theorem 1 ([29]) Let $M : [0, 1]^2 \rightarrow [0, 1]$ satisfy $M(x, y) \leq x, M(x, 1) = x, M(0, y) = 0$ and let M be $(1, 0)$ -increasing. Finally, let, for any fuzzy measure g , a generalized Choquet integral be a function of the form

$$\text{Ch}' \int h \circ g(M) = \sum_{i=1}^n M((h(x_i) - h(x_{i+1})), g(A_i)), \tag{2}$$

where $h(x_{n+1}) = 0$. Then, there exists such a nonzero vector \mathbf{r} that Ch' is \mathbf{r} -increasing and it satisfies the conditions: $f(0, \dots, 0) = 0, f(1, \dots, 1) = 1$, and $\min(x_1, \dots, x_n) \leq \text{Ch}' \leq \max(x_1, \dots, x_n)$.

This theorem suggests a potential applicability of the function (2) as a suitable aggregation operator, particularly in the decision-making problems. Hence, we discuss here a large collection of t-norm families which can be applied as the function $M(x, y)$ [see formula (2)]. As the reference, we have used the monograph [3, p. 72, Table 2.6] and paper [29] where simple functions as minimum $T_M(x, y) = \min(x, y)$, product $T_P(x, y) = xy$, Łukasiewicz t-norm $T_L(x, y) = \max(0, x + y - 1)$, drastic product $T_D(x, y) = y$ or x or 0 for $x = 1, y = 1$, or $x, y \neq 1$, respectively, nilpotent minimum $T_{NM}(x, y) = \min(x, y)$ for $x + y > 1$ or 0 otherwise, and Hamacher product $T_H(x, y) =$

$xy/(x + y - xy)$ for $x, y \neq 0$ and 0 otherwise were considered. For instance, the first family of the t-norms present in [3] is $T_\alpha(x, y) = (\max[x^{-\alpha} + y^{-\alpha} - 1, 0])^{-\frac{1}{\alpha}}, \alpha \in (-\infty, 0) \cup (0, \infty)$. The range of a value of parameter α may depend upon a choice of t-norm.

Example 1 Let us consider the reference data coming from [39]. Namely, let $g_1 = 0.6, g_2 = 0.35, g_3 = 0.05, g_4 = 0.21, g_5 = 0.72$ and $h_1 = 0.1, h_2 = 0.4, h_3 = 0.3, h_4 = 0.7, h_5 = 0.05$. Then, $\text{Ch}' \int h \circ g (T_P) \approx 0.31, \text{Ch}' \int h \circ g (T_M) \approx 0.61, \text{Ch}' \int h \circ g (T_L) \approx 0.05, \text{Ch}' \int h \circ g (T_D) \approx 0.31, \text{Ch}' \int h \circ g (T_{NM}) \approx 0.05$, and $\text{Ch}' \int h \circ g (T_H) \approx 0.5$.

Now, let us present a general processing scheme. First, a face image is partitioned onto subimages corresponding to specific facial parts or this facial image (a whole face) is treated by a few transformation algorithms. Each of the algorithms or feature-based classifiers produces the distances between the testing image and training images, of course in a new feature space such as, for instance, new vectors generated by PCA transform corresponding to particular images. The distances obtained in the classification processes form an input to the Choquet integral with an assumption that the values $h(\cdot)$ are defined as mentioned in Remark 1. The final decision about the belongingness to a particular class is made on the basis of the values of the Choquet integral. The highest value is suspected to point at the sought class, see Fig. 1. It is worth noting here that the function $M(\cdot, \cdot)$ appearing in (2) is replaced by any two-argument t-norm.

3 Experimental Studies

In the first series of experiments, we used the AT&T set of facial images. The faces were initially cropped and scaled. Next, we asked 18 people (members of our lab or friends) to provide the weights of saliency of six chosen facial parts considered in the processes of human or computer face

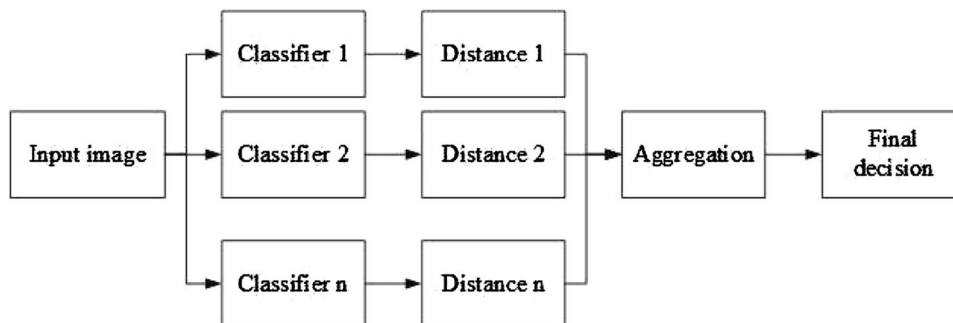


Fig. 1 An overall processing scheme

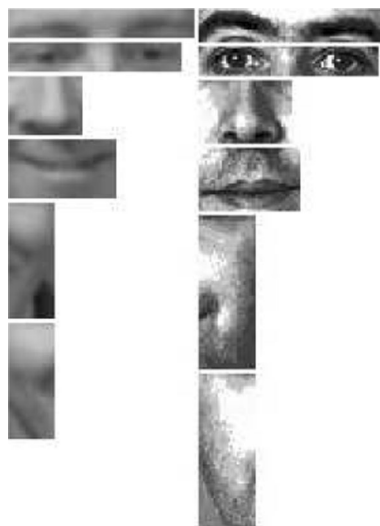


Fig. 2 Facial regions selected from the AT&T (*left*) and FERET (*right*) images, for details, see [24]

recognition. According to their answers, we obtained the average weights shown in Table 1. Examples of related facial regions taken from the AT&T and FERET gray-scale subset (after scaling, cropping the face, and histogram equalization) are depicted in Fig. 2. Next, we conducted a series of 100 classification processes for the well-known methods such as PCA (separately, for the two norms serving as distances between vectors representing facial features after the image transformation, which are the Euclidean and Canberra distances). In each series of experiments, five images of each person were randomly selected to the training set and the rest five to the testing set. In Table 2, listed are the values of parameter α for corresponding t-norm families for which the average classification results were higher than for the classic Choquet integral with product t-norm. Moreover, the maximal difference and its corresponding argument are presented. It is worth to stress here that the obtained values of classification are not important in this comparison, since the Choquet integral has demonstrated to be one of the best aggregation operators used in face recognition, see, for instance [26]. Hence, only the positive differences are discussed to find an alternative operator, if any. Additionally, two functions which can be considered as aggregation

Table 1 Average weights of facial parts

Part	Weight
Eyebrows	0.13
Eyes	0.32
Nose	0.2
Mouth	0.21
Left cheek	0.07
Right cheek	0.07

Table 2 AT&T PCA

No	Euclidean			Canberra		
	Interval	Arg	Max	Interval	Arg	Max
1	[-0.7, -0.1]	-0.6	2.23	[0.1, 0.2]	0.1	0.23
2	[1.2, 1.6]	1.4	0.85	-	-	-
3	[-10.0, -0.1]	-4.2	4.25	[0.1, 0.3]	0.2	0.23
4	[0.6, 0.9]	0.7	3.23	1.1	1.1	0.15
5	[-10.0, -0.1]	-5.7	3.18	[0.2, 0.8]	0.4	0.25
6	[0.2, 0.9]	0.4	2.98	[1.1, 1.3]	1.2	0.15
7	[0.1, 0.5]	0.3	1.58	-	-	-
8	[1.5, 3.4]	1.9	1.43	-	-	-
9	[0.3, 10.0]	0.8	2.93	-	-	-
10	[0.1, 10.0]	1.4	1.98	-	-	-
11	[0.1, 0.6]	0.3	2.48	-	-	-
12	[0.3, 0.5]	0.4	1.95	-	-	-
14	[0.1, 0.6]	0.4	2.78	-	-	-
15	[0.1, 0.9]	0.1	2.58	[1.1, 1.3]	1.2	0.23
17	[1.1, 1.4]	1.2	1.65	-	-	-
20	[-10.0, -1.1]	-9.8	2.98	[-0.8, 0.1]	-0.3	0.25
24	[1.1, 1.5]	1.2	1.55	-	-	-
25	[0.1, 1.4]	0.8	2.60	-	-	-
m	-	-	14.25	-	-	-
v	-	-	8.95	-	-	-

operators, namely median and voting (denoted in the resulting tables as m and v, respectively), were compared. Similar statistics are presented in Table 3, where LDA method with three norms (Euclidean, cosine and Canberra) was incorporated. Tables 4 and 5 present the analogous results for six local descriptors: LBP, MBLBP with 3, 5 and 7 pixel width block sizes, full ranking and CCBLD, respectively. The descriptors here were used in their simplest forms without division of the images into subregions. Relatively similar tests were conducted for the FERET database. In 100 repetitions, two images were randomly selected to the training and one image was chosen to the testing set, respectively. The results for selected three norms, namely Euclidean, cosine and correlation, are listed (see Table 6). Finally, a series of experiments for Yale dataset, where all above-mentioned local descriptors were compared with their best settings for the whole images, is established. Five images of each person were randomly selected to constitute the training set in each of the 100 repetitions. Eight methods (PCA and LDA with Euclidean norm) and six local descriptors were tested for the PCA whole faces with the same protocol as previously. Finally, the LFW dataset was checked for CCBLD, full ranking, MBLBP with 5 px width blocks and simple LBP. The images of people having six photographs were selected.

Table 3 LDA (AT&T)

No	Euclidean			Cosine			Canberra		
	Interval	Arg	Max	Interval	Arg	Max	Interval	Arg	Max
1	[0.1, 0.2]	0.1	0.23	[0.1, 0.3]	0.2	0.40	[0.1, 0.2]	0.1	0.48
3	[0.1, 0.4]	0.2	0.28	[0.3, 0.7]	0.4	0.15	[0.1, 0.5]	0.4	0.28
4	1.1	1.1	0.28	[1.1, 1.3]	1.2	0.20	[1.1, 1.2]	1.1	0.15
5	[0.1, 1.3]	0.4	0.28	[0.5, 2.4]	0.8,	0.13	[0.2, 1.3]	0.6	0.18
6	[1.1, 1.5]	1.3	0.30	[1.2, 1.6]	1.4	0.10	[1.3, 1.4]	1.3	0.03
10	[5.8, 10.0]	8.5	0.13	–	–	–	–	–	–
12	–	–	–	[0.5, 0.6]	0.5	0.28	0.4	0.4	0.45
14	–0.1	–0.1	0.30	–0.1	–0.1	0.03	–0.1	–0.1	0.03
15	[1.1, 1.3]	1.2	0.28	[1.2, 2.0]	1.4	0.28	[1.1, 1.4]	1.3	0.35
20	[–0.8, 0.8]	–0.1	0.38	[–0.3, 2.4]	0.6	0.18	[–0.7, 1.0]	0.1	0.18
23	0.1	0.1	0.05	0.1	0.1	0.38	0.1	0.1	0.13
m	–	–	1.25	–	–	–	–	–	–
v	–	–	1.65	–	–	–	–	–	–

Table 4 Local descriptors (AT&T)

No	LBP			MBLBP (3px block)			MBLBP (5px block)		
	Interval	Arg	Max	Interval	Arg	Max	Interval	Arg	Max
1	–	–	–	–0.1	–0.1	0.45	[–0.2, –0.1]	–0.1	0.28
3	[–0.3, –0.2]	–0.2	0.08	[–1.3, –0.1]	–0.7	0.50	[–2.3, –0.1]	–1.3	0.85
4	0.9	0.9	0.05	0.9	0.9	0.53	[0.8, 0.9]	0.8	0.78
5	[–0.6, –0.3]	–0.5	0.10	[–1.8, –0.1]	–1.0	0.48	[–2.6, –0.1]	–0.8	0.73
6	0.8	0.8	0.13	[0.6, 0.8]	0.8	0.43	[0.5, 0.9]	0.7	1.03
9	–	–	–	[0.1, 10.0]	7.9	0.53	[2.9, 10.0]	5.4	0.60
10	[4.2, 10.0]	7.3	0.15	[0.1, 10.0]	0.2	0.55	[0.1, 10.0]	2.2	0.93
11	–	–	–	0.1	0.1	0.05	0.1	0.1	0.43
14	–	–	–	0.1	0.1	0.33	0.1	0.1	0.28
15	0.9	0.9	0.13	[0.6, 0.9]	0.7	0.50	[0.5, 0.9]	0.6	0.55
20	[–1.9, –1.4]	–1.5	0.13	[–3.8, –1.1]	–2.6	0.53	[–5.1, –1.1]	–3.5	0.75
25	–	–	–	[0.1, 0.2]	0.1	0.53	[0.1, 0.2]	0.2	0.45
m	–	–	1.68	–	–	2.43	–	–	4.15
v	–	–	–	–	–	–	–	–	0.93

Four of the images were randomly selected to the training set and one was selected to the testing set. The weights needed to construct the Choquet integral were initially obtained in the pretests for each of the classifiers separately. The three last sets of test results are gathered in Table 7. It is worth noting that we checked all the 25 functions with the parameter α not lower than -10 and not higher than 10 by considering the successive values $-10, -9.9, \dots, 9.9, 10$, assuming that such range for the parameter represents a satisfactory level of covering of its values. The results gathered in Tables 2, 3, 4, 5, 6 and 7 show the potential value of a few families of functions, namely 1st, 3rd, 4th, 5th, 6th, 9th, 10th, 11th, 14th, 15th,

20th and 25th (their indexes correspond to the indexes introduced in [3, p. 72, Table 2.6]) function as the aggregation operator. Moreover, the median and voting functions show that can be substitutes of Choquet integral. However, Table 8 shows the finding which can be relatively surprising in this context. Namely, only the function no. 10, i.e.,

$$T_\alpha(x, y) = \frac{xy}{[1 + (1 - x^\alpha)(1 - y^\alpha)]^{\frac{1}{\alpha}}}, \alpha \in (0, \infty), \tag{3}$$

for $\alpha > 3.1$ and median give slightly higher recognition rates when the averages of all 17 considered test cases are taken into consideration. This means that the functions

Table 5 Local descriptors 2 (AT&T)

No	MBLBP (7px block)			Full ranking			CCBLD		
	Interval	Arg	Max	Interval	Arg	Max	Interval	Arg	Max
1	[-0.2, -0.1]	0.1	0.9	-	-	-	0.1	0.1	0.20
3	[-3.6, -0.1]	-0.9	1.48	[-0.4, -0.3]	-0.3	0.05	[-0.2, 0.3]	0.3	0.30
4	[0.7, 0.9]	0.8	1.28	-	-	-	1.1	1.1	0.23
5	[-3.4, -0.1]	-1.4	1.53	[-0.6, -0.5]	-0.6	0.10	[-0.5, 1.0]	0.7	0.23
6	[0.4, 0.9]	0.7	1.55	[0.7, 0.9]	0.7	0.05	[0.8, 1.3]	1.1	1.13
9	[2.1, 10.0]	4.5	1.38	-	-	-	-	-	-
10	[0.1, 10.0]	0.5	1.55	[1.9, 10.0]	8.7	0.10	[3.3, 10.0]	9.7	0.18
11	0.1	0.1	1.05	-	-	-	-	-	-
12	0.3	0.3	0.05	-	-	-	-	-	-
14	0.1	0.1	1.00	-	-	-	-0.1	-0.1	0.03
15	[0.3, 0.9]	0.6	1.43	0.9	0.9	0.03	[0.9, 1.3]	1.2	0.25
20	[-1.9, -1.8]	-1.9	0.08	-	-	-	[-1.7, 0.4]	0.2	0.25
25	[0.1, 0.3]	0.2	1.30	-	-	-	-	-	-
m	-	-	3.83	-	-	4.48	-	-	3.10
v	-	-	2.18	-	-	-	-	-	1.85

Table 6 LDA (FERET)

No	Euclidean			Cosine			Correlation		
	Interval	Arg	Max	Interval	Arg	Max	Interval	Arg	Max
1	[0.1, 0.4]	0.2	1.08	[0.1, 0.5]	0.2	0.78	[0.1, 0.4]	0.1	0.63
3	[0.1, 0.6]	0.5	0.80	[0.1, 0.6]	0.2	0.45	[0.1, 0.7]	0.3	0.33
4	[1.1, 1.2]	1.2	0.43	[1.1, 1.2]	1.1	0.15	[1.1, 1.2]	1.1	0.18
5	[0.1, 1.9]	0.9	0.60	[0.1, 1.6]	1.1	0.45	[0.1, 1.7]	0.5	0.33
6	1.3	1.3	0.05	-	-	-	-	-	-
10	[5.4, 10.0]	10.0	0.35	-	-	-	-	-	-
12	[0.4, 0.6]	0.5	1.15	[0.4, 0.7]	0.5	0.75	[0.4, 0.6]	0.5	0.78
14	-0.1	-0.1	0.03	-0.1	-0.1	0.40	-0.1	-0.1	0.20
15	[1.1, 1.9]	1.4	0.93	[1.1, 2.0]	1.4	0.68	[1.1, 1.9]	1.3	0.50
18	0.1	0.1	0.43	[0.1, 0.3]	0.1	1.03	0.1	0.1	0.80
20	[-0.9, 2.0]	0.7	0.83	[-0.9, 1.7]	0.2	0.60	[-0.9, 1.8]	-0.1	0.40
23	0.1	0.1	1.2	[0.1, 0.2]	0.1	0.73	[0.1, 0.2]	0.2	0.75

produced the best accuracy rates when substituted into (2) in the place of the function $M(\cdot, \cdot)$. For each method (classifier), the interval for which a given aggregating function gives the best results varies. Therefore, it is difficult to explicitly predict its value. Even median does not produce the results totally outclassing the rest of functions. On the other hand, the functions no. 16, 19 and 22 have not appeared in any tables; namely, they have never hit the results produced by the classic Choquet integral.

4 Conclusions and Future Work

In the study, we have investigated 25 families of well-known t-norms which can replace the product used in the definition of Choquet integral serving as the aggregation operator for the classifiers in face recognition. The classifiers used various facial parts such as eyebrows, eyes, nose, mouth, left and right cheek or the methods such as principal component analysis, linear discriminant analysis,

Table 7 Various techniques

No	Yale			FERET			LFW		
	Interval	Arg	Max	Interval	Arg	Max	Interval	Arg	Max
1	−0.1	−0.1	0.15	[−0.6, −0.1]	−0.3	0.11	–	–	–
2	[3.3, 10.0]	5.8	2.13	–	–	–	[2.9, 5.6]	3.9	0.03
3	[−0.3, 0.3]	−0.2	0.02	[−10, −0.1]	−9.4	0.09	–	–	–
5	[−10, −0.1]	0.5	0.03	[−10, −0.1]	−5.8	0.10	–	–	–
6	2.3	2.3	0.02	–	–	–	–	–	–
7	–	–	–	0.1	0.1	0.05	–	–	–
8	[7.7, 10.0]	10.0	0.95	[2.7, 6.2]	4.8	0.08	[8.2, 10]	10.0	0.02
9	[4.6, 10.0]	7.9	0.09	[0.3, 10]	1.3	0.11	[5.5, 5.9]	5.5	0.01
10	[0.1, 0.2]	0.2	0.03	[0.1, 10]	0.5	0.06	–	–	–
11	–	–	–	[0.1, 0.4]	0.2	0.11	–	–	–
13	–	–	–	0.1	0.1	0.01	–	–	–
14	[−10, −6.5]	−7.4	1.39	[−0.1, 0.5]	0.2	0.11	–	–	–
15	[0.7, 0.9]	0.9	0.04	[0.1, 0.9]	0.1	0.11	–	–	–
17	[1.8, 2.7]	2.1	1.36	2.0	2.0	0.01	[1.7, 2.2]	1.8	0.01
20	[−1.8, −1.1]	−1.8	0.04	[−10, −1.1]	−10.0	0.11	–	–	–
21	–	–	–	–	–	–	[1.4, 4.1]	2.2	0.02
24	[2.1, 4.6]	2.9	1.37	–	–	–	[1.9, 2.8]	2.2	0.02
25	[0.1, 0.2]	0.1	0.09	[0.1, 1.1]	0.5	0.11	–	–	–
m	–	–	2.04	–	–	0.05	–	–	0.01
v	–	–	2.51	–	–	0.08	–	–	0.03

Table 8 Average results

No	Interval	Arg	Max
10	[3.1, 10.0]	9.8	0.15
m	–	–	0.77

local descriptors based on binary or string words. We selected a single family of functions (3) which can potentially replace the product. Moreover, a collected set of data can serve as the initial values of parameters for the functions considered as potential aggregation operators. The future work can be an application of optimization methods to find the proper parameter value, an application of generalized Choquet integrals to other than face recognition classification problems, and improvement in other fuzzy integrals such as Sugeno or Shilkret [13]. Furthermore, a problem of optimal choice of classifiers taking part in the aggregation process is still open; for instance, the question about the number of facial parts and their saliency in relation to the kind of aggregation operator has not been fully addressed in the literature of the area.

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).

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Ahonen, T., Hadid, A., Pietikäinen, M.: Face recognition with local binary patterns. In: Proceedings of 8th Europ. Conf. Computer Vision, LNCS, 3021, pp. 469–481 (2004)
- Al-Hmouz, R., Pedrycz, W., Daqrouq, K., Morfeq, A.: Development of multimodal biometric systems with three-way and fuzzy set-based decision mechanisms. *Int. J. Fuzzy Syst.* (2017). doi:10.1007/s40815-017-0299-9
- Alsina, C., Frank, M.J., Schweizer, B.: *Associative Functions. Triangular Norms and Copulas.* World Scientific, Hackensack (2006)
- AT&T Laboratories Cambridge: The database of faces. <http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html> (2017). Accessed 6 April 2017
- Asadi, H., Kaboli, S.H.A., Mohammadi, A., Oladazimi, M.: Fuzzy-control-based five-step Li-ion battery charger by using AC impedance technique. In: Fourth International Conference on Machine Vision: Machine Vision, Image Processing, and Pattern Analysis, pp. 1–6 (2012)

6. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eigenfaces vs. fisherfaces: recognition using class specific linear projection. *IEEE Trans. Pattern Anal. Mach. Intell.* **19**, 711–720 (1997)
7. Beliakov, G., Pradera, A., Calvo, T.: *Aggregation Functions: A Guide for Practitioners*. Springer, Berlin (2007)
8. Brunelli, R., Poggio, T.: Face recognition: features versus templates. *IEEE Trans. Pattern Anal. Mach. Intell.* **15**, 1042–1052 (1993)
9. Bustince, H., Sanz, J.A., Lucca, G., Dimuro, G.P., Bedregal, B., Mesiar, R., Kolesárová, A., Ochoa, G.: Pre-aggregation functions: definition, properties and construction methods. In: 2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), pp. 294–300 (2016)
10. Chan, C.-H., Kittler, J., Messer, K.: Multi-scale local binary pattern histograms for face recognition. In: ICB 2007, LNCS 4642, pp. 809–818 (2007)
11. Chan, C.H., Yan, F., Kittler, J., Mikolajczyk, K.: Full ranking as local descriptor for visual recognition: a comparison of distance metrics on sn. *Pattern Recognit.* **48**, 1328–1336 (2015)
12. Dolecki, M., Karczmarek, P., Kiersztyn, A., Pedrycz, W.: Utility functions as aggregation functions in face recognition. In: 2016 IEEE Symposium Series on Computational Intelligence (SSCI), pp. 1–6 (2016)
13. Gagolewski, M.: *Data Fusion. Theory, Methods, and Applications*. Institute of Computer Science, Polish Academy of Sciences, Warsaw (2015)
14. Grabisch, M., Marichal, J.-L., Mesiar, R., Pap, E.: *Aggregation Functions*. Cambridge University Press, Cambridge (2009)
15. Haddadnia, J., Ahmadi, M.: N-feature neural network human face recognition. *Image and Vis. Comput.* **22**, 1071–1082 (2004)
16. Hu, X., Pedrycz, W., Wang, X.: Comparative analysis of logic operators: a perspective of statistical testing and granular computing. *Int. J. Approx. Reason.* **66**, 73–90 (2015)
17. Huang, G.B., Lee, H., Learned-Miller, E.: Learning hierarchical representations for face verification with convolutional deep belief networks. In: *Computer Vision and Pattern Recognition (CVPR)*, 2012 IEEE Conference on, pp. 2518–2525 (2012)
18. Huang, G.B., Ramesh, M., Berg, T., Learned-Miller, E.: *Labeled Faces in the wild: a database for studying face recognition in unconstrained environments*, University of Massachusetts, Amherst, Technical Report 07–49 (2007)
19. Jarillo, G., Pedrycz, W., Reformat, M.: Aggregation of classifiers based on image transformations in biometric face recognition. *Mach. Vis. Appl.* **19**, 125–140 (2008)
20. Kakadiaris, I.A., Passalis, G., Theoharis, T., Toderici, G., Konstantinidis, I., Murtuza, N.: Multimodal face recognition: combination of geometry with physiological information. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), vol. 2, pp. 1022–1029 (2005)
21. Karczmarek, P., Kiersztyn, A., Pedrycz, W.: An evaluation of fuzzy measure for face recognition. In: *ICAISC 2017: Artificial Intelligence and Soft Computing*, pp. 668–676 (2017)
22. Karczmarek, P., Kiersztyn, A., Pedrycz, W., Dolecki, M.: An application of chain code-based local descriptor and its extension to face recognition. *Pattern Recognit.* **65**, 26–34 (2017)
23. Karczmarek, P., Pedrycz, W., Kiersztyn, A., Rutka, P.: A study in facial features saliency in face recognition: an analytic hierarchy process approach. *Soft Comput.* (2016). doi:10.1007/s00500-016-2305-9
24. Karczmarek, P., Pedrycz, W., Reformat, M., Akhouni, E.: A study in facial regions saliency: a fuzzy measure approach. *Soft Comput.* **18**, 379–391 (2014)
25. Kwak, K.-C., Pedrycz, W.: Face recognition using fuzzy integral and wavelet decomposition method. *IEEE Trans. Syst. Man Cybern. Part B Cybern.* **34**, 1666–1675 (2004)
26. Kwak, K.-C., Pedrycz, W.: Face recognition: a study in information fusion using fuzzy integral. *Pattern Recognit. Lett.* **26**, 719–733 (2005)
27. Liao, S., Zhu, X., Lei, Z., Zhang, L., Li, S.: Learning multi-scale block local binary patterns for face recognition. In: *Advances in Biometrics. International Conference, ICB 2007, LNCS 4642*, pp. 828–837 (2007)
28. Liu, Z., Liu, C.: Fusion of color, local spatial and global frequency information for face recognition. *Pattern Recognit.* **43**, 2882–2890 (2010)
29. Lucca, G., Sanz, J.A., Dimuro, G.P., Bedregal, B., Mesiar, R., Kolesárová, A., Bustince, H.: Preaggregation functions: construction and an application. *IEEE Trans. Fuzzy Syst.* **24**, 260–272 (2016)
30. Mansouri, M., Kaboli, S.H.A., Ahmadian, J., Selvaraj, J.: A hybrid neuro-fuzzy P.I. speed controller for B.L.D.C. enriched with an integral steady state error eliminator. In: 2012 IEEE International Conference on Control System, Computing and Engineering, pp. 234–237 (2012)
31. Martínez, G.E., Melin, P., Mendoza, O.D., Castillo, O.: Face recognition with Choquet integral in modular neural networks. In: Castillo, O., Melin, P., Pedrycz, W., Kacprzyk, J. (eds.) *Recent Advances on Hybrid Approaches for Designing Intelligent Systems. Part III*, pp. 437–449. Springer, Cham (2014)
32. Martínez, G.E., Melin, P., Mendoza, O.D., Castillo, O.: Face recognition with a Sobel edge detector and the Choquet integral as integration method in a modular neural networks. In: Melin, P., Castillo, O., Kacprzyk, J. (eds.) *Design of Intelligent Systems Based on Fuzzy Logic Neural Networks and Nature-Inspired Optimization. Part I*, pp. 59–70. Springer, Cham (2015)
33. Martínez, G.E., Mendoza, O.D., Castro, J.R., Melin, P., Castillo, O.: Choquet integral and interval type-2 fuzzy Choquet integral for edge detection. In: Melin, P., Castillo, O., Kacprzyk, J. (eds.) *Nature-Inspired Design of Hybrid Intelligent Systems. Part I*, pp. 79–97. Springer, Cham (2017)
34. Martínez, G.E., Mendoza, O., Castro, J.R., Rodríguez-Díaz, A., Melin, P., Castillo, O.: Comparison between Choquet and Sugeno integrals as aggregation operators for pattern recognition. In: 2016 Annual Conference of the North American Fuzzy Information Processing Society (NAFIPS), pp. 1–6 (2016)
35. Martínez, G.E., Mendoza, O., Castro, J.R., Rodríguez-Díaz, A., Melin, P., Castillo, O.: Response integration in modular neural networks using Choquet Integral with Interval type 2 Sugeno measures. In: *Proceedings of NAFIPS*, pp. 1–6 (2015)
36. Martínez, G.E., Mendoza, O., Melin, P., Gaxiola, F.: Comparison between Choquet and Sugeno integrals as aggregation operators for modular neural networks. In: 2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), pp. 2331–2336 (2016)
37. Melin, P., Felix, C., Castillo, O.: Face recognition using modular neural networks and the fuzzy Sugeno integral for response integration. *Int. J. Intell. Syst.* **20**, 275–291 (2005)
38. Mirhosseini, A.R., Yan, H., Lam, K.-M., Pham, T.: Human face image recognition: an evidence aggregation approach. *Comput. Vis. Image Underst.* **71**, 213–230 (1998)
39. Pedrycz, W., Gomide, F.: *An Introduction to Fuzzy Sets: Analysis and Design*. The MIT Press, Cambridge (1998)
40. Pentland, A., Moghaddam, B., Starner, T.: View-based and modular eigenspaces for face recognition. In: *Computer Vision and Pattern Recognition. Proc. CVPR*, pp. 84–91 (1994)
41. Saghafinia, A., Kahourzade, S., Mahmoudi, A., Hew, W.P., Uddin, M.N.: On line trained fuzzy logic and adaptive continuous wavelet transform based high precision fault detection of IM with broken rotor bars. In: 2012 IEEE Industry Applications Society Annual Meeting, pp. 1–8 (2012)
42. Sugeno, M.: *Theory of fuzzy integral and its applications*. Dissertation, Tokyo Institute of Technology, Tokyo (1974)

43. Turk, M., Pentland, A.: Eigenfaces for recognition. *J. Cogn. Neurosci.* **3**, 71–86 (1991)
44. Phillips, P.J., Wechsler, J., Huang, J., Rauss, P.: The FERET database and evaluation procedure for face recognition algorithms. *Image Vis. Comput.* **16**, 295–306 (1998)
45. Sanderson, C., Lovell, B.C.: Multi-region probabilistic histograms for robust and scalable identity inference. In: *ICB 2009*. LNCS 5558, pp. 199–208 (2009)
46. Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S., Ma, Y.: Robust face recognition via sparse representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **31**, 210–227 (2009)
47. Yale Face Database. <http://vision.ucsd.edu/content/yale-face-database>. Accessed 6 April 2017
48. Zhai, H., Liu, C., Dong, H., Ji, Y., Guo, Y., Gong, S.: Face verification across aging based on deep convolutional networks and local binary patterns. In: *Intelligence Science and Big Data Engineering. Image and Video Data Engineering*, LNCS 9242, pp. 341–350 (2015)

Paweł Karczmarek received his M.Sc. degree in mathematics from The John Paul II Catholic University of Lublin, Poland (2005), and the Ph.D. degree in mathematics from the University of Gdańsk, Poland (2010). He is currently an assistant professor at Institute of

Mathematics and Computer Science of The John Paul II Catholic University of Lublin. He was a postdoctoral fellow in the Department of Electrical and Computer Engineering of the University of Alberta (2011–2012). His current research areas are pattern recognition, computer vision, granular computing, soft computing, multicriteria decision-making theory and computational mathematics.

Adam Kiersztyn received Ph.D. degree in mathematics from Faculty of Mathematics, Physics and Computer Science, Maria Curie-Skłodowska University, Lublin, Poland. His current research interests include fuzzy measures, face recognition, quantitative methods and decision-making theory. Moreover, he is interested in the application of mathematics in other areas of science like biology or demography.

Witold Pedrycz is Professor and Canada Research Chair in the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, Canada. He is also with the Systems Research Institute of the Polish Academy of Sciences, Warsaw, Poland. His main research directions involve computational intelligence, fuzzy modeling and granular computing, knowledge discovery and data mining, pattern recognition, knowledge-based neural networks, relational computing and software engineering.